

# Symposium on Statistics and Risk Management Science: Data Analytics

## About the Symposium

In this information blooming age, to cope with various challenges from the influx of huge amount of data, it is crucial for having a proper use of and developing the novel data analytical methods. These methods should root in the deep interplay between mathematics, statistics and computer science. To celebrate the 55th anniversary of CUHK, with the generous support from the Faculty of Science, we aim to organize this symposium on data analytics to provide an avenue for researchers from academia and practitioners to exchange ideas, especially on the availability of cutting-edge technologies and the most recent calls for resolution from industries.

The key topics covered range from genomics, health analytics, high dimensional statistics, to FinTech and InsurTech.

**(Fri) November 30, 2018**

Venue: The Chinese University of Hong Kong  
LT, William M W Mong Engineering Building (Morning)  
LT6, Lady Shaw Building (Afternoon)



## Invited Speakers (in alphabetical order)

Ka Chun CHEUNG	Department of Statistics & Actuarial Science, The University of Hong Kong
Yen-Tsung HUANG	Institute of Statistical Science, Academia Sinica
Alfred MA	CASH Algo Finance Group Limited
Ben NG	Coherent Capital Advisors, Ltd
Howell TONG	University of Electronic Science and Technology of China/London School of Economics and Political Science
Maggie Haitian WANG	The Centre for Clinical Research and Biostatistics, The Chinese University of Hong Kong
Can YANG	Department of Mathematics, Hong Kong University of Science and Technology
Xiang ZHOU	School of Data Science and Department of Mathematics, City University of Hong Kong

## Session Chairs

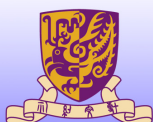
Ngai Hang CHAN	Department of Statistics, The Chinese University of Hong Kong
Qi Man SHAO	Department of Statistics, The Chinese University of Hong Kong
Hoi Ying WONG	Department of Statistics, The Chinese University of Hong Kong
Phillip YAM	Department of Statistics, The Chinese University of Hong Kong



## Organizing Committee

Department of Statistics, The Chinese University of Hong Kong

Free Tea Receptions. No Registration Required.  
Detailed event information are available at  
<http://www.sta.cuhk.edu.hk/Symposium/2018>



香港中文大學理學院  
**FACULTY OF SCIENCE**  
THE CHINESE UNIVERSITY OF HONG KONG

# Symposium on Statistics and Risk Management Science: Data Analytics

## Programme (Part 1)

Date: 30<sup>th</sup> November, 2018 (Friday)

Venue: LT, William M W Mong Engineering Building, The Chinese University of Hong Kong

### Morning Session

Time	Event	Speaker
0855 - 0900	Opening Remark	Xinyuan SONG (The Chinese University of Hong Kong)
0900 - 0940	Session Chair: Ngai Hang CHAN, The Chinese University of Hong Kong	
	Jackknife Approach to the Estimation of Mutual Information	Howell TONG (University of Electronic Science and Technology of China/ London School of Economics and Political Science)
0940 - 1020	A Statistical Method to Quantify the Impact of Genetically Regulated Gene Expression on Complex Traits	Can YANG (The Hong Kong University of Science and Technology)
1020 - 1050	<i>Coffee Break</i>	
1050 - 1130	Session Chair: Hoi Ying WONG, The Chinese University of Hong Kong	
	Implementation Shortfall for Algorithmic Trading	Alfred MA (CASH Algo Finance Group Limited)
1130 - 1210	Control Perspective for Rare Events	Xiang ZHOU (City University of Hong Kong)
1210 - 1230	<i>Photo Session</i>	
1230 - 1430	<i>Lunch (By Invitation only)</i>	

# Symposium on Statistics and Risk Management Science: Data Analytics

## Programme (Part 2)

Date: 30<sup>th</sup> November, 2018 (Friday)

Venue: Lecture Theatre 6, Lady Shaw Building, The Chinese University of Hong Kong

### Afternoon Session

Time	Event	Speaker
1430 - 1510	Session Chair: Qiman SHAO, The Chinese University of Hong Kong	
	Enhancing Power of Association Test in Whole Genome Sequencing Data by a Fuzzy Zoom-Focus Algorithm	Maggie Haitian WANG (Centre for Clinical Research and Biostatistics, The Chinese University of Hong Kong)
1510 - 1550	Causal Mediation of Semicompeting Risks	Yen-Tsung HUANG (Institute of Statistical Science, Academia Sinica)
1550 - 1630	<i>Coffee Break</i>	
1630 - 1710	Session Chair: Phillip YAM, The Chinese University of Hong Kong	
	Prepare for the New Trends in Insurance Analytics	Ben NG (Coherent Capital Advisors, Ltd)
1710 - 1740	Evolutionary Credibility Risk Premium	Ka Chun CHEUNG (The University of Hong Kong)
1740 - 1745	Closing Remark	Phillip YAM (The Chinese University of Hong Kong)
1800 - 2030	<i>Dinner (By Invitation only)</i> <i>The shuttle bus will depart from Lady Shaw Building at 18:00.</i>	

List of Abstracts (in alphabetical order)



**Ka Chun CHEUNG** (*Department of Statistics & Actuarial Science, The University of Hong Kong*)

**Title: Evolutionary Credibility Risk Premium**

*Abstract:* In this talk, we study the problem of the risk premium calibration under an evolutionary credibility model in which both the hypothetical mean and the process variance are simultaneously estimated. The procedure is carried out by minimizing the mean squared loss with respect to both quantities by solving the corresponding normal equations. It is worth noticing that our formulation and procedure are model-free and are different from both the SURE estimate proposed by Xie et al. (2012), which assumes heteroscedastic but known variances, and that by Jing et al. (2016), which focuses on single time period only. We devise an effective recursive \$LU\$ algorithm for solving the sets of normal equations, and study several special cases of the evolutionary model. The superior performance of our procedure compared to existing methodologies will be illustrated through simulation studies. This talk is based on a joint work with Yongzhao Chen, Hugo Choi, Tze Leung Lai, and Phillip Yam.



**Yen-Tsung HUANG** (*Institute of Statistical Science, Academia Sinica*)

**Title: Causal Mediation of Semicompeting Risks**

*Abstract:* The semi-competing risk problem arises when one is interested in the effect of an exposure or treatment on both intermediate (e.g., having cancers) and primary (e.g., death) events where the intermediate event may be censored by the primary event but not vice versa. Here we propose a nonparametric approach by casting the semi-competing risk problem in the framework of causal mediation modeling. We set up a mediation model with the intermediate and primary events, respectively as the mediator and the outcome, and define indirect effect (IE) as the effect of the exposure on the primary event mediated by the intermediate event and direct effect (DE) as that not mediated by the intermediate event. A time-varying weighted Nelson-Aalen type of estimators are proposed for direct and indirect effects where the counting process at time  $t$  of the primary event  $N_{2n_1}(t)$  and its compensator  $A_{n_1}(t)$  are both defined conditional on the status of the intermediate event right before  $t$ ,  $N_1(t^-) = n_1$ . We show that  $N_{2n_1}(t) - A_{n_1}(t)$  is a zero-mean martingale. Based on this, we further establish asymptotic unbiasedness, consistency and asymptotic normality for the proposed estimators. Numerical studies including simulation and data application are presented to illustrate the finite sample performance and utility of the proposed method.

**Keywords:** Causal inference; Causal mediation model; Martingale; Nelson-Aalen estimator; Semi-competing risk

## Symposium on Statistics and Risk Management Science: Data Analytics

 **Alfred MA** (*CASH Algo Finance Group Limited*)

**Title: Implementation Shortfall for Algorithmic Trading**

*Abstract:* Implementation shortfall (Perold, 1988) is defined as the profit and loss difference between paper and real portfolio, and it is composed of execution cost and opportunity cost. However, the framework of Perold (1988) is not directly applicable to algorithmic trading because the method has a rigid requirement on the time stamp of the trade records of the paper and real portfolio. In this paper, we propose a framework to compute implementation shortfall for algorithmic trading. We employ an efficient algorithm inspired by DNA sequence alignment techniques to compute the implementation shortfall with a breakdown of execution cost and opportunity cost. Our proposed framework is simple and computationally efficient. In particular, the complexity of our framework only grows linearly with the number of trades on backtesting or live trading. Hence our framework is applicable to even high frequency trading data.

 **Ben NG** (*Coherent Capital Advisors, Ltd*)

**Title: Prepare for the New Trends in Insurance Analytics**

*Abstract:* Insurance industry hasn't changed in the last 30 years. The product offer and distribution model are basically the same and the adoption of technology is slow compared with other industries. Ironically, there are many Actuaries working in this industry who are experts to apply statistics and mathematic to business. I would like to share my thought about the problem and how Actuaries or Statisticians can disrupt the insurance industry.

Intermediaries are critical in the insurance ecosystem as the communication is not effective in the past. They serve as the bridge between customers and insurance companies. Most insurance executives will think Life Insurance is a people business and reluctant to adopt technology to partially replace the role of intermediaries. It is quite different to other financial institutions and can explain why they can move faster in the digitalization journey.

In retail business, e-commerce grows dramatically in the last 10 years. Alibaba and Amazon becomes the largest retail business companies in the world. Their business model is to cut down the intermediaries and deliver the products to customers directly at low price. Most importantly, these companies analyze the customer data and able to offer a right product at right time. These two companies started to develop their insurance business in a completely different ways of business model. They don't need intermediaries but they need data expert to analyze the data. Unlike other industries, insurance heavily involve statistics and life contingency, only Actuaries and Statisticians really know how to handle those data. The role of Actuaries will not be limited to traditional pricing and valuation roles. Instead, they need understand the customer through data and helping insurance to serve the customers directly. It

## Symposium on Statistics and Risk Management Science: Data Analytics

sounds reasonable and promising but how they can make it. Let me share my thought in the following 30 minutes.



**Howell TONG** (*University of Electronic Science and Technology of China / London School of Economics and Political Science*)

**Title: Jackknife Approach to the Estimation of Mutual Information**

**Abstract:** Quantifying the dependence between two random variables is a fundamental issue in data analysis, and thus many measures have been proposed. Recent studies have focused on the renowned mutual information (MI) [Reshef DN, et al. (2011) *Science* 334:1518–1524]. However, “Unfortunately, reliably estimating mutual information from finite continuous data remains a significant and unresolved problem” [Kinney JB, Atwal GS (2014) *Proc Natl Acad Sci USA* 111:3354–3359]. In this paper, we examine the kernel estimation of MI and show that the bandwidths involved should be equalized. We consider a jackknife version of the kernel estimate with equalized bandwidth and allow the bandwidth to vary over an interval. We estimate the MI by the largest value among these kernel estimates and establish the associated theoretical underpinnings.



**Maggie Haitian WANG** (*Division of Biostatistics and Centre for Clinical Research and Biostatistics (CCRB), JC School of Public Health and Primary Care, The Chinese University of Hong Kong / Centre for Clinical Trials and Biostatistics, CUHK Shenzhen Research Institute*)

**Title: Enhancing Power of Association Test in Whole Genome Sequencing Data by a Fuzzy Zoom-Focus Algorithm**

**Abstract:** The increasing amount of whole exome or genome sequencing data brings forth the challenge of analyzing the association of rare variants that have extremely small minor allele frequencies. Various statistical tests have been proposed, which are specifically configured to increase power for rare variants by conducting the test within a certain bin, such as a gene or a pathway. However, a gene may contain from several to thousands of markers, and not all of them are related to the phenotype. Combining functional and non-functional variants in arbitrary genomic region could impair the testing power. We propose a Fuzzy Zoom-Focus algorithm (fZFA) to locate the optimal testing region within a given genomic region. It can be applied as a wrapper function of existing rare variant association tests to increase testing power. The algorithm is very efficient and the complexity is linear to the number of variants. Simulation studies showed that fZFA substantially increased the statistical power of rare variants tests, including the burden test, SKAT, SKAT-O, and the W-test. The algorithm was applied on real exome sequencing data of hypertensive disorder, and identified biologically relevant genetic markers to metabolic disorder that were undiscoverable by gene-based method. The proposed algorithm is an efficient and powerful tool to enhance the power of association study for whole exome or genome sequencing data.

## Symposium on Statistics and Risk Management Science: Data Analytics



**Can YANG** (*Department of Mathematics, Hong Kong University of Science and Technology*)

**Title: A Statistical Method to Quantify the Impact of Genetically Regulated Gene Expression on Complex Traits**

**Abstract:** Genome-wide association studies (GWAS) have identified many risk variants associated with human complex phenotypes since 2005. However, nearly 90% of the risk variants are located in the non-coding region, highlighting the regulatory role of genetic variants. A scientific hypothesis is that a substantial proportion of risk variants affect complex traits/diseases through regulating expression of their target gene. In this talk, we discuss how to formulate the examination of this scientific hypothesis as a statistical problem and then develop a statistical method to address the challenge associated with this problem.



**Xiang ZHOU** (*School of Data Science and Department of Mathematics, City University of Hong Kong*)

**Title: Control Perspective for Rare Events**

**Abstract:** The rare events in randomly perturbed dynamical systems are very important in physics, chemistry and biology, since they describe the random and infrequent hoppings between metastable states. The traditional studies are based on the large deviation and the underlying variational structure, which have been proved to be successful in understanding the transition mechanism in many applications. Now the dynamical programming and optimal control play a new role in understanding the importance sampling and constructing potentially more useful numerical schemes, in particular in combination with the machine learning techniques. This talk will introduce this perspective, elaborate controlled diffusion process from rest point and general proposals in future.

# Symposium on Statistics and Risk Management Science: Data Analytics

**Note**



# Symposium on Statistics and Risk Management Science: Data Analytics

**Note**

# Symposium on Statistics and Risk Management Science: Data Analytics

**Note**

# Symposium on Statistics and Risk Management Science: Data Analytics

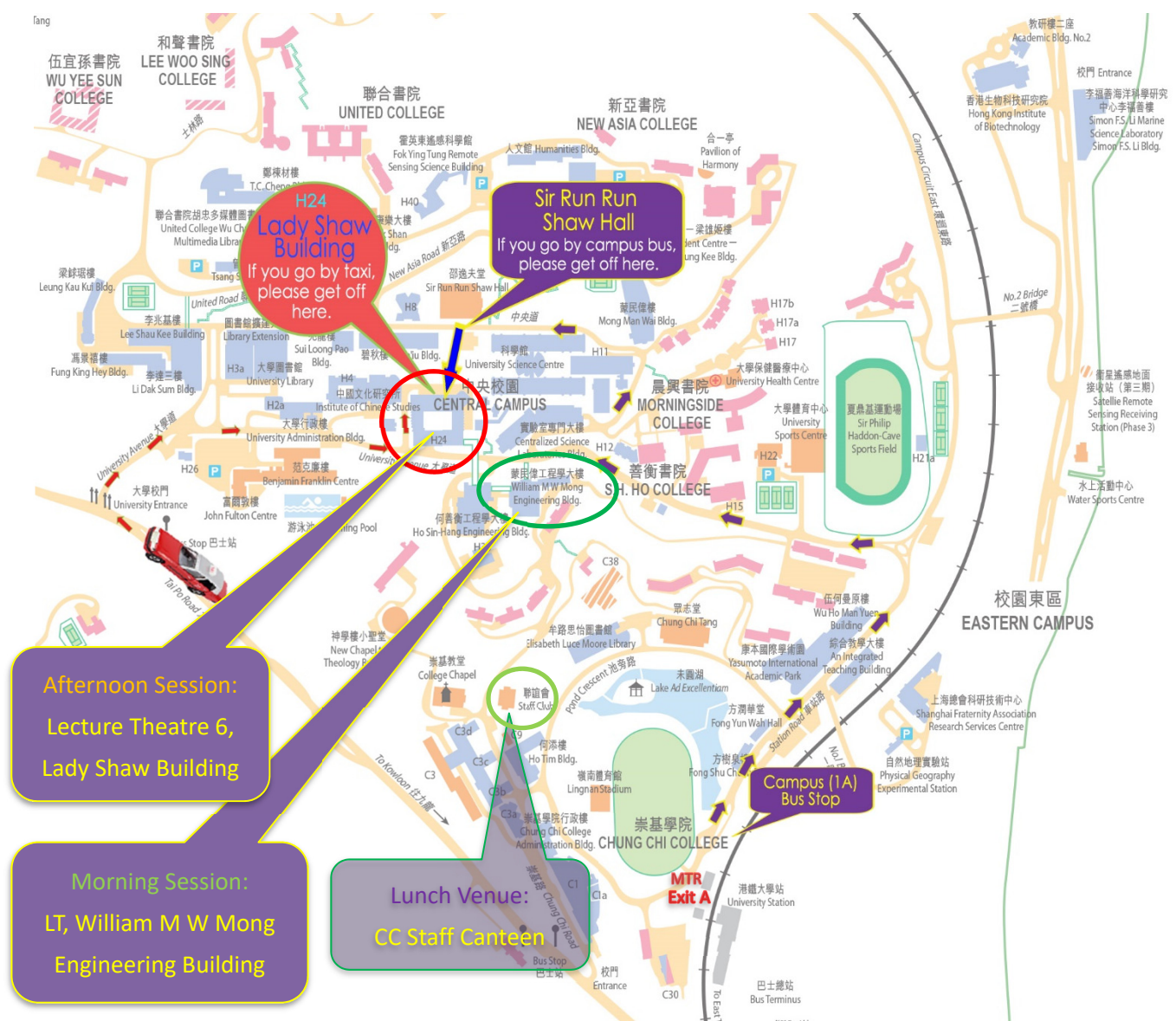
## Campus Map

### Symposium Venue

Morning Session: LT, William M W Mong Engineering Building, The Chinese University of Hong Kong

Afternoon Session: Lecture Theatre 6, Lady Shaw Building, The Chinese University of Hong Kong

Lunch Venue: CC Staff Canteen, The Chinese University of Hong Kong



# Symposium on Statistics and Risk Management Science: Data Analytics

## Map of Banquet Venue

Banquet Venue: Lei Garden Restaurant – 6/F, New Town Plaza, Phase 1, Centre Sha Tin Street, Sha Tin, NT



### *Organizing Committee:*

Department of Statistics, The Chinese University of Hong Kong

### *Email:*

[statistics\\_sym@sta.cuhk.edu.hk](mailto:statistics_sym@sta.cuhk.edu.hk)

### *Phone:*

852-3943 7932 (Ms. Wendy TANG)

852-3943 7952 (Ms. Yanny NG)

### *Correspondence:*

Rm 119, Lady Shaw Building

Department of Statistics

The Chinese University of Hong Kong

Shatin, N.T.

Hong Kong