

HKU-NUS-STANFORD Conference in Statistical Science and Decision Analytics

In Celebration of the 50th Anniversary of the HKU
Department of Statistics and Actuarial Science

March 10 – 11(AM), 2017 (Friday and Saturday)

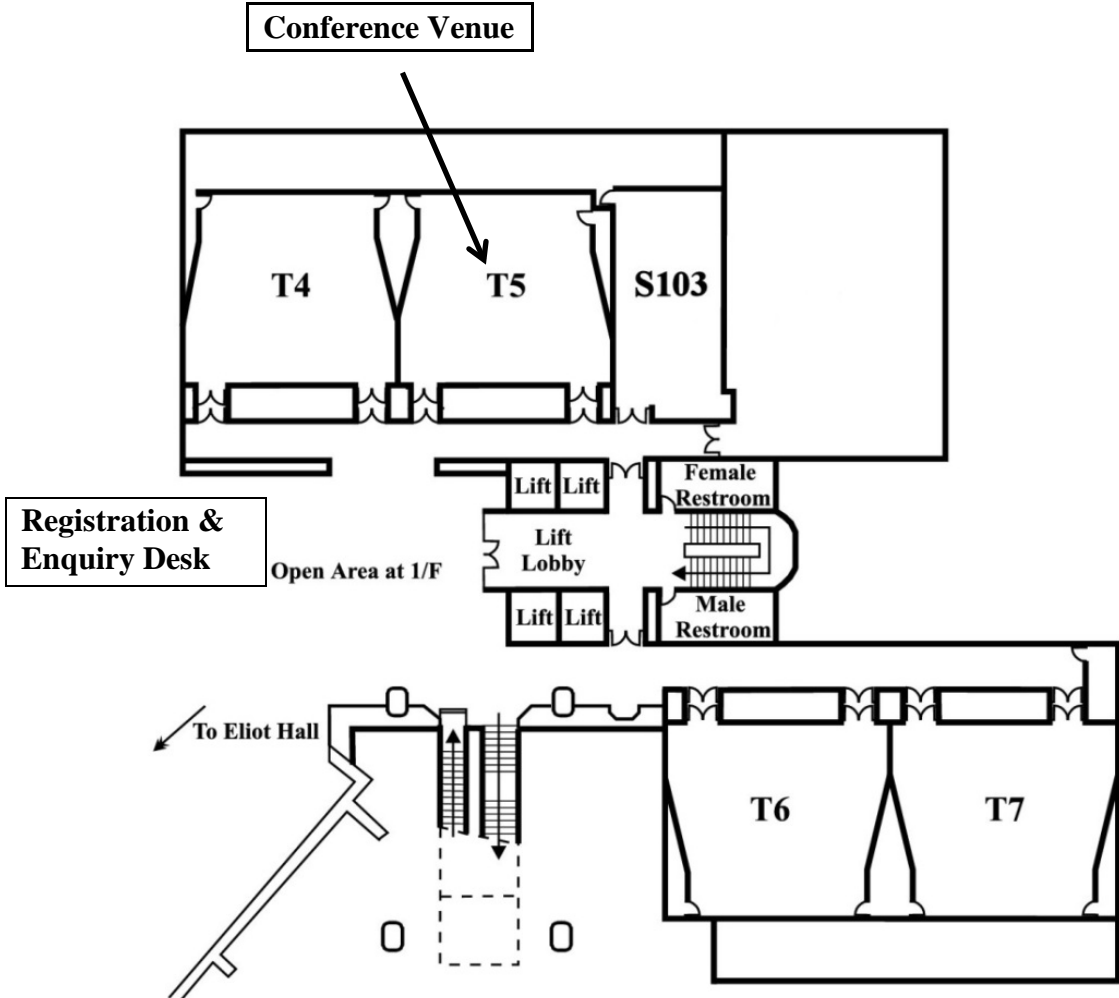
Lecture Theatre 5 (T5), 1/F, Meng Wah Complex
The University of Hong Kong

Programme and Abstracts



MAPS OF HKU CAMPUS

1/F, Meng Wah Complex, HKU



Organizing Committee

LI Wai Keung (Chairman)

CHAN Hock Peng

LAI Tze Leung

The University of Hong Kong, Hong Kong

National University of Singapore, Singapore

Stanford University, USA

GENERAL INFORMATION

Registration & Enquiry Desk

The Registration & Enquiry Desk will operate at the following times and venue:

<u>Date</u>	<u>Time</u>	<u>Venue</u>
Friday, 10 March	08:30 – 09:00	Open area, 1/F, Meng Wah Complex
Saturday, 11 March (AM)	08:45 – 09:00	Open area, 1/F, Meng Wah Complex

If participants require urgent assistance outside the operating times of the Registration & Enquiry Desk, they may visit the Department of Statistics and Actuarial Science, Room 303, 3/F, Run Run Shaw Building, HKU.

Opening Ceremony

Date: Friday, 10 March, 2017

Time: 09:00 – 09:10

Venue: Theatre 5 (T5), 1/F, Meng Wah Complex, HKU

Internet Facilities

The Wi-Fi.HK via HKU programme provides free WiFi service to public and visitors of the University so that users can surf the Internet freely for instructional, learning, research or administrative purposes whenever they are on campus. Connection to Wi-Fi.HK via HKU is easy and no registration is required. However, Wi-Fi.HK service has limited capacity and provides web browsing service only. The service uses unencrypted channel and hence, may not be secure. For further details on Wi-Fi.HK service, please visit link: <http://www.its.hku.hk/documentation/guide/network/wifi/wifihk>

Please read the Terms and Conditions of Use and Disclaimers carefully before using the Wi-Fi.HK via HKU Service provided by the Information Technology Services of the University of Hong Kong (HKU). By using this service, users agree to be bound by the Terms and Conditions which may be amended at any time by the Information Technology Services without prior notice.

Car Parking

Complimentary parking is available on campus. Please take the parking ticket to the Registration & Enquiry Desk for arranging complimentary parking.

No Smoking Policy

Smoking is prohibited in all areas of the campus, including corridors and restrooms.

Safety and Security

Please do not leave your belongings unattended at any times inside or outside the venue.

Disclaimer of Liability

The Organizer of the HKU-NUS-STANFORD Conference in Statistical Science and Decision Analytics will not accept any liability for damages of any nature sustained by participants or their accompanying persons, or loss of or damages to their personal property during the conference.

The Secretariat

For enquiries, please contact the Secretariat of the HKU-NUS-STANFORD Conference in Statistical Science and Decision Analytics:

The Secretariat,
HKU-NUS-STANFORD Conference in Statistical Science and Decision Analytics
Department of Statistics and Actuarial Science
Room 303, 3/F, Run Run Shaw Building
The University of Hong Kong
Pokfulam Road, Hong Kong
Telephone : (852) 3917-8312
Facsimile : (852) 2858-9041
E-mail : saas@hku.hk
Website : <http://www.saasweb.hku.hk/conference/saas50/>

HKU-NUS-STANFORD Conference in Statistical Science and Decision Analytics

In Celebration of the 50th Anniversary of the HKU
Department of Statistics and Actuarial Science

Program Schedule

March 10, 2017 (Friday)

08:30 – 09:00	Registration
09:00 – 09:10	Welcoming Address Professor Andy HOR, Pro-Vice-Chancellor (Research), HKU
09:10 – 09:15	Opening Address Professor Wai Keung LI, Head of Department of Statistics & Actuarial Science, HKU
09:15 – 09:20	Group Photo Taking
Session I	Chair: Wai Keung LI
09:20 – 10:00	Tze Leung LAI (Stanford University) <i>Valid P-values in Post-selection Multiple Testing: A 50-year Journey</i>
10:00 – 10:40	Hock Peng CHAN (National University of Singapore) <i>An Efficient Non-parametric Solution to the Multi-Armed Bandit Problem</i>
10:40 – 11:00	Coffee Break
Session II	Chair: Hock Peng CHAN
11:00 – 11:40	Wai Keung LI (The University of Hong Kong) <i>On Buffered Time Series Models</i>
11:40 – 12:20	Guodong LI (The University of Hong Kong) <i>Hybrid Conditional Quantile Inference for Conditional Heteroscedastic Time Series Models</i>
12:30 – 14:00	Lunch (by invitation only)
Session III	Chair: Stephen M.S. LEE
14:00 – 14:40	Yik-Ying TEO (National University of Singapore) <i>Detecting and Characterizing Signatures of Positive Natural Selection in the Human Genome</i>

14:40 – 15:20	Philip Leung Ho YU (The University of Hong Kong) <i>Big Data Decision Analytics</i>
15:20 – 16:00	Anthony Yung Cheung KUK (National University of Singapore) <i>Function Compositional Adjustments of Conditional Quantile Curves</i>
16:00 – 16:20	Coffee Break
Session IV	Chair: Yik Ying TEO
16:20 – 17:00	Stephen Man Sing LEE (The University of Hong Kong) <i>Consistency of a Hybrid Block Bootstrap Procedure for Sample Quantiles under Weak Dependence</i>
17:00 – 17:40	Kin LAM (The University of Hong Kong) <i>From Interday Trading, Intraday Trading to High Frequency Trading</i>
	Dinner (by invitation only)

March 11, 2017 (AM) (Saturday)

08:45 – 09:00	Registration
Session V	Chair: Anthony Yung Cheung KUK
09:00 – 09:40	Jeff Jianfeng YAO (The University of Hong Kong) <i>On a Spiked Model for Large Volatility Matrix Estimation from Noisy High-frequency Data</i>
09:40 – 10:20	Yingcun XIA (National University of Singapore) <i>Kernel Estimation of Mutual Information</i>
10:20 – 10:40	Coffee Break
Session VI	Chair: Jeff Jianfeng YAO
10:40 – 11:20	Eva TSUI (Hospital Authority, HKSARG) <i>Leveraging Hospital Authority's Service Demand Management on Decision Analytics with Clinical Data Repository</i>
11:20 – 12:00	Matthew Siu Fung YIU (Hong Kong Monetary Authority) <i>Effects of Capital Flow on the Equity and Housing Markets in Hong Kong</i>
12:00 – 12:40	Samuel Po Shing WONG (The University of Hong Kong) <i>Trading, Fast and Slow</i>
	Lunch (by invitation only)

ABSTRACTS

Hock Peng CHAN

Department of Statistics and Applied Probability, National University of Singapore

An Efficient Non-parametric Solution to the Multi-Armed Bandit Problem

Lai and Robbins (1985) and Lai (1987) provided an efficient parametric solution to the multi-armed bandit problem, showing that arm allocation via upper confidence bounds (UCB) achieves optimal regret. The UCB are constructed from the Kullback-Leibler information of the reward distributions, estimated from within a specified parametric family.

In recent years there has been renewed interest in the multi-armed bandit problem due to new applications in machine learning algorithms and data analytics. Many non-parametric arm allocation procedures like epsilon-greedy and Boltzmann exploration were carefully studied, and the UCB arm allocation procedure that is designed for normal rewards distributions was also analyzed in a non-parametric setting. However unlike UCB these arm allocation procedures do not achieve optimal regret.

In this talk, we propose a subsample comparison arm allocation procedure that is non-parametric in nature, and achieves optimal regret under an exponential family setting.

Anthony Y.C. KUK

Department of Statistics and Applied Probability, National University of Singapore

Function Compositional Adjustments of Conditional Quantile Curves

To adjust nonparametrically the quantile function estimate obtained from a parametric regression, we estimate the quantile function of the probability integral transformed data and compose the two functions. Compared with the customary additive adjustment approach, the proposed approach pays more attention to the fact that we are estimating a quantile function and as a result, it possesses some desirable properties such as monotonicity and invariance to increasing transformation. One round of bandwidth selection suffices as compositional adjustments at all quantile levels can be obtained by smoothing the same set of probability integral transformed data. Another advantage of the proposed method is that it yields a diagnostic plot useful in assessing the goodness of fit of the assumed model. Taylor series approximation and results from two simulation studies suggest that the compositionally adjusted estimator is robust to model misspecification, and can be more efficient than direct nonparametric estimation. We illustrate the proposed method using two examples from water resources and human biomonitoring studies.

Tze Leung LAI

Department of Statistics, Stanford University

Valid P-values in Post-selection Multiple Testing: A 50-year Journey

Test-based variable selection, as exemplified by partial F-tests in forward stepwise or backward elimination procedures in linear regression and their extensions to generalized linear and Cox regression models, preceded information-criterion selection and prediction-based selection procedures, and is still popular in software packages. After a brief review of these variable selection methodologies, highlighting their differences and similarities, we focus on applications in which test-based variable selection fits nicely into the goal of the study but falls short of giving a valid overall test. By a novel application of the closed testing principle and resampling methods, we show how a valid test can be developed. We then extend the methodology to develop valid post-selection multiple testing procedures and thereby address the reproducibility crisis of scientific findings in the big data era. Applications to fault diagnosis in multistage manufacturing processes, which motivated this work, are also discussed. This recent development is then traced back to my personal journey in this subject, which dated back 50 years ago when I graduated from HKU and the Statistics Department was established, and which is also interwoven into the Department's phenomenal growth in the past 50 years.

Kin LAM

Department of Statistics and Actuarial Science, The University of Hong Kong

From Interday Trading, Intraday Trading to High Frequency Trading

The speaker will present his views on quantitative trading strategies based on time series of daily closing prices, on time series of intraday data like minute by minute prices, or on high frequency tick-by-tick prices. Special attention will be paid to the filter trading rule together with its mathematical properties which distinguish it from other trading rules familiar to market participants.

Stephen M.S. LEE

Department of Statistics and Actuarial Science, The University of Hong Kong

Consistency of a Hybrid Block Bootstrap Procedure for Sample Quantiles under Weak Dependence

Consistency and optimality of moving block bootstrap procedures for distribution estimation for dependent data have been well studied under a smooth function model. Much less is known, however, for nonsmooth functionals such as quantiles. Existing results focus primarily on the standard moving block bootstrap (MBB), which requires each bootstrap series be of

approximately the same length as the observed data series. Consistency may also be achieved by subsampling, which can be viewed as an extreme case of MBB with each bootstrap series formed by only one single block. We show that, in the context of distribution estimation for sample quantiles, weak consistency indeed holds for any “hybrid” block bootstrap scheme lying between the two extremes corresponding to the standard MBB and subsampling. Our empirical results examine the performance of such hybrid block bootstrap distribution estimators, and illustrate the extent of benefit which can be gained by suitably adjusting the number of blocks contained in each bootstrap series.

Guodong LI

Department of Statistics and Actuarial Science, The University of Hong Kong

Hybrid Conditional Quantile Inference for Conditional Heteroscedastic Time Series Models

In financial time series analysis, it is an essential task to estimate the conditional quantiles for conditional heteroscedastic models, among which, the generalized autoregressive conditional heteroscedastic (GARCH) model has the greatest popularity. So far, feasible quantile regression methods in this area have been confined to the linear GARCH model owing to its tractable conditional quantile structure. We show that, surprisingly, the GARCH process admits a linear GARCH representation after a simple albeit nontrivial transformation. This result allows us to first estimate the conditional variances by the Gaussian quasi-maximum likelihood estimation for the GARCH model and then fit the conditional quantiles based on the corresponding linear GARCH form. An easy-to-implement hybrid conditional quantile estimation procedure is thereby developed for the conditional heteroscedastic model. To approximate the asymptotic distribution of the proposed hybrid estimator, we introduce a new bootstrapping procedure, where a time-consuming optimization is replaced by a direct averaging. Moreover, a goodness-of-fit test based on the residual quantile autocorrelation function is also constructed to check the adequacy of the fitted conditional quantiles. Simulation experiments are carried out to assess the finite-sample performance of the proposed inference tools, and real examples are discussed to further illustrate the usefulness of our approach.

Wai Keung LI

Department of Statistics and Actuarial Science, The University of Hong Kong

On Buffered Time Series Models

The progress as of today of a new class of threshold time series models known as buffered processes is reviewed. In this new class of models switching back and forth between two regimes depends on two different thresholds. We first investigate the self-excited buffered autoregressive (BAR) process to some extent including an identification procedure and the asymptotic properties of the least squares estimators. We then extend the class of models to cover conditional heteroscedasticity resulting in the buffered GARCH and buffered AR-

GARCH models. Simulation studies and applications to real data are considered to illustrate the potential of this new type of threshold models.

Yik Ying TEO

Professor, Department of Statistics and Applied Probability, National University of Singapore

Vice Dean (Research), Head of Biostatistics Domain, Leader of Public Health Genomics Programme, Saw Swee Hock School of Public Health

Integrative Omics Programme Leader, Life Sciences Institute

Associate Faculty Group Leader, Genome Institute of Singapore

Detecting and Characterizing Signatures of Positive Natural Selection in the Human Genome

Natural selection is a significant force that shapes the architecture of the human genome and introduces diversity across global populations. This process of Darwinian selection leaves a footprint in the genetic code of modern humans, which can be statistically identified and quantified. The question of whether advantageous mutations have arisen in the human genome as the result of single or multiple mutation events remains unanswered except for a handful of genes such as those that confer lactase persistence, affect skin pigmentation or cause sickle cell anemia. Here, I will describe a methodology development where I utilized a long-range haplotype method for identifying genomic signatures of positive selection, which is found to be considerably more sensitive and specific than existing methods such as iHS or XP-EHH. Our method also locates the founder haplotypes that are carrying the advantageous variants and infers their corresponding population frequencies. This presents an unprecedented opportunity to systematically interrogate the whole human genome whether a selection signal that is shared across different populations is the consequence of a single mutation process followed subsequently by gene flow between populations, or convergent evolution due to the occurrence of multiple independent mutation events either at the same variant or within the same gene. I will then demonstrate this method by applying it to fourteen populations across the entire globe, and separately to another collection of 30 populations from Asia.

Eva TSUI

Statistics & Workforce Planning Department, Hospital Authority, HKSARG

Leveraging Hospital Authority's Service Demand Management on Decision Analytics with Clinical Data Repository

It is a sharing of a few pieces of application within the Hospital Authority (HA), illustrating how to transform the huge volume and wide variety of data in HA's IT systems into information with the use of analytics techniques and tools to facilitate decision making and formulation of strategies to manage growing services demand. In the face of an ageing population in Hong

Kong, HA is facing a challenge regarding how to better manage the growing service demand. Under the HA's Strategic Service Framework for Elderly People, appropriate level of care should be respectively provided to persons of the three different risk strata under the conceptual pyramid of healthcare needs of the elderly population. Like a corporate tool in support of the formulation, implementation and evaluation of strategies, measures and action plans within this framework, decision analytics are applied to leverage service demand management on (a) emergency care services, (b) chronic disease care services and (3) episodic illness services. Illustrative examples include the implementation of statistical models to predict the risk of elderly emergency medical admissions within a month following an index hospital visit and the surge in emergency admissions during winter; the development of chronic diseases trajectories to enhance services and facilities planning through profiling, association and projection; as well as the development of quantitative measures on access to general outpatient episodic illness service and the application of spatial analysis with the use of Geographic Information System (GIS) to facilitate planning to cope with services demand across geographical areas in Hong Kong. All in all, given the accumulated and rapidly growing data inside and outside HA, decision analytics has been playing a pivotal role to transform data into information and then into knowledge, hence facilitating a better decision making within HA, which ultimately aims for a better system to meet the population healthcare needs in Hong Kong.

Samuel Po Shing WONG

Department of Statistics and Actuarial Science, The University of Hong Kong

Trading, Fast and Slow

Said the Nobel prize-winning economist Daniel Kahneman in his best-seller *Thinking, Fast and Slow*: "Overconfidence is fed by the illusory certainty of hindsight". This summarizes how human decision-making systems can be corrupted by over-generalizing data analysis results. Such delusion is commonly encountered in trading algorithm developments and can lead to highly unfavorable consequences. Two typical examples are (A) confusing the statistical significance with the predictive power of time series models, and (B) back-testing trading algorithms without considering market microstructure. More detailed discussions in this direction will be delivered in this talk. Some suggestions on how to derive high and low frequency trading strategies will also be presented under the framework of reinforcement learning.

Yingcun XIA

Department of Statistics and Applied Probability, National University of Singapore

Kernel Estimation of Mutual Information

Quantifying the strength of dependence between two random variables is a fundamental issue in data analysis. The mutual information is arguably the most popular measure of the dependence. "Unfortunately, reliably estimating mutual information from finite continuous data

remains a significant and unresolved problem” (Kinney and Atwal, PNAS, 2014). In this talk, we propose a new estimation of MI. The estimator has a unique global maximum values with respect to the bandwidth in the estimation, and that the maximum value is the most efficient under some conditions. Simulation study also suggests that numerical performance of the estimation is more stable than the existing methods.

Jeff J.F. YAO

Department of Statistics and Actuarial Science, The University of Hong Kong

On a Spiked Model for Large Volatility Matrix Estimation from Noisy High-frequency Data

Recently, inference about high-dimensional integrated covariance matrices (ICVs) based on noisy high-frequency data has emerged as a challenging problem. In the literature, a pre-averaging estimator (PA-RCov) is proposed to deal with the microstructure noise. Using the large-dimensional random matrix theory, it has been established that the eigenvalue distribution of the PA-RCov matrix is intimately linked to that of the ICV through the Marcenko-Pastur equation. Consequently, the spectrum of the ICV can be inferred from that of the PA-RCov. However, extensive data analyses demonstrate that the spectrum of the PA-RCov is spiked, that is, a few large eigenvalues (spikes) stay away from the others which form a rather continuous distribution with a density function (bulk). Therefore, any inference on the ICVs must take into account this spiked structure. As a methodological contribution, we propose a spiked model for the ICVs where spikes can be inferred from those of the available PA-RCov matrices. The consistency of the inference procedure is established and is checked by extensive simulation studies. In addition, we apply our method to the real data from the US and Hong Kong markets. It is found that our model clearly outperforms the existing one in predicting the existence of spikes and in mimicking the empirical PA-RCov matrices.

This is a joint work with Keren Shen and Wai Keung Li.

Yin-Wong CHEUNG¹, Kenneth K. CHOW², Matthew S. YIU³

¹City University of Hong Kong

²Hong Kong Monetary Authority

³Hong Kong Institute for Monetary Research

Effects of Capital Flow on the Equity and Housing Markets in Hong Kong*

The revival of strong capital flows to emerging economies in the aftermath of the 2008-9 Global Financial Crisis has rekindled the debate on the adverse effects of excessive capital inflows. In this study, we study effects of official and illicit capital flows on Hong Kong, which is a small and open economy with minimal restrictions on cross-border fund movements. To illustrate the impacts of different types of capital flows, we study the effects of official and illicit flows on Hong Kong's equity and residential housing markets. It is found that the official and illicit

capital flow measures reflect different facets of flow movements and exhibit differential effects on the equity and residential housing markets. The results highlight the complexity of managing capital flows, and the relevance of policies targeting specific sectors.

* The views expressed in this paper are those of the authors, and do not necessarily reflect those of the Hong Kong Monetary Authority, Hong Kong Institute for Monetary Research, its Council of Advisers, or the Board of Directors.

Philip L.H. YU

Department of Statistics and Actuarial Science, The University of Hong Kong

Big Data Decision Analytics

The world is now experiencing a proliferation of big data whose volume, variety and velocity continue to reach unprecedented levels. Statistics are critical to making good use of big data for decision making. But how exactly? In this talk, I will briefly introduce recent research projects which demonstrate the essential contribution of statistics and highlight several challenges in big data research and applications.

HKU FOOD MAP

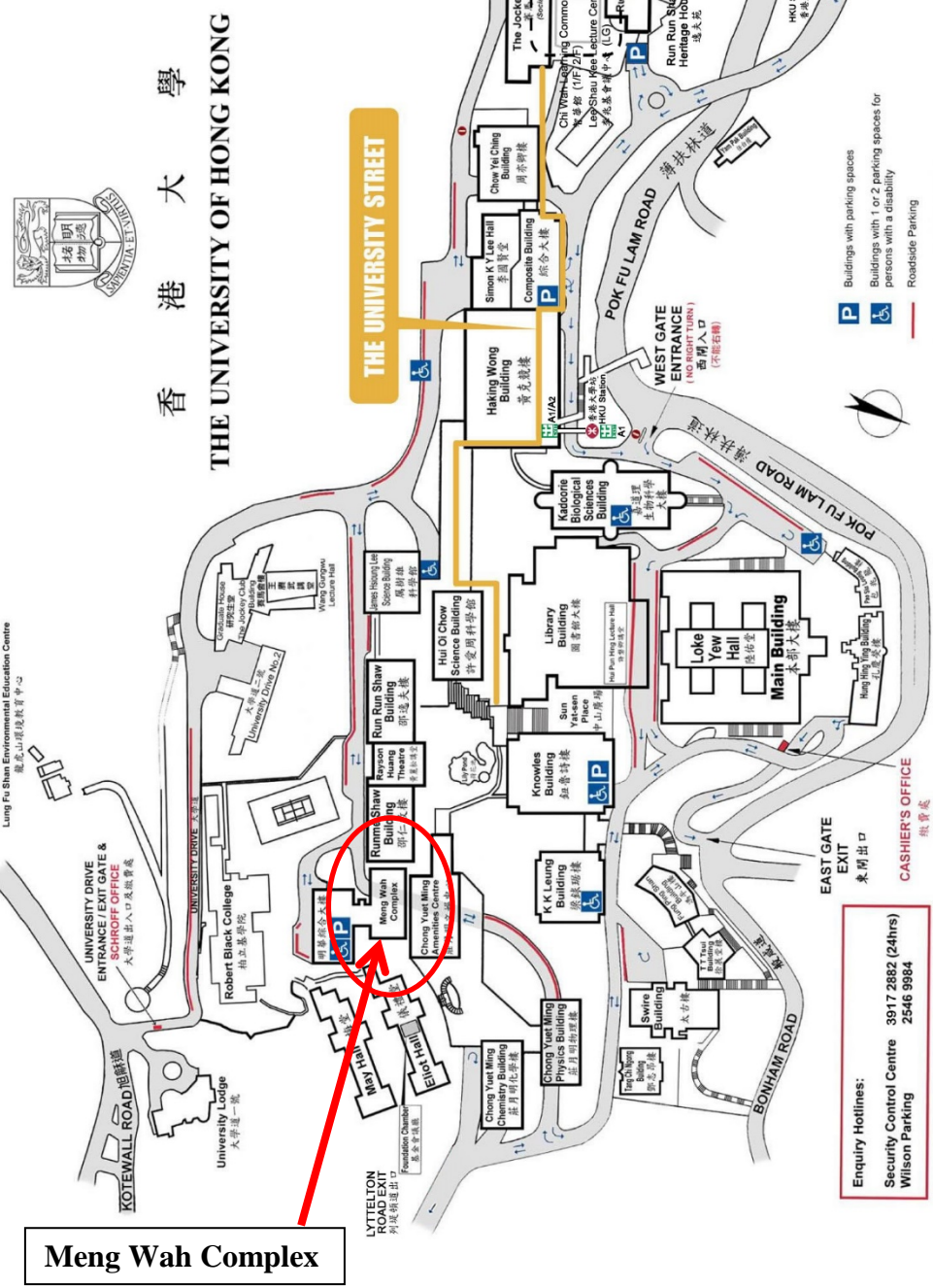


Last update: Jan 2016

MAPS OF HKU CAMPUS

ATTENTION:
 Vehicles transiting through campus will not be charged for parking, provided that:
 1) They leave the campus within a 30-minute period.
 2) Vehicles entering at the West Gate must leave by the East Gate.
 3) Vehicles entering at the University Drive Gate must leave by the same gate.
 Vehicles contravening these regulations will pay a parking charge.

注意:
 時時車輛須利用同一門口進出方可享有30分鐘免費時限：
 1) 由港林道西門進入校園，必須經由該西門離開校園；同
 樣地，由大學道進入校園，必須經由該西門離開；否則，
 一律不會獲得免費的停車及須繳付泊車費。



Meng Wah Complex