PROGRAM

Tuesday, 26 June 2018

① LT27	9:00 AM – 9:30 AM	Opening Ceremony
	9:30 AM – 10:30 AM	Plenary Lecture (PL01)
	Tea Break — Foyer of ①	
	11:10 AM – 12:10 PM	Plenary Lecture (PL02)
	Lunch — Foyer of ①	

② LT28	③ LT29	④ LT26	⑤ LT33	⑥ LT34	⑦ LT31	⑧ MD9-0102 GH	⑨ S16-06118	⑩ S17-0405

Tuesday, 26 June 2018

1:30 PM – 3:10 PM	DL16	IP01	IP13	IP51	IP08	IP53	TCP03	CP01	
Tea Break — Foyers of ①, ⑥, ⑦									
3:30 PM – 5:10 PM	DL06	IP04	IP05	IP42	IP17	IP23	TCP05	CP07	

Wednesday, 27 June 2018

8:30 AM – 10:10 AM	DL01	DL14	IP21	IP35	IP46	IP09	IP41	CP02	
Tea Break — Foyers of ①, ⑥, ⑦									
10:30 AM – 12:10 PM	DL11	IP03	IP29	IP30	IP40	IP55	IP14	CP04	
Lunch — Foyer of ①									
1:30 PM – 3:10 PM	DL07	DL15	IP22	IP39	IP47	IP26	IP16	TCP04	
Tea Break — Foyers of ①, ⑥, ⑦									
3:30 PM – 5:10 PM	DL13	IP18		IP33	IP44	IP59	IP61	CP05	
Buses begin departing NUS Kent Ridge Campus at 5:30 PM for Reception (7:00 PM) and Banquet Dinner (7:30 PM – 10:30 PM) at Carlton Hotel									

Thursday, 28 June 2018

8:30 AM – 10:10 AM	DL03	DL05	IP06	IP15	IP25	IP43	IP56	CP10	TCP09
Tea Break — Foyers of ①, ⑥, ⑦									
10:30 AM – 12:10 PM	DL04	DL08	IP57	IP10	IP20	IP32	IP12	IP49	TCP10
Lunch — Foyer of ①									
Excursion: "Singapore City Tour" (1:00 PM – 6:00 PM) [optional, tickets required]									

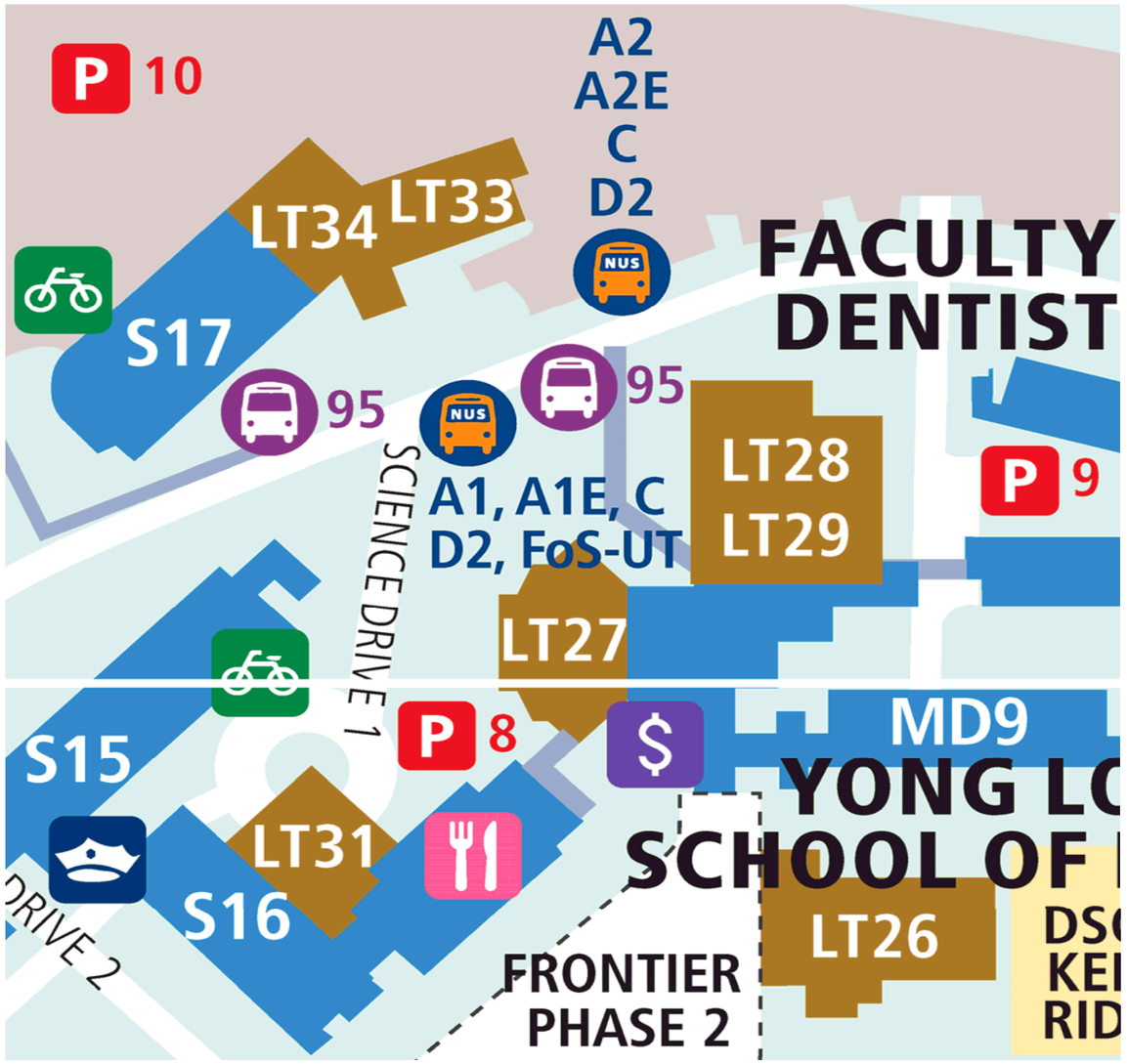
Friday, 29 June 2018

8:30 AM – 10:10 AM	DL12	DL10	IP11	IP07	IP34	IP48	TCP06	CP06	
Tea Break — Foyers of ①, ⑥, ⑦									
10:30 AM – 12:10 PM	IP45	IP28	IP37	IP50	IP58	TCP07	TCP08	CP08	
Lunch — Foyer of ①									
1:30 PM – 3:10 PM	DL02	IP02	IP31	IP36	IP52	IP60		CP03	
Tea Break — Foyers of ①, ⑥, ⑦									
3:30 PM – 5:10 PM	DL09	IP24	IP27	IP38	IP54	TCP02		CP09	

LOCALITY MAP

Getting to NUS (Kent Ridge Campus):
Ride the MRT Circle Line to Kent Ridge (CC24) Station, take Exit A, go to Bus Stop 3, transfer to the NUS

① LT27	Opening ceremony & plenary lectures; registration, tea break & lunch (foyer)
② LT28	Parallel sessions; same foyer as LT27
③ LT29	Parallel sessions; same foyer as LT27
④ LT26	Parallel sessions; up a ramp and then a short flight of stairs from LT27, on the left
⑤ LT33	Parallel sessions; cross the road from LT28 to building S17, level 2
⑥ LT34	Parallel sessions; cross the road from LT28 to building S17, level 3; tea break (foyer)
⑦ LT31	Parallel sessions; up a ramp from LT27, turn right to building S16, lvl 3; t/break (foyer)
⑧ MD9-0102GH	Parallel sessions; up a ramp from LT27, turn left to building MD9, level 1
⑨ S16-06118	Parallel sessions; up a ramp from LT27, turn right to building S16, level 6
⑩ S17-0405	Parallel sessions; cross the road from LT28 to building S17, level 4



PLENARY LECTURES

Session ID	Full Name Affiliation	Title Abstract
DL01:	**Tuesday, 26 June 2018, 9:30 AM to 10:30 AM, ① LT27**	
Chair	Qi-Man Shao, The Chinese University of Hong Kong	
Plenary Speaker	Louis H. Y. Chen National University of Singapore	"FROM STEIN'S LEMMA TO STEIN'S METHOD AND BEYOND" Stein's Stein's lemma gives a simple characterization of the normal distribution. In this talk, I will explain how Stein discovered this lemma and how it leads to Stein's method. I will also discuss the generator approach, the connection between Stein's method and Palm theory in random measures, the connection between Stein's method and Malliavin calculus, and applications to stochastic geometry.
DL02:	**Tuesday, 26 June 2018, 11:10 AM to 12:10 PM, ① LT27**	
Chair	Xuming He, University of Michigan	
Plenary Speaker	Bin Yu University of California, Berkeley	"THREE PRINCIPLES OF DATA SCIENCE: PREDICTABILITY, STABILITY, AND COMPUTABILITY" In this talk, I'd like to discuss the intertwining importance and connections of three principles of data science in the title. They will be demonstrated in the context of two collaborative projects in neuroscience and genomics, respectively. The first project in neuroscience uses transfer learning to integrate fitted convolutional neural networks (CNNs) on ImageNet with regression methods to provide predictive and stable characterizations of neurons from the challenging primary visual cortex V4. The second project proposes iterative random forests (iRF) as a stablized RF to seek predictable and interpretable high-order interactions between biomolecules.

Last updated
6/25/2018

DISTINGUISHED LECTURE SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
DL01:	**Controlled Diffusion Processes, Wednesday, 27 June 2018, 8:30 AM to 10:10 AM, ② LT28**	
DL / Chair	Vivek S. Borkar (Distinguished Lecturer), Indian Institute of Technology / Vladimir Gaitsgory, Macquarie University	
8:30 AM to 9:20 AM:	Vivek S. Borkar Indian Institute of Technology	"CONTROLLED DIFFUSION PROCESSES" This talk will be an overview of some developments in controlled diffusion processes, with emphasis on the probabilistic (as opposed to p.d.e.) aspects. In particular it will spell out the issues in existence theory and characterisation of optimal policies for the harder problems such as the degenerate case, the ergodic and risk-sensitive costs, partially observed control, etc., and touch upon multiple time scale phenomena.
9:20 AM to 9:45 AM:	Anup Biswas Indian Institute of Science Education and Research	"RISK SENSITIVE ERGODIC CONTROLS" Gaussian assumption but retains the simplicity of the Gaussian dependence structure, which is particularly attractive for large data.
9:45 AM to 10:10 AM:	Vladimir Gaitsgory Macquarie University	"ON AVERAGING OF SINGULARLY PERTURBED CONTROL SYSTEMS" Models of real life dynamical systems are often characterized by the fact that their state variables are decomposed into groups of slow and fast ones, this decomposition being formalized with a special introduction of a small parameter. Such systems are commonly called singularly perturbed (SP). In this talk, we will discuss the use of averaging techniques in dealing with SP control systems, our emphasis will be on results obtained in collaboration with Prof V. Borkar.
DL02:	**Methods for Gene-Environment Interactions and Physical Activity in Retrospective Studies, Friday, 29 June 2018, 1:30 PM to 3:10 PM, ② LT28**	
DL / Chair	Raymond J. Carroll (Distinguished Lecturer), Texas A&M University / Yanyuan Ma, Pennsylvania State University	
1:30 PM to 2:20 PM:	Raymond J. Carroll Texas A&M University	"SEMIPARAMETRIC ANALYSIS OF COMPLEX POLYGENIC GENE-ENVIRONMENT INTERACTIONS IN CASE-CONTROL STUDIES" Many methods have been proposed recently for efficient analysis of case-control studies of gene-environment interactions using a retrospective likelihood framework that exploits the natural assumption of gene-environment independence in the underlying population. We will review some of this literature and discuss some of the fairly astonishing gains in efficiency that are possible for understanding the interactions. However, for polygenic modeling of gene-environment interactions, a topic of increasing scientific interest, applications of retrospective methods have been limited due to a requirement in the literature for parametric modeling of the distribution of the genetic factors, which is difficult because of the complex nature of polygenic data. We propose a fully general, computationally simple, efficient semiparametric method for analysis of case-control studies that allows exploitation of the assumption of gene-environment independence without any further parametric modeling assumptions about the marginal distributions of any of the two sets of factors. The method relies on the key observation that an underlying efficient profile likelihood depends on the distribution of genetic factors only through certain expectation terms that can be evaluated empirically. We develop asymptotic inferential theory for the estimator and evaluate numerical performance using simulation studies. An application of the method is illustrated using a case-control study of breast cancer.
2:20 PM to 2:45 PM:	Chongzhi Di Fred Hutchinson Cancer Research Center	"TESTING GENE-ENVIRONMENT INTERACTIONS IN THE PRESENCE OF MEASUREMENT ERROR" Complex diseases result from an interplay between genetic and environmental risk factors, and it is of great interest to study the gene-environment interaction (GxE) to understand the etiology of complex diseases. However, one difficulty with GxE arises from the fact that environmental exposures are often measured with error. We focus on testing GxE when the environmental exposure is subject to measurement error. In contrast to the well-established results that the naive test ignoring measurement error is valid in testing the main effects, we find that the naive test for GxE leads to inflated type I error under the null hypothesis of no interaction. The analytic form of the bias term for general linear models is obtained, which is shown to be closely related to regression calibration. We then propose an approach to correct measurement error for testing GxE\nwhen either validation data or replicates are available. Extensive simulation studies are conducted to illustrate the performance of various tests in finite samples. The proposed methods are applied to the Women's Health Initiative.
2:45 PM to 3:10 PM:	Yanyuan Ma Pennsylvania State University	"A SUPERPOPULATION TREATMENT TO CASE-CONTROL DATA ANALYSIS" We study the regression relationship among covariates in case-control data, an area known as the secondary analysis of case-control studies. The context is such that only the form of the regression mean is specified, so that we allow an arbitrary regression error distribution, which can depend on the covariates and thus can be heteroscedastic. Under mild regularity conditions we establish the theoretical identifiability of such models. Previous work in this context has either (a) specified a fully parametric distribution for the regression errors, (b) specified a homoscedastic distribution for the regression errors, (c) has specified the rate of disease in the population (we refer this as true population), or (d) has made a rare disease approximation. We construct a class of semiparametric estimation procedures that rely on none of these. The estimators differ from the usual semiparametric ones in that they draw conclusions about the true population, while technically operating in a hypothetic superpopulation. We also construct estimators with a unique feature, in that they are robust against the misspecification of the regression error distribution in terms of variance structure, while all other nonparametric effects are estimated despite of the biased samples. We establish the asymptotic properties of the estimators and illustrate their finite sample performance through simulation studies.
DL03:	**Stochastic Analysis for Non-Local Operators, Thursday, 28 June 2018, 8:30 AM to 10:10 AM, ② LT28**	
DL / Chair	Zhen-Qing Chen (Distinguished Lecturer), University of Washington / Jian Wang, Fujian Normal University	

DISTINGUISHED LECTURE SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:30 AM to 9:20 AM:	Zhen-Qing Chen University of Washington	"TIME FRACTIONAL EQUATIONS AND PROBABILISTIC REPRESENTATION" Time-fractional diffusion equations have been actively studied in several fields including mathematics, physics, chemistry, engineering, hydrology and even finance and social sciences as they can be used to model the anomalous diffusions exhibiting subdiffusive behavior, due to particle sticking and trapping phenomena. In this talk, I will report some recent progress in the study of general fractional-time parabolic equations of mixture type, including existence and uniqueness of the solutions and their probabilistic representations in terms of the corresponding inverse subordinators with or without drifts. Sharp two-sided estimates on the fundamental solution will be given. Fractional-time parabolic equations with source term will also be discussed. In particular, a new representation formula for the solution of time fractional Poisson equation will be presented, which does not involve fractional time derivative of the fundamental solution.
9:20 AM to 9:45 AM:	Xin Chen Shanghai Jiao Tong University	"DIRICHLET FRACTIONAL HEAT KERNEL ESTIMATES ON A HORN-SHAPED DOMAIN" In this talk, we will give a full time Dirichlet heat kernel estimates for α-stable process on a horn-shaped domain. In particular, we will obtain different estimates according to different reference functions of the domain. This talk is based on a joint work with Panki Kim and Jian Wang.
9:45 AM to 10:10 AM:	Jian Wang Fujian Normal University	"GRADIENT ESTIMATES FOR STOCHASTIC HAMILTONIAN SYSTEMS WITH LEVY NOISES " Consider general stochastic Hamiltonian system with locally Lipschitz drift terms and Levy nosies having a rotationally invariant component. We establish gradient estimates for the assciated semigroups by using the coupling by reflection for Levy processes.
DL04:	**Recent Advances in the Analysis of High-Dimensional Dependent Data, Thursday, 28 June 2018, 10:30 AM to 12:10 PM, ② LT28**	
DL / Chair	Ching-Kang Ing (Distinguished Lecturer), National Tsing Hua University / Chun Yip Yau, Chinese University of Hong Kong	
10:30 AM to 11:20 AM:	Ching-Kang Ing National Tsing Hua University	"VARIABLE SELECTION USING ORTHOGONAL GREEDY ALGORITHMS IN SPARSE HIGH-DIMENSIONAL TIME SERIES MODELS" We investigate prediction capability of the orthogonal greedy algorithm (OGA) in high-dimensional time series models. We derive the prediction error rate of OGA in terms of the number of iterations and the number of candidate variables, under various sparsity conditions. We further introduce a high-dimensional model selection method, high-dimensional Akaike's information criterion (HDAIC), to determine the number of OGA iterations. We show that when used together with HDAIC, OGA can achieve the desired error rate, which is sometimes better than the minimax-optimal rate, e.g., Raskutti et al. (2011) and Negahban et al. (2012), in the special case where observations are independent over time.
11:20 AM to 11:45 AM:	Shinpei Imori Hiroshima University	"CONVERGENCE RATE OF OGA UNDER COVARIATE SHIFT" In applied statistics and machine learning, prediction in regression analysis is one of the most important problem. The present paper attempts to construct an accurate predictor in a high-dimensional linear regression model under covariate shift (Shimodaira, 2000, Journal of Statistical Planning and Inference). The covariate shift situation assumes that distributions of covariate between training and test data are different, while conditional distributions of response given covariate are common. Our attempt is based on orthogonal greedy algorithm (OGA), of which a convergence rate of conditional prediction error (CPE) has been already derived by Ing and Lai (2011, Statistica Sinica) when the training and test distributions are the same. In this paper, we consider using a weighted version of OGA and derive its convergence rate of CPE under the covariate shift.
11:45 AM to 12:10 PM:	Chun Yip Yau Chinese University of Hong Kong	"GROUP ORTHOGONAL GREEDY ALGORITHM FOR CHANGE-POINT ESTIMATION OF MULTIVARIATE TIME SERIES" This paper proposes a three-step method for the detection of multiple structural breaks in piecewise stationary vector autoregressive processes. The number of the structural breaks can be large and unknown. Moreover, the number and the location of the breaks are not necessarily the same in different components. The proposed method is based on a connection between structural break problem and high dimensional regression problem. With such connection, we develop a group orthogonal greedy algorithm, originally from high dimensional variable selection context, for efficient estimation of structural breaks. A high-dimensional information criterion is proposed to detect different structural breaks in different components. We prove the consistency of the estimators and provide Monte Carlo experiments for the finite sample performance.
DL05:	**Some Developments in High-Dimensional and High-Frequency Data Analysis, Thursday, 28 June 2018, 8:30 AM to 10:10 AM, ③ LT29**	
DL / Chair	Bing-Yi Jing (Distinguished Lecturer), Hong Kong University of Science and Technology / Zhi Liu, University of Macau	
8:30 AM to 9:20 AM:	Bing-Yi Jing Hong Kong University of Science and Technology	"LEARNING NETWORK DATA" Network data has been well studied over the years and yet there are still many challenging interesting problems. First, we will investigate the problem of community detection for sparse networks, whereas its counterparts for dense network have been intensively studied. We propose a new similarity matrix, SLIM, which have been shown to work very well in both simulations and real examples. Secondly, we will investigate semi-supervised learning with partially labeled network data. We propose a new method, referred to as global information integration. The method is shown to be consistent and work particularly well for unbalanced and/or heterogeneous networks.
9:20 AM to 9:45 AM:	Zhi Liu University of Macau	"DATA-EFFICIENT REALIZED COVARIANCE MATRIX" The variate features of high frequency data, such as, market microstructure effect, asynchronous trading, multiple records, etc, bring challenges in the estimation of covariance matrices among assets. In this paper, we proposed a data-efficient estimator which utilizes all of data, whereas all existing approaches have to discard part of data, to deal with the asynchronicity and multiple records. By congregating the data points within the synchronized time intervals as average, we found that the realized covariance is still consistent. We have established the related central limit theorem and its studentized version. Compared to the existing approaches, our estimator achieves much improvement in the estimation efficient, in particular for those highly traded liquid assets. Through a variety of synthetic data experiments, we assess the finite sample performance of proposed estimator and make comparison with other existing methods. Finally we illustrate the estimator via some real data analyses.
9:45 AM to 10:10 AM:		#N/A
DL06:	**Statistical Prediction with Big Data: From Digital Disease Forecasting to Protein Folding, Tuesday, 26 June 2018, 3:30 PM to 5:10 PM, ② LT28**	

DISTINGUISHED LECTURE SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
DL / Chair	S.C. Samuel Kou (Distinguished Lecturer), Harvard University / Shihao Yang, Harvard University	
3:30 PM to 4:20 PM:	S.C. Samuel Kou Harvard University	"BIG DATA, GOOGLE AND DISEASE DETECTION: THE STATISTICAL STORY" Big data collected from the internet have generated significant interest in not only the academic community but also industry and government agencies. They bring great potential in tracking and predicting massive social activities. We focus on tracking disease epidemics in this talk. We will discuss the applications, in particular, Google Flu Trends, some of the fallacy and the statistical implications. We will propose a new model that utilizes publicly available online data to estimate disease epidemics. Our model outperforms all previous real-time tracking models for influenza epidemics at the national level of the US. An extended version of the model gives accurate tracking of Dengue fever in Asian and South American countries/states. We will also draw some lessons for big data applications.
4:20 PM to 4:45 PM:	Samuel Wong University of Florida	"IMPROVING PROTEIN FOLDING PREDICTIONS BY DATA-DRIVEN MONTE CARLO SAMPLING AND SCORING METHODS" The problem of predicting the 3-D structure of a protein from its amino acid sequence using computer algorithms has challenged scientists for nearly a half century. The structure of a protein is essential for understanding its function, and hence accurate structure prediction is of vital importance in modern applications such as protein design in biomedicine. A powerful approach for structure prediction is to search for the conformation of the protein that has minimum potential energy. The success of this approach hinges on having an efficient search method and an accurate scoring function -- both of which are challenging to develop in practice due to the size of the conformational space. In this talk we overview our sequential Monte Carlo method for searching for low energy conformations and describe its application to learning improved scoring functions using the Protein Data Bank.
4:45 PM to 5:10 PM:	Shihao Yang Harvard University	"BIG DATA, GOOGLE, AND INFECTIOUS DISEASE PREDICTION: A STATISTICAL PERSPECTIVE" Big data generated from the internet have great potential in tracking and predicting massive social activities, in particular infectious diseases, whose accurate real-time prediction could help public health officials make timely decisions to save lives. We introduce and extend model ARGO (AutoRegression with General Online data) that has successfully utilized publicly available Google search data, combined with cloud-based electronic health records, to estimate current and near-future influenza-like illness activity level and/or dengue fever activity level for United States and other countries around the globe. Our regularized multivariate regression model dynamically selects the most appropriate variables for prediction every week, and significantly outperforms all previous internet-based tracking models, including Google Flu Trends and Google Dengue Trends. We further extend the model to multiple geographical resolution, tracking infectious disease not only at national level but also at regional level, with spatial-temporal information pooling, making it flexible, self-correcting, robust and scalable.
DL07:	**The Euler Characteristic in Multivariate Distribution Theory and Random Fields, Wednesday, 27 June 2018, 1:30 PM to 3:10 PM, ② LT28**	
DL / Chair	Satoshi Kuriki (Distinguished Lecturer), Institute of Statistical Mathematics, Japan / Akimichi Takemura, Shiga University	
1:30 PM to 2:20 PM:	Satoshi Kuriki Institute of Statistical Mathematics, Japan	"THE EULER CHARACTERISTIC METHOD FOR MULTIVARIATE ANALYSIS AND RANDOM MATRICES" The Euler characteristic method (the expected Euler characteristic heuristic) is an integral-geometric methodology for approximating the tail probability of the supremum of a smooth random field. This method is equivalent to the volume-of-tube method when the random field is zero-mean, unit-variance Gaussian. In the first part of my talk, these methods are briefly reviewed and shown to provide simple and very accurate tail probability formulas. The approximate error is determined by a geometric quantity referred to as the critical radius (or reach) of the tube. In the second half, we apply this method to random matrices arising in multivariate analysis, primarily to Wishart matrices. Noting that the largest eigenvalue of a symmetric random matrix is characterized as the maximum of the corresponding quadratic form, the tail probabilities of the largest eigenvalues are approximated by the Euler characteristic method. We discuss the case where the matrix size is finite and also the case when the size tends to infinity. In both cases, the approximation errors are shown to be exponentially small. Some variations and analogues of the largest eigenvalue, such as the partial sum of several larger eigenvalues, the range of the eigenvalues, the largest singular value of a skew-symmetric matrix, and the largest singular-value of a three-way tensor, are dealt with in the same framework; these distributions appear as the limiting distribution of the likelihood ratio test statistic in singular (locally conic) multivariate models. We also show that the Euler characteristic method can be used to assess p-values in such models. Besides, we treat briefly the problem of constructing simultaneous confidence bands, a major application of the Euler characteristic method.
2:20 PM to 2:45 PM:	Donald Richards Pennsylvania State University	"RECENT DEVELOPMENTS FOR THE SINGULAR VALUES OF SKEW-SYMMETRIC GAUSSIAN RANDOM MATRICES" Skew-symmetric random matrices, despite their mathematically beautiful structure and many potential applications in data analysis, have attracted less attention than their symmetric counterparts. We review in this talk the skew-symmetric Gaussian random matrices and the distributions of their singular values, which were obtained by Kuriki (1992). We review Kuriki's derivation and explain how crucial consequences of his formula are related to an integral evaluated by Harish-Chandra (Amer. J. Math., 1956). By means of that connection, we apply results of K. I. Gross and the speaker (J. Approx. Theory, 1995) to derive new total positivity properties of the joint distribution of those singular values.
2:45 PM to 3:10 PM:	Tomoyuki Shirai Kyushu University	"LIMIT THEOREMS FOR PERSISTENCE DIAGRAMS" Persistent homology appeared around 2000 as an algebraic method which measures topological features of objects or point cloud data. Recently, much attention has been paid to it in the context of Topological Data Analysis (TDA). Persistent homology describes, roughly speaking, the birth and death of homologies (connected components, holes, voids, and so on) of an increasing family of topological objects. Persistent homology is visualized by plotting birth-death pairs in the plane, which is called a persistence diagram. In this talk, I would like to discuss limit theorems for persistent diagrams for stationary point processes.
DL08:	**New Perspectives for Predicative Inference and Modeling, Thursday, 28 June 2018, 10:30 AM to 12:10 PM, ③ LT29**	
DL / Chair	Regina Y. Liu (Distinguished Lecturer), Rutgers University / Alan Wan, City University of Hong Kong	

DISTINGUISHED LECTURE SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
10:30 AM to 11:20 AM:	Regina Y. Liu Rutgers University	"PREDICTION WITH CONFIDENCE — A GENERAL FRAMEWORK FOR PREDICTIVE INFERENCE" We propose a general framework for prediction in which a prediction is in the form of a distribution function, called 'predictive distribution function'. This predictive distribution function is well suited to prescribing the notion of confidence under the frequentist interpretation, and it can provide meaningful answers for prediction-related questions. A general approach under this framework is formulated and illustrated using the so-called confidence distributions (CDs). This CD-based prediction approach inherits many desirable properties of CD, including its capacity to serve as a common platform for directly connecting the existing procedures of predictive inference in Bayesian, fiducial and frequentist paradigms. We discuss the theoretical properties, including efficiency and optimality issues, of the CD-based predictive distribution. We also propose a simple yet broadly applicable Monte-Carlo algorithm for implementation of the proposed approach. This concrete algorithm together with the proposed definition and associate theoretical development provide a comprehensive statistical inference framework for prediction. Finally, this approach is demonstrated by simulation studies and a real project on predicting the incoming volume of application submissions to an agency. The latter in fact shows the applicability of the proposed approach to dependent data settings. This is joint work with Jieli Shen, Goldman Sachs, and Minge Xie, Rutgers University.
11:20 AM to 11:45 AM:	Ying Chen National University of Singapore	"REGULARIZED PARTIALLY FUNCTIONAL AUTOREGRESSIVE MODELING WITH APPLICATION TO NATURAL GAS FLOW FORECASTING" We propose a partially functional autoregressive model to describe the dynamics of serial correlated functional data dependent on its own lagged values and ultra-high dimensional exogenous scalar covariates. Estimation is performed under the two-layer sparsity assumption, where both group and elementary sparsities are simultaneously imposed on the functional and scalar covariates respectively. We established consistency properties of the estimator under certain conditions and investigated the finite sample performance along with simulation studies. In real data analysis, we illustrate the application to natural gas flows forecasting in Germany energy market, with both powerful interpretability and good out-of-sample forecast accuracy. This is a join work with Thorsten Koch and Xiaofei Xu.
11:45 AM to 12:10 PM:	Alan Wan City University of Hong Kong	"MODEL AVERAGING IN A MULTIPLICATIVE HETEROSCEDASTIC MODEL" In recent years, the body of literature on frequentist model averaging in Statistics has grown signi cantly. Most of this work focuses on models with different mean structures but leaves out the variance consideration. In this paper, we consider a regression model with multiplicative heteroscedasticity and develop a model averaging method that combines maximum likelihood estimators of unknown parameters in both the mean and variance functions of the model. Our weight choice criterion is based on a minimisation of a plug-in estimator of the model average estimator's squared prediction risk. We prove that the new estimator possesses an asymptotic optimality property. Our investigation of finite sample performance by simulations demonstrates that the new estimator frequently exhibits very favourable properties compared to some existing heteroscedasticity-robust model average estimators. The model averaging method hedges against the selection of very bad models and serves as a remedy to variance function mis-speci cation, which often discourages practitioners from modeling heteroscedasticity altogether. The proposed model average estimator is applied to the analysis of two data sets on housing and economic growth. This is a joint work with Yanyuan Ma, Pennsylvania State University, and Xinyu Zhang and Shouyang Wang, Chinese Academy of Sciences.
DL09:	**High-Dimensional Bayesian Computation, Friday, 29 June 2018, 3:30 PM to 5:10 PM, ② LT28**	
DL / Chair	Eric Moulines (Distinguished Lecturer), École Polytechnique / Randal Douc, Telecom SudParis	
3:30 PM to 4:20 PM:	Eric Moulines École Polytechnique	"THE LANGEVIN MCMC: THEORY AND METHODS" In machine learning literature, a large number of problems amount to simulate a density which is log-concave (at least in the tails) and perhaps non smooth. Most of the research efforts so far has been devoted to the Maximum A posteriori problem, which amounts to solve a high-dimensional convex (perhaps non smooth) program. The purpose of this lecture is to understand how we can use ideas which have proven very useful in machine learning community to solve large scale optimization problems to design efficient sampling algorithms, with convergence guarantees (and possibly « usable » convergence bounds). In high dimension, only first order method (exploiting exclusively gradient information) is a must. Most of the efficient algorithms know so far may be seen as variants of the gradient descent algorithms, most often coupled with « partial updates » (coordinates descent algorithms). This of course suggests to study methods derived from Euler discretization of the Langevin diffusion, which may be seen as a noisy version of the gradient descent. Partial updates may in this context as « Gibbs steps » where some components are frozen. This algorithm may be generalized in the non-smooth case by « regularizing » the objective function. The Moreau-Yosida inf-convolution algorithm is an appropriate candidate in such case, because it does not modify the minimum value of the criterion while transforming a non smooth optimization problem in a smooth one. We will prove convergence results for these algorithms with explicit convergence bounds both in Wasserstein distance and in total variation.
4:20 PM to 4:45 PM:	Alexandre Thiery National University of Singapore	"SCALING ANALYSIS OF LOCAL CONDITIONAL SMC ALGORITHMS IN HIGH-DIMENSIONS" The analysis of MCMC algorithms through diffusion limits has been applied in many different settings. By approximating an appropriately time-rescaled version of a Markov chain by an ergodic diffusion, one can easily obtain some insight into the mixing behaviour of this Markov chain. In this talk, I will describe ongoing work towards understanding the high-dimensional behaviour of a class of local conditional SMC algorithms. This is joint work with Axel Finke (NUS).

DISTINGUISHED LECTURE SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
4:45 PM to 5:10 PM:	Randal Douc Telecom SudParis	"NUMERICALLY STABLE ONLINE ESTIMATION OF VARIANCE IN PARTICLE FILTERS" Joint work with Jimmy Olsson (KTH). This paper discusses variance estimation in sequential Monte Carlo methods, alternatively termed particle filters. The variance estimator that we propose is a natural modification of that suggested by H. P. Chan and T. L. Lai [A general theory of particle filters in hidden Markov models and some applications. Ann. Statist., 41(6):2877--2904, 2013], which allows the variance to be estimated in a single run of the particle filter by tracing the genealogical history of the particles. However, due particle lineage degeneracy, the estimator of the mentioned work becomes numerically unstable as the number of sequential particle updates increases. Thus, by tracing only a part of the particles' genealogy rather than the full one, our estimator gains long-term numerical stability at the cost of a bias. The scope of the genealogical tracing is regulated by a lag, and under mild, easily checked model assumptions, we prove that the bias tends to zero geometrically fast as the lag increases. As confirmed by our numerical results, this allows the bias to be tightly controlled also for moderate particle sample sizes.
DL10:	**Genomic Hypothesis Testing, Friday, 29 June 2018, 8:30 AM to 10:10 AM, ③ LT29**	
DL / Chair	Art B. Owen (Distinguished Lecturer), Stanford University / ,	
8:30 AM to 9:20 AM:	Art B. Owen Stanford University	"PERMUTATION P-VALUE APPROXIMATION VIA GENERALIZED STOLARSKY INVARIANCE" It is common for genomic data analysis to use p-values from a large number of permutation tests. The multiplicity of tests may require very tiny p-values in order to reject any null hypotheses and the common practice of using randomly sampled permutations then becomes very expensive. We propose an inexpensive approximation to p-values for two sample linear test statistics, derived from Stolarsky's invariance principle. The method creates a geometrically derived set of approximate p-values for each hypothesis. The average of that set is used as a point estimate $\hat{p}$ and our generalization of the invariance principle allows us to compute the variance of the p-values in that set. We find that in cases where the point estimate is small the variance is a modest multiple of the square of the point estimate, yielding a relative error property similar to that of saddlepoint approximations. On a Parkinson's disease data set, the new approximation is faster and more accurate than the saddlepoint approximation. We also obtain a simple probabilistic explanation of Stolarsky's invariance principle.
9:20 AM to 9:45 AM:		#N/A
9:45 AM to 10:10 AM:		#N/A
DL11:	**Recent Advances in Non- and Semi-parametric Structured Models, Wednesday, 27 June 2018, 10:30 AM to 12:10 PM, ② LT28**	
DL / Chair	Byeong Uk Park (Distinguished Lecturer), Seoul National University / Stefan Sperlich, Universite de Geneve	
10:30 AM to 11:20 AM:	Byeong Uk Park Seoul National University	"NONPARAMETRIC ADDITIVE REGRESSION" We discuss statistical methods of estimating structured nonparametric regression models. Our discussion is focused on the additive models where the regression function (map) is expressed as a sum of unknown univariate functions (maps), but it also covers some other non- and semi-parametric models. We present the state of the art in the subject area with an extension to non-Euclidean data objects.
11:20 AM to 11:45 AM:	Shujie Ma University of California at Riverside	"A ROBUST AND EFFICIENT APPROACH TO CAUSAL INFERENCE BASED ON SPARSE SUFFICIENT DIMENSION REDUCTION" A fundamental assumption used in causal inference with observational data is that treatment assignment is ignorable given measured confounding variables. This assumption of no missing confounders is plausible if a large number of baseline covariates are included in the analysis, as we often have no prior knowledge of which variables can be important confounders. Thus, estimation of treatment effects with a large number of covariates has received considerable attention in recent years. Most existing methods require specifying certain parametric models involving the outcome, treatment and confounding variables, and employ a variable selection procedure to identify confounders. However, selection of a proper set of confounders depends on correct specification of the working models. The bias due to model misspecification and incorrect selection of confounding variables can yield misleading results. We propose a robust and efficient approach for inference about the average treatment effect via a flexible modeling strategy incorporating penalized variable selection. Specifically, we consider an estimator constructed based on an efficient influence function that involves a propensity score and an outcome regression. We then propose a new sparse sufficient dimension reduction method to estimate these two functions without making restrictive parametric modeling assumptions. The proposed estimator of the average treatment effect is asymptotically normal and semiparametrically efficient without the need for variable selection consistency. The proposed methods are illustrated via simulation studies and a biomedical application.
11:45 AM to 12:10 PM:	Stefan Sperlich Universite de Geneve	"VARYING COEFFICIENT MODELS FOR MODELLING HETEROGENEITY" The flexibility of semiparametric varying coefficient models is exploited to model heterogeneity that cannot be captured by additive isotonic random deviations from the mean. This is of particular interest for causal analysis when heterogeneity in returns is important. Although they are more restrictive than fully nonparametric methods, varying coefficient models are a powerful compromise between the still popular and typically used linear methods and nonparametrics. Where unobserved heterogeneity is a major concern, they allow to substantially improve in both the interpretation of their estimates and the credibility of the so-called instrumental methods where applied which presently are predominant in empirical economic studies. To this aim, all considerations and developments should be embedded in the modeling of structural equations for which varying coefficients can play a key role. While spline methods are very popular for computational reasons, kernel based smooth backfitting is the maybe most developed method regarding asymptotic theory.
DL12:	**Limit Theorems and Application, Friday, 29 June 2018, 8:30 AM to 10:10 AM, ② LT28**	
DL / Chair	Giovanni Peccati (Distinguished Lecturer), Université du Luxembourg / Mark Podolskji, Aarhus University	

DISTINGUISHED LECTURE SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:30 AM to 9:20 AM:	Giovanni Peccati Université du Luxembourg	"VARIATIONS AROUND DE JONG'S THEOREM" In a groundbreaking 1990 paper, Peter de Jong proved a remarkable central limit theorem, stating that a sequence of degenerate U-statistics - satisfying a Lindeberg type condition - verifies a central limit theorem whenever their fourth cumulants converge to zero. After having described some recent improvements of such a result that I have recently obtained with Ch. Döbler, I will show how de Jong's result has been the seed of an extremely fruitful line of research, related in several ways to the "Malliavin-Stein" method for probabilistic approximations. This connection will allow us to survey some recent quantitative limit theorems for non-linear functionals of Gaussian fields and geometric random graphs. Based on joint works with Ch. Döbler and I. Nourdin (and many others).
9:20 AM to 9:45 AM:	Domenico Marinucci University of Tor Vergata	"THE GEOMETRY OF RANDOM SPHERICAL EIGENFUNCTIONS" The talk will survey some recent results on the high-frequency behaviour of geometric functionals for excursion sets of random spherical eigenfunctions. In particular, we shall discuss how the limiting behaviour of these functionals is asymptotically dominated (in the limit where the eigenvalues diverge to infinity) by a single component, corresponding to the projection of these functionals on the second-order Wiener chaos. A similar behaviour has been established also on other manifolds (such as the torus) and is conjectured to hold in much greater generality. Based on joint papers with Valentina Cammarota, Giovanni Peccati, Maurizia Rossi and Igor Wigman.
9:45 AM to 10:10 AM:	Mark Podolskji Aarhus University	"ON LIMIT THEORY FOR FUNCTIONALS OF STATIONARY INCREMENTS LEVY DRIVEN MOVING AVERAGES" In this talk we present some new limit theorems for variational functionals of high frequency observations of stationary increments Levy driven moving average processes. We will see that the asymptotic behaviour of such variational functionals heavily depends on the kernel, the driving pure jump Levy motion and the properties of the function under consideration. We show the ``law of large numbers'' for our class of statistics, which consists of three different limiting results. For one of the appearing limits, which we refer to as ergodic type limit, we also prove the associated weak limit theory. Finally, we indicate some statistical applications of our theoretical results.
DL13:	**Asymptotic Results for Some Nonparametric Tests, Wednesday, 27 June 2018, 3:30 PM to 5:10 PM, ② LT28**	
DL / Chair	John Robinson (Distinguished Lecturer), University of Sydney / John Kolassa, Rutgers University	
3:30 PM to 4:20 PM:	John Robinson University of Sydney	"ASYMPTOTIC APPROXIMATIONS FOR SOME PERMUTATION AND BOOTSTRAP TESTS" Permutation tests, rank tests and bootstrap tests are nonparametric methods used to give accurate test procedures when doubts of the validity of parametric models exist. Exact permutation tests valid for overall tests in analysis of variance and regression automatically provide valid tests. Bootstrap methods are more flexible, applying to a wider variety of problems, but they depend on both an asymptotic validation of the bootstrap accuracy and on approximations to the distribution of the bootstrap statistic. Frequently these tests are based on test statistics chosen for a parametric problem. We describe an approach in the nonparametric case, where the statistic is chosen based on the empirical cumulant generating function of some score functions used to obtain M-estimates. We will give a brief indication of the conditions and proof of results justifying the relative second order approximation for the distribution of the test statistic under a multiparameter hypothesis. These provide a nonparametric approach for linear models and generalized linear models. We will show numerically that this approach can provide a test which is more powerful under deviations from the parametric model and maintains the power of the parametric test under the model.
4:20 PM to 4:45 PM:	Alan Huang University of Queensland	"TWO FLAVOURS OF EMPIRICAL LIKELIHOOD" Empirical likelihood is a powerful yet flexible nonparametric method for model inferences in the presence of an infinite-dimensional distribution parameter. When coupled with estimating equations, the asymptotic behaviour of empirical likelihood can often be reduced to a problem in M-estimation. However, in cases in which the number of constraints increase with sample size, asymptotic analysis of empirical likelihood requires tools from semiparametric efficiency theory. We will compare and contrast both approaches in the context of semiparametric generalized and vector generalized linear models.
4:45 PM to 5:10 PM:	John Kolassa Rutgers University	"SIZE AND POWER CONSIDERATIONS FOR MULTIVARIATE RANK STATISTICS" Rank statistics that are multivariate, perhaps because of a sequential testing scheme, or because multiple response variable are ranked, present inferential challenges when accuracy beyond simple normal theory approximations is required. This talk presents applications of Edgeworth and Cornish-Fisher approaches to more accurate asymptotic inference.
DL14:	**Recent Advances in Survival Analysis, Wednesday, 27 June 2018, 8:30 AM to 10:10 AM, ③ LT29**	
DL / Chair	Ingrid Van Keilegom (Distinguished Lecturer), KU Leuven / Valentin Patilea, CREST, Ensai	
8:30 AM to 9:20 AM:	Ingrid Van Keilegom KU Leuven	"GOODNESS-OF-FIT TESTS IN SURVIVAL MODELS WITH RANDOM EFFECTS" The paper deals with testing the functional form of the covariate effects in a Cox model with random effects, like for instance a shared frailty model. We assume that the responses are clustered and incomplete due to right censoring. The estimation of the model under the null (parametric covariate effect) and the alternative (non-parametric effect) is performed using full marginal likelihood. Under the alternative, the non-parametric covariate effects are represented using an orthogonal polynomial expansion. The test statistic is the likelihood ratio statistic, and its distribution is approximated using a bootstrap method. The proposed testing procedure is studied theoretically and through simulations.
9:20 AM to 9:45 AM:	Wenbin Lu North Carolina State University	"TESTING AND ESTIMATION OF SOCIAL NETWORK DEPENDENCE WITH TIME TO EVENT DATA" In this work, we propose a novel latent spatial autocorrelation Cox model to study social network dependence with time-to-event data. The proposed model introduces a latent indicator to characterize whether a person's survival time might be affected by his or her friends' features. We first propose a score-type test for detecting the existence of social network dependence. If it exists, we further develop an EM-type algorithm to estimate the model parameters. The performance of the proposed test and estimators are illustrated by simulation studies and an application to a time-to-event data set about playing a popular QQ game from Tencent.

DISTINGUISHED LECTURE SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
9:45 AM to 10:10 AM:	Valentin Patilea CREST, Ensai	"INFERENCE WITH CONDITIONAL MOMENT EQUATIONS IN THE PRESENCE OF CENSORING: APPLICATION TO CURE MODELS" In survival analysis it often happens that some subjects under study do not experience the event of interest; they are considered to be `cured'. The population is thus a mixture of two subpopulations: the one of cured subjects, and the one of `susceptible' subjects. Based on a conditional estimating equation, we propose a new regression approach for the cure rate, that is for the conditional probability of being cured, when the data are subject to random right censoring. Both continuous and discrete covariates are allowed and the parameters of interest are estimated by least squares, without any reference to the law of the susceptible subjects. A penalized version of the least squares criterion allows for variable selection. A new model check procedure that avoids the curse of dimensionality is also proposed. The asymptotic results and the inference are based on the i.i.d. representation of the conditional survival function estimator of the censoring variable. Such representations could be derived under mild conditions, several examples are provided. Simulation experiments illustrate the effectiveness of the new modeling approach. The talk is based on joint work with Kevin Burke from University of Limerick.
DL15:	**Regularity and Ergodicity for Singular Stochastic Differential Equations, Wednesday, 27 June 2018, 1:30 PM to 3:10 PM, ③ LT29**	
DL / Chair	Fengyu Wang (Distinguished Lecturer), Beijing Normal University / Jinghai Shao, Tianjin University	
1:30 PM to 2:20 PM:	Fengyu Wang Beijing Normal University	"ESTIMATES ON INVARIANT PROBABILITY MEASURES FOR SINGULAR SDES" In terms of a nice reference probability measure, integrability conditions on the path-dependent drift are presented for (infinite-dimensional) degenerate PDEs to have regular positive solutions. To this end, the corresponding stochastic (partial) differential equations are proved to possess the weak existence and uniqueness of solutions, as well as the existence, uniqueness and entropy estimates of invariant probability measures. When the reference measure satisfies the log-Sobolev inequality, Sobolev estimates are derived for the density of invariant probability measures. Some results are new even for non-degenerate SDEs with path-independent drifts. The main results are applied to nonlinear functional SPDEs and degenerate functional SDEs/SPDEs.
2:20 PM to 2:45 PM:	Xing Huang Tianjin University	#N/A
2:45 PM to 3:10 PM:	Jinghai Shao Tianjin University	"INVARIANT MEASURES AND EULER-MARUYAMA'S APPROXIMATIONS OF STATE-DEPENDENT REGIME-SWITCHING DIFFUSION" Regime-switching processes contain two components: continuous component and discrete component, which can be used to describe a continuous dynamical system in a random environment. Such processes have many different properties than general diffusion processes, and much more difficulties are needed to be overcome due to the intensive interaction between continuous and discrete component. We give conditions for the existence and uniqueness of invariant measures for state-dependent regime-switching diffusion processes. We also establish the strong convergence in the L^1-norm of the Euler-Maruyama's approximation and estimate the order of error. A refined application of Skorokhod's representation of jumping processes plays a substantial role in this work.
DL16:	**Recent Advances in Statistical Genetics, Tuesday, 26 June 2018, 1:30 PM to 3:10 PM, ② LT28**	
DL / Chair	Hongyu Zhao (Distinguished Lecturer), Yale University / Can Yang, Hong Kong University of Science and Technology	
1:30 PM to 2:20 PM:	Hongyu Zhao Yale University	"DATA INTEGRATION METHODS TO UNDERSTAND COMPLEX DISEASE GENETICS" Genome-wide association study (GWAS) has been a great success in the past decade. However, significant challenges still remain in both identifying new risk loci and interpreting results. Complex structure of linkage disequilibrium also makes it challenging to separate causal variants from nonfunctional ones in large haplotype blocks. In this presentation, I will describe our recent efforts to integrate genomic functional annotations from computational predictions (e.g. genomic conservation) and high-throughput experiments (e.g. the ENCODE and Roadmap Epigenomics Projects) with GWAS summary statistics. Tissue and cell specific annotations allow us to infer relevant tissue/cell types at each risk locus. The usefulness of our methods will be demonstrated through their applications to several large GWASs. I will also discuss our approach to inferring genetic correlations from summary statistics. Joint analysis of multiple GWAS results allows us to infer genetic correlations among many complex traits. Finally, I will brief discuss the improvement of genetic risk prediction using annotation data.
2:20 PM to 2:45 PM:	Dongjun Chung Medical University of South Carolina	"A STATISTICAL FRAMEWORK FOR THE INTEGRATION OF MULTIPLE GWAS RESULTS WITH LITERATURE MINING DATA" Integration of genetic studies for multiple diseases with biomedical big data has been shown to be a powerful approach to improving the identification of risk genetic variants. However, it still remains challenging to effectively integrate genome-wide association studies (GWAS) datasets for multiple diseases and fully utilize information in biomedical big data. In this presentation, I will discuss our novel DDNet-graph-GPA framework which addresses this challenge. graph-GPA is a novel Bayesian approach to integrate multiple GWAS datasets using a latent Markov random field architecture, which also allows to incorporate various external prior knowledge. To further improve genetic analyses using biomedical big data, we implemented an effective text mining of biomedical literature utilizing gene ontology knowledge and further developed DDNet, a database and web interface that allow researchers mine relationships among diseases based on the biomedical literature. The disease-disease graph obtained from DDNet can be used as prior information for graph-GPA. The proposed approach improves the identification of risk variants and facilitates understanding of genetic relationships among complex diseases. I will illustrate the proposed approach with simulation studies and its application to real GWAS datasets. Finally, I will discuss our current work for more effective utilization of biomedical literature data, including GAIL and BayesGO.

DISTINGUISHED LECTURE SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
2:45 PM to 3:10 PM:	Can Yang Hong Kong University of Science and Technology	"COMM: A COLLABORATIVE MIXED MODEL TO DISSECTING GENETIC CONTRIBUTIONS TO COMPLEX TRAITS" Genome-wide association studies (GWASs) have been successful in identifying genetic variants associated with complex traits. However, the mechanistic links underlying how these genetic variants affect complex traits remain elusive. A scientific hypothesis is that genetic variants influence complex traits at the organismal level via affecting cellular traits, such as regulating gene expression and altering protein abundance. Although earlier works have already presented some scientific insights about this hypothesis and their findings are very promising, statistical methods that effectively harness multilayer data (e.g., genetic variants, cellular traits and organismal traits) on a large scale for functional and mechanistic exploration are highly demanding. In this article, we propose a collaborative mixed model (CoMM) to dissect genetic contributions to complex traits by leveraging regulatory information in transcriptome data so that the mechanistic role of associated variants could be fully revealed. To demonstrate the advantages of CoMM over existing methods, we conducted extensive simulation studies and also applied CoMM to analyze 25 traits in NFBC1966 and Genetic Epidemiology Research on Aging (GERA) studies. The results indicate that by leveraging regulatory information, CoMM can effectively improve the power of prioritizing risk variants. This is the join work of Xiang Wan, Xinyi Lin, Mengjie Chen, Xiang Zhou and Jin Liu.

Last updated
6/25/2018

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
IP01	**New Techniques in High-Dimensional Data Analysis, Tuesday, 26 June 2018, 1:30 PM to 3:10 PM, ③ LT29**	
Org / Chair	Makoto Aoshima, University of Tsukuba / Makoto Aoshima, University of Tsukuba	
1:30 PM to 1:55 PM	Haipeng Shen University of Hong Kong	"TO WAIT OR NOT TO WAIT: TWO-WAY FUNCTIONAL HAZARDS MODEL FOR UNDERSTANDING WAITING IN CALL CENTERS" Telephone call centers offer a convenient communication channel between businesses and their customers. Efficient management of call centers needs accurate modeling of customer waiting behavior, which contains important information about customer patience (how long a customer is willing to wait) and service quality (how long a customer needs to wait to get served). Hazard functions offer dynamic characterization of customer waiting behavior, and provide critical inputs for agent scheduling. Motivated by this application, we develop a two-way functional hazards (tF-Hazards) model to study customer waiting behavior as a function of two timescales, waiting duration and the time of day that a customer calls in. The model stems from a two-way piecewise constant hazard function, and imposes low-rank structure and smoothness on the hazard rates to enhance interpretability. We exploit an alternating direction method of multipliers (ADMM) algorithm to optimize a penalized likelihood function of the model. We carefully analyze the data from a US bank call center, and provide informative insights about customer patience and service quality patterns along waiting time and across different times of a day. The findings provide primitive inputs for call center agent staffing and scheduling, as well as for call center practitioners to understand the effect of system protocols on customer waiting behavior.
1:55 PM to 2:20 PM	Cheng Wang Shanghai Jiao Tong University	"ON THE DIMENSION EFFECT OF REGULARIZED LINEAR DISCRIMINANT ANALYSIS" This paper studies the dimension effect of the linear discriminant analysis (LDA) and the regularized linear discriminant analysis (RLDA) classifiers for large dimensional data where the observation dimension p is of the same order as the sample size n. More specifically, built on properties of the Wishart distribution and recent results in random matrix theory, we derive explicit expressions for the asymptotic misclassification errors of LDA and RLDA respectively, from which we gain insights of how dimension affects the performance of classification and in what sense. Motivated by these results, we propose adjusted classifiers by correcting the bias brought by the dimension effect. Several interesting examples are discussed in detail to verify our theoretical results. This is a joint work with Binyan Jiang.
2:20 PM to 2:45 PM	Hirokazu Yanagihara Hiroshima University	"CONSISTENT GENERALIZED CP IN HIGH-DIMENSIONAL MULTIVARIATE LINEAR MODELS UNDER NONNORMALITY" In this paper, we deal with a variable selection in multivariate linear regression models, based on minimization of the generalized Cp (GCp) criterion when the dimension of the response variables vector may be large. Recently, Yanagihara (2016) proposed the high-dimensionality-adjusted consistent GCp (HCGCp) that is the consistent GCp criterion for which consistency can be achieved whenever the dimension of the response variables vector is fixed or goes to infinity. A high probability of selecting the true subset of explanatory variables can be expected under a moderate sample size when the HCGCp criterion is used to select variables, even when there is a high-dimensional response variables vector. Unfortunately, Yanagihara (2016) showed the consistency of HCGCp under the assumption that the true distribution of response variables is the multivariate normal distribution. Needless to say, nobody knows the true distribution of response variables. Hence, we show that the robustness to nonnormality of the HCGCp, i.e., the HCGCp has a consistency property under the violation of normality of the true distribution.
2:45 PM to 3:10 PM	Kazuyoshi Yata University of Tsukuba	"REGULARIZED PCA FOR HIGH-DIMENSIONAL DATA BASED ON THE NOISE-REDUCTION METHODOLOGY" In this talk, we consider PCA methods in high-dimensional settings. We first show that the conventional PCA is affected by the high-dimensional noise structure directly. In order to overcome the difficulty, Yata and Aoshima (2012, JMVA) developed a new PCA method called the noise-reduction (NR) methodology. We show that the NR method can enjoy consistency properties for both eigenvalues and PC directions in high-dimensional settings. The estimator of the PC directions by the NR method has a consistency property in terms of an inner product. However, it does not hold a consistency property in terms of the angle. With the help of a thresholding method, we modify the estimator and propose a regularized PCA method. We show that it holds the consistency property of the angle. Finally, we check the performance of the regularized PCA method by using microarray data sets.
IP02	**Stochastic Analysis and Applications, Friday, 29 June 2018, 1:30 PM to 3:10 PM, ③ LT29**	
Org / Chair	Sayan Banerjee, University of North Carolina at Chapel Hill / Sayan Banerjee, University of North Carolina at Chapel Hill	
1:30 PM to 1:55 PM	Siva Athreya Indian Statistical Institute, Bangalore	"STRONG EXISTENCE AND UNIQUENESS FOR STABLE SDES" We consider the stochastic differential equation dX_t = b(X_t) dt + dL_t where the drift b is a generalized function and L is a symmetric one dimensional α-stable Lévy process, 1 < α < 2. We define the notion of solution to this equation and establish strong existence and uniqueness whenever b belongs to the Besov–Hölder space C^β for β > 1/2 − α/2.
1:55 PM to 2:20 PM	Sayan Banerjee University of North Carolina at Chapel Hill	"ANALYZING THE DIFFUSION LIMIT OF JOIN THE SHORTEST QUEUE POLICY" Consider a system of N parallel single-server queues with unit-exponential service time distribution and a single dispatcher where tasks arrive as a Poisson process of rate Î»(N). When a task arrives, the dispatcher assigns it to one of the servers according to the Join-the-Shortest Queue (JSQ) policy. Eschenfeldt and Gamarnik (2015) established that in the Halfin-Whitt regime where (N â^’ Î»(N))/ â^š N â†’ Î² > 0 as N â†’ â^ž, the appropriately scaled occupancy measure of the system under the JSQ policy converges weakly on any finite time interval to a certain non-elliptic reflected diffusion process as N â†’ â^ž. Recently, it was established by Braverman (2018) that the convergence result extends to the steady state as well, i.e., the stationary occupancy measure of the system converges weakly to the steady state of the diffusion process as N â†’ â^ž, proving the interchange of limits result. I will talk about analyzing the detailed behavior of the steady state of this diffusion process using tools from renewal theory. The tails and bulk behavior of the steady state distribution and sample path fluctuations of the diffusion process will be explored. We will also see how the steady state shows a stark difference in behavior between the small and large Î² regimes. This is joint work with Debankur Mukherjee.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
2:20 PM to 2:45 PM	Amarjit Budhiraja University of North Carolina at Chapel Hill	"LARGE DEVIATIONS FROM THE HYDRODYNAMIC LIMIT FOR A SYSTEM WITH NEAREST NEIGHBOR INTERACTIONS" We present a new proof of the large deviation principle from the hydrodynamic limit for the Ginzberg-Landau model studied in Donsker-Varadhan(1989) using techniques from the theory of stochastic control and weak convergence methods. The proof is based on characterizing subsequential hydrodynamic limits of controlled diffusions with nearest neighbor interaction that arise from a variational representation of certain Laplace functionals. The approach taken here does not require superexponential probability estimates or estimation of exponential moments that are central ingredients in previous proofs. Instead, proof techniques are very similar to those used for the law of large number analysis, namely in the proof of convergence to the hydrodynamic limit (cf. Guo-Papanicolaou-Varadhan(1988)). Specifically, the key step in the proof is establishing suitable bounds on relative entopies and Dirichlet forms associated with certain controlled laws. This general approach has the promise to be applicable to other interacting particle systems as well and to the case of non-equilibrium starting configurations, and to infinite volume systems. This is joint work with S. Banerjee and M. Perlmutter.
2:45 PM to 3:10 PM	Rongfeng Sun National University of Singapore	"MOMENTS OF THE (2+1)-DIMENSIONAL DIRECTED POLYMER IN THE CRITICAL WINDOW" Recently, we have shown that the partition function of the directed polymer model on Z^{2+1} admits a phase transition in a suitable continuum and weak disorder limit. In particular, the partition function converges in law to a log-normal distribution below the critical point, and converges to 0 at and above the critical point. Here we focus on a suitable window around the critical point, and we prove that the space-averaged point-to-plane partition function has a uniformly bounded third moment. As a consequence, when interpreted as a random measure on R^2, the rescaled point-to-plane partition functions have non-trivial limit points, and each limit point has the same explicit covariance structure.
IP03	**Statistical Inference Based on Divergence Measures, Wednesday, 27 June 2018, 10:30 AM to 12:10 PM, ③ LT29**	
Org / Chair	Ayanendranath Basu, Indian Statistical Institute, Kolkata / Ayanendranath Basu, Indian Statistical Institute, Kolkata	
10:30 AM to 10:55 AM	Ayanendranath Basu Indian Statistical Institute, Kolkata	"ROBUST WALD-TYPE TESTS UNDER RANDOM CENSORING" Randomly censored survival data are frequently encountered in applied sciences including biomedical and reliability applications. We propose Wald-type tests for testing parametric statistical hypothesis, both simple as well as composite, for randomly censored data using the M-estimators under a fully parametric set-up. We propose a consistent estimator of the asymptotic variance of the M-estimators based on sample data without any assumption on the form of the censoring scheme. General asymptotic and robustness properties of the proposed Wald-type tests are developed. Their advantages and usefulness are demonstrated in detail for Wald-type tests based on a particular M-estimator, namely the minimum density power divergence estimator. This is a joint work with Abhik Ghosh and Leandro Pardo.
10:55 AM to 11:20 AM	Michel Broniatowski University Pierre et Marie Curie	"A MONTE CARLO APPROACH TO DIVERGENCE MINIMIZATION PROBLEMS" Sanov type results hold for some weighted versions of empirical measures, and the rates for those Large Deviation principles can be identi ed as divergences between measures, which in turn characterize the form of the weights. This correspondence is considered within the range of the Cressie-Read family of statistical divergences, which covers most of the usual statistical criterions. We propose a weighted bootstrap procedure in order to estimate these rates. To any such rate we produce an explicit procedure which defines the weights, therefore replacing a variational problem in the space of measures by a simple Monte Carlo procedure. Examples include estimation of divergences between the distribution of the sample and tubular neighborhoods of parametric models or classes of distributions defined by ordering properties.
11:20 AM to 11:45 AM	Hironori Fujisawa Institute of Statistical Mathematics	"ON DIFFERENCE BETWEEN TWO TYPES OF GAMMA-DIVERGENCE IN ROBUST REGRESSION MODELING" Two types of gamma-divergence are proposed in regression modeling. When the response variable follows a location-scale family where the explanatory variables affects on the location but not on the scale, the parameter estimations based on two types of gamma-divergence are the same. However, two types of gamma-divergence presents different parameter estimations in general. In particular, when the outlier ratio depends on the explanatory variables, one is superior to the other in the sense of bias reduction.
11:45 AM to 12:10 PM	Wolfgang Stummer University of Erlangen Nurnberg	"SOME UNIVERSAL INSIGHTS ON DIVERGENCES FOR STATISTICS, MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE" Dissimilarity quantifiers such as divergences (e.g. Kullback-Leibler information, relative entropy, Pearson's chisquare distance, Hellinger distance) and distances between probability distributions are widely used in statistics, machine learning, information theory and adjacent artificial intelligence. Within these fields, in contrast, some applications deal with divergences between other-type real-valued functions and vectors. We present a correspondingly unifying framework and discuss some specificalities, subtleties as well as pitfalls when e.g. one ``moves away'' from the probability context. Several subcases and examples are given, including a new approach to obtain parameter estimators in continuous models which is based on noisy divergence minimization. This talk is based on joint work with Michel Broniatowski from Sorbonne Universite Pierre et Marie Curie Paris.
IP04	**New Advances in Statistical Methods for Complex Data in Biomedical Studies, Tuesday, 26 June 2018, 3:30 PM to 5:10 PM, ③ LT29**	
Org / Chair	Hongyuan Cao, University of Missouri-Columbia / Jialiang Li, National University of Singapore	
3:30 PM to 3:55 PM	Yu Shen MD Anderson Cancer Center	"AFT REGRESSION ANALYSES OF SURVIVAL DATA SUBJECT TO BIASED SAMPLING" Methodologic development in semiparametric modeling of length-biased data has made considerable progress in recent years in many different directions. We will give an overview on recent semiparametric modeling for right-censored survival data under length-biased sampling. The various approaches will be reviewed for commonly used AFT model for time-to-event outcomes. The estimation methods cover both estimating equation approaches and likelihood-based methods using EM algorithm. Some related software for the implementation of such methods will be illustrated. (Team members on the multiple projects include Jing Ning and Jing Qin).
3:55 PM to 4:20 PM	Antai Wang New Jersey Institute of Technology	#N/A

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
4:20 PM to 4:45 PM	Wanjie Wang National University of Singapore	"FUNCTIONAL DEBIAS ESTIMATION OF GENETIC RELATEDNESS IN HIGHÂ€ DIMENSIONAL LINEAR MODELS" Motivated by the co-heritability estimation problem in genomics, we study the inner product, quadratic functional and ratio in high-dimensional linear regression. Observing (X_i,Y_i) for N_1 samples, and (Z_j, W_j) for the other N_2 samples, the feature vectors Xi and Zj are observed on the same features, and the response scalars Y_i and W_j are observed on different responses. The quantity of interest is the inner product of parameters vectors θ_1 for response 1 and θ_2 for response 2. The regime of interest is that the number of features is larger than the sample size, and the non-zero coordinates in θ_1 and θ_2 are sparse. We propose a functional de-biased estimator for the inner product, quadratic functional and ratio. Beginning with the lasso estimator, we correct it by the estimation of the inner product between θ_1 (respectively, θ_2) and error of the estimation for θ_2 (respectively, θ_1), and take the summation as the corrected estimator. We prove that this method yields the optimal convergence rate, and apply it to a yeast segregant data set with multiple traits to estimate the genetic relatedness among them.
4:45 PM to 5:10 PM		#N/A
IP05	**Recent Developments in Clinical Trials, Tuesday, 26 June 2018, 3:30 PM to 5:10 PM, ④ LT26**	
Org / Chair	Yuh Ing Chen, National Central University / Yuh Ing Chen, National Central University	
3:30 PM to 3:55 PM	Yuh Ing Chen National Central University	"ADAPTIVE DOSE-FINDING METHODS FOR PHASE I/II CLINICAL TRIALS" In early phase clinical trials for drug development, one-stage designs taking into account both the toxicity and efficacy are usually employed to identify the optimal dose at which the drug reaches its highest efficacy with acceptable toxicity. However, the one-stage designs may not be suitable when the efficacy takes a long time to evaluate. Therefore, we consider a two-stage design to find the optimal dose, where an adaptive procedure is used at the first stage to tentatively estimate the maximum tolerated dose (MTD) and batches of patients are then sequentially assigned to receive the adjusted doses. A simulation study is conducted to investigate the performance of the early phase designs on the identification of the optimal dose and the required trial time. The results show that the two-stage design is competitive to the one-stage design on identifying the optimal dose and treating patients, but generally requires a shorter time to find the optimal dose.
3:55 PM to 4:20 PM	Lu Cui Abbvie	"ADAPTIVE GROUP SEQUENTIAL TRIAL WITH FLEXIBLE TIMING OF SAMPLE SIZE DETERMINATION" Flexible sample size design as an adaptive clinical trial approach vs. fixed sample size design may lead to better trial efficiency and more robust statistical power. This presentation is to further investigate the so called 'sample size re-estimation method' and reclassify it as adaptive group sequential design with open-ended final sample size which is to be projected based on unblinded interim data. Multiple design parameters, including information time or weight, can be varied for design optimization against a chosen performance criterion. As the result, an efficient design with robust statistical power can be obtained. Comments on design optimization as well as performance comparison of the proposed design vs. fixed sample group sequential design are provided to illustrate the viewpoint presented.
4:20 PM to 4:45 PM	Toshimitsu Hamasaki National Cerebral and Cardiovascular Center	"DESIGNING EFFICIENT CLINICAL TRIALS WITH MULTIPLE EVENT- TIME OUTCOMES" In clinical trials, a use of multiple time-to-event outcomes as primary endpoints have been used in disease areas such as oncology, infectious disease, and cardiovascular and cerebral disease, in order to provide a more complete assessment of the interventions, to answer more complex clinical questions, or to evaluate concerns about balancing risks with benefits of the interventions. Due to multiple outcome evaluation, such clinical trials could be expensive and resource intensive as they often require enrollment of large numbers of participants, collection of large amounts of data, and long term follow-up. Use of group-sequential designs has the potential to streamline such clinical trials making them more efficient, i.e., offering potentially fewer required trial participants, shortening the duration of clinical trials, and reducing costs. In clinical trials with a single time-to-event as the primary endpoint, the clinical cutoff is often event-driven and the logrank test is the most commonly used statistical method for evaluating intervention effect. However, methods for multiple time-to-event outcomes are more complex. The number of required events can be different among the endpoints with varying follow-up duration for each endpoint, and the amount of information for the endpoints may vary at a particular interim time-point of the trial monitoring. The censoring mechanism further complicates the design of these trials. We illustrate these practical issues using real clinical trials and explore sample size recommendations, alternative strategies for clinical cutoffs, and efficient testing strategies with the highest power to detect effects on: (i) at least one of the endpoint, and (2) joint effects on all of the endpoints.
4:45 PM to 5:10 PM	Chin Fu Hsiao National Health Research Institutes	"AN ADAPTIVE PHASE I/II DESIGN" In the traditional drug development, the objective of the phase I clinical trial is to find the maximum tolerated dose (MTD) of the drug administrated to human, while the objective of the Phase II clinical trial is to determine whether the drug has promising effects at the MTD. Pharmaceutical development is a risky, complex, costly and time-consuming endeavor, and thus it is desired to shortening the duration of drug development. Instead of conducting the phase I and phase II clinical trials separately, we propose a phase I/II design based on Bayesian approach. That is, in the phase I/II design we developed, our objective is to find the MTD and confirm the efficacy at each dose level simultaneously. In our design, the trial starts with the lowest dose level. The beta-binomial Bayesian posterior probabilities are calculated for the response rates of toxicity and efficacy. The decision rules of De-escalation/Stop/Escalation dose depend on the maximum unit probability mass (UPM) defined as the ratio of probability interval and the length of interval.
IP06	**Recent Developments in Change-Point Problems, Thursday, 28 June 2018, 8:30 AM to 10:10 AM, ④ LT26**	
Org / Chair	Haeran Cho, University of Bristol / Hao Chen, University of California, Davis	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:30 AM to 8:55 AM	Rebecca Killick Lancaster University	"MULTIVARIATE CHANGEPOINT DETECTION WITH SUBSETS" Historically much of the research on changepoint analysis has focused on the univariate setting. However, increasingly data found in contemporary scientific fields are multivariate in nature, with each observation in a sequence containing the values of multiple variables which have been observed simultaneously. The multivariate changepoints which may be observed within such time series can be categorized as either fully-multivariate or subset-multivariate. Fully-multivariate changepoints refer to those changes in structure which occur simultaneously in all variables. Conversely, subset-multivariate changepoints refer to those which occur in only a subset of the observed variables. Traditionally, multivariate changepoint detection methods typically assume that all changes within a series are fully-multivariate. Some recent papers have taken into account that all variables may not change but either do not explicitly output the subsets or do a fully multivariate analysis followed by subset determination. The work in this presentation is the first to create a dynamic program specifically for detecting changes in subset-multivariate time series. We present both exact and approximate optimization methods for determining the number, location and affected subsets of changepoints. Simulation studies demonstrate the performance of the approach on fully multivariate, as well as large, medium and small proportions of affected subsets and we apply the approach to acoustic sensing data.
8:55 AM to 9:20 AM	Wai Leong Ng Chinese University of Hong Kong	"OPTIMAL ESTIMATION OF CHANGE-POINT IN TIME SERIES" This paper establishes asymptotic theory for optimal estimation of change-points in time series. We show that the Bayes estimator is asymptotically efficient for change-point estimation under mean-squared error loss. Two bootstrap procedures are developed to construct confidence intervals for the change-points. Approximated limiting distribution of the change-point estimation under small change is also derived. Simulations and real data applications are presented to investigate the finite sample performance of the Bayes estimator and the two bootstrap procedures.
9:20 AM to 9:45 AM	Tengyao Wang University of Cambridge	"HIGH-DIMENSIONAL CHANGEPOINT ESTIMATION VIA SPARSE PROJECTION" Changepoints are a very common feature of Big Data that arrive in the form of a data stream. In this paper, we study high-dimensional time series in which, at certain time points, the mean structure changes in a sparse subset of the coordinates. The challenge is to borrow strength across the coordinates in order to detect smaller changes than could be observed in any individual component series. We propose a two-stage procedure called \texttt{inspect} for estimation of the changepoints: first, we argue that a good projection direction can be obtained as the leading left singular vector of the matrix that solves a convex optimisation problem derived from the CUSUM transformation of the time series. We then apply an existing univariate changepoint detection algorithm to the projected series. Our theory provides strong guarantees on both the number of estimated changepoints and the rates of convergence of their locations, and our numerical studies validate its highly competitive empirical performance for a wide range of data generating mechanisms.
9:45 AM to 10:10 AM		#N/A
IP07	**Markov Chains and Related Topics, Friday, 29 June 2018, 8:30 AM to 10:10 AM, ⑤ LT33**	
Org / Chair	Michael Choi, Cornell University / Michael Choi, Cornell University	
8:30 AM to 8:55 AM	Swee Hong Chan Cornell University	"IN BETWEEN RANDOM WALK AND ROTOR WALK IN THE SQUARE LATTICE" How much randomness is needed to prove a scaling limit result? In this talk we consider this question for a family of random walks on the square lattice. When the randomness is turned to the maximum, we have the symmetric random walk, which is known to scale to a two-dimensional Brownian motion. When the randomness is turned to zero, we have the rotor walk, for which its scaling limit is an open problem. This talk is about random walks that lie in between these two extreme cases and for which we can prove their scaling limit. This is a joint work with Lila Greco, Lionel Levine, and Boyao Li.
8:55 AM to 9:20 AM	Michael Choi Cornell University	"HITTING TIME AND MIXING TIME BOUNDS OF STEIN'S FACTORS" For any discrete target distribution, we exploit the connection between Markov chains and Stein's method via the generator approach and express the solution of Stein's equation in terms of expected hitting time. This yields new upper bounds of Stein's factors in terms of the parameters of the Markov chain, such as mixing time and the gradient of expected hitting time. We compare the performance of these bounds with those in the literature, and in particular we consider Stein's method for discrete uniform, binomial, geometric and hypergeometric distribution. As another application, the same methodology applies to bound expected hitting time via Stein's factors. This article highlights the interplay between Stein's method, modern Markov chain theory and classical fluctuation theory.
9:20 AM to 9:45 AM	Daniel Jerison Tel Aviv University	"BEYOND THE MONTE CARLO STANDARD ERROR: HONEST MCMC CONVERGENCE GUARANTEES" Is MCMC estimation as trustworthy as sampling directly from the target probability distribution (if that were feasible)? Usually not: Monte Carlo standard errors, which purport to measure the uncertainty introduced by the Markov chain, are asymptotically valid but provide no finite-time guarantees. The few nonasymptotic results are difficult to apply in practice. I will discuss new MCMC estimation theorems for Markov chains with a regenerative structure. These theorems give accuracy guarantees of the same type that sampling directly from the target would provide.
9:45 AM to 10:10 AM		#N/A
IP08	**Functional Data Analysis and Related Topics, Tuesday, 26 June 2018, 1:30 PM to 3:10 PM, ⑥ LT34**	
Org / Chair	Aurore Delaigle, University of Melbourne / ,	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
1:30 PM to 1:55 PM	Gery Geenens University of New South Wales	"THE HELLINGER DEPENDENCE MEASURE AND ITS FUNCTIONAL ANALOGUE" In his seminal paper, Renyi (1959) formulated 7 axioms that an ideal dependence measure between two random variables should satisfy. However, these axioms were proved too strong for general use. In this work, we first propose a list of more reasonable properties that any dependence measure between two random variables X and Y should fulfil, and we exhibit a measure which does. That measure is essentially the Hellinger distance between the joint distribution of (X,Y) and its independence base case. We show how to obtain a root-n consistent estimator of it by borrowing ideas from k-nearest-neighbours techniques. In a second time, we extend those ideas to the case of two functional random variables.
1:55 PM to 2:20 PM	Naisyin Wang University of Michigan	"ESTIMATION AND INFERENCE OF TIME-VARYING COEFFICIENTS IN NON-LINEAR ODE MODELS" The use of ordinary differential equations (ODEs) in modeling dynamic systems has gained high popularity in the recent decade. The physiological meanings of ODE parameters are often useful in enabling scientists to gain better understanding of the underlying system. On this regard, both estimation and inference procedures are essential. Even though various time-varying coefficient ODE model has been considered previously. The inference procedures considered earlier tends to be similar to what has been used in parametric models. We propose a new set of estimation and inference procedures for time-varying ODE coefficients. Our methods take into account features that are unique for ODE estimation and, as such, are adaptive in nature. The validity of the proposed procedures is justified through asymptotic properties. The numerical efficacy of the methodologies is illustrated using both synthetic and real-world data-sets.
2:20 PM to 2:45 PM	Zhigang Yao National University of Singapore	"PRINCIPAL SUB-MANIFOLDS" We invent a novel method of finding principal components to the multivariate data sets, that lie on an embedded nonlinear Riemannian manifold within a higher-dimensional space. Our aim is to extend the geometric interpretation of PCA, while being able to capture non-geodesic mode of variation in the data. We introduce the concept of a principal sub-manifold, a manifold passing through the center of the data, and at any point on the manifold, it moves in the direction of highest variation in the space spanned by the eigenvectors of the local tangent space PCA. Compared to the recent work in the case where the sub-manifold is of dimension one [Panaretos, Pham and Yao, 2014]--essentially a curve lying on the manifold attempting to capture the one-dimensional variation--the current setting is much more general. The principal sub-manifold is therefore an extension of the principal flow, accommodating to capture the higher dimensional variation in the data. We show the principal sub-manifold yields the ball spanned by the usual principal components in Euclidean space. By means of examples, we illustrate how to find, use and interpret principal sub-manifold with an extension of using it in shape analysis.
2:45 PM to 3:10 PM	Jin-Ting Zhang National University of Singapore	"NEW TESTS FOR EQUALITY OF SEVERAL COVARIANCE FUNCTIONS FOR FUNCTIONAL DATA" In this paper, we propose two new tests for the equality of the covariance functions of several functional populations, namely a quasi GPF test and a quasi Fmax test whose test statistics are obtained via Globalizing a Point-wise quasi F-test statistic with integration and taking its supremum over some time interval of interest, respectively. Unlike several existing tests, they are scale-invariant in the sense that their test statistics will not change if we multiply each of the observed functions by any non-zero function of time. We derive the asymptotic random expressions of the two tests under the null hypothesis and show that under some mild conditions, the asymptotic null distribution of the quasi GPF test is a chi-squared-type mixture whose distribution can be well approximated by a simple scaled chi-squared distribution. We also propose a random permutation method for approximating the null distributions of the quasi GPF and Fmax tests. The asymptotic distributions of the two tests under a local alternative are also investigated and the two tests are shown to be root-n consistent. A theoretical power comparison between the quasi GPF test and the L2-norm based test proposed in the literature is also given. Simulation studies are presented to demonstrate the finite-sample performance of the new tests against five existing tests. An illustrative example is also presented. (This is a joint work with Drs. Jia Guo and Bu Zhou)
IP09	**Recent Developments in Bayesian Statistics, Wednesday, 27 June 2018, 8:30 AM to 10:10 AM, ⑦ LT31**	
Org / Chair	Aurore Delaigle, University of Melbourne / ,	
8:30 AM to 8:55 AM	Howard Bondell University of Melbourne	"BAYESIAN REGRESSION FOR HIGH-DIMENSIONAL DATA USING A PRIOR ON THE MODEL FIT" We introduce a new class of prior distributions for linear regression, particularly the high dimensional case. Instead of placing a prior on the coefficients themselves, we place a prior on the regression R-squared. This is then distributed to the coefficients conditional on the value of R-squared. In addition to a convenient interpretation, compared to existing shrinkage priors, we show that the use of this prior can provide a higher degree of shrinkage on the irrelevant coefficients, along with less bias in estimation of the larger signals.
8:55 AM to 9:20 AM	John Ormerod University of Sydney	"A PARTICLE BASED COLLAPSED VARIATIONAL APPROXIMATION FOR BAYESIAN LINEAR MODEL AVERAGING" Bayesian model averaging has several desirable properties, it is computationally expensive unless the number of models to be averaged over is small. Typically the number of models to be averaged grows exponentially in the number of covariates and some form of approximation is required. In this paper we explore a novel particle based collapsed variational approximation for Bayesian model averaging. The resulting objective function can be optimized in a highly parallel manner. We explore several different prior specifications which lead to Bayes factors with closed forms. We show empirically that our approach is fast and effective for moderately large problems on several simulated and publicly available datasets, particularly when parallel computing resources are available. For example we fit a n=600 and p=7200 problem using 100 particles in around 8 seconds using 20 cores. An R package is available implementing our approach.
9:20 AM to 9:45 AM		#N/A
9:45 AM to 10:10 AM		#N/A
IP10	**Statistical Methods for Functional Data, Thursday, 28 June 2018, 10:30 AM to 12:10 PM, ⑤ LT33**	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
Org / Chair	Subhajit Dutta, Indian Institute of Technology / ,	
10:30 AM to 10:55 AM	Karthik Bharath University of Nottingham	"SAMPLING OF WARP MAPS FOR CURVE ALIGNMENT" Alignment of functional or curve data in the presence of important landmark information is integral to analysis of biomedical data. I will discuss a sampling scheme for warp maps used in alignment of open and closed curves, possibly with landmark constraints. The scheme provides a point process-based constructive definition of a probability measure on the set of warp maps of [0, 1] and the unit circle. The measure is used (i) as a proposal distribution in a stochastic algorithm to solve a variational formulation of curve alignment, and (ii) as a prior on warp maps in a Bayesian model for alignment.
10:55 AM to 11:20 AM	Joydeep Chowdhury Indian Statistical Institute	"ON NONPARAMETRIC REGRESSION USING DATA DEPTH" We investigate nonparametric regression methods based on statistical depth functions. These nonparametric regression procedures can be used in situations, where the response is multivariate and the covariate is a random element in a metric space. This includes regression with functional covariate as a special case. Our objective is to study different features of the conditional distribution of the response given the covariate. We construct measures of the center and the spread of the conditional distribution using depth based nonparametric regression procedures. We establish the asymptotic consistency of those measures and develop a test for heteroscedasticity based on the measure of conditional spread. The usefulness of the methodology is demonstrated in some real datasets.
11:20 AM to 11:45 AM		#N/A
11:45 AM to 12:10 PM		#N/A
IP11	**Recent Advances in Change-Point Detection and Data Segmentation, Friday, 29 June 2018, 8:30 AM to 10:10 AM, ④ LT26**	
Org / Chair	Piotr Fryzlewicz, London School of Economics / Piotr Fryzlewicz, London School of Economics	
8:30 AM to 8:55 AM	Hock Peng Chan National University of Singapore	"THE LOCAL TEST STATISTIC APPROACH IN MULTI-SEQUENCE SEGMENTATION" Niu and Zhang (2012) and Fang, Li and Siegmund (2016) developed the local test statistic approach for the segmentation of a single sequence. We first refine their approach to improve detection sensitivity, and show that the local approach is able to achieve a sensitivity lower bound up to 2 times less than the of the SMUCE algorithm. We then show how the local test statistic approach can be extended to multi-sequence segmentation, and apply it on a genetic dataset.
8:55 AM to 9:20 AM	Hao Chen University of California, Davis	"CHANGE-POINT DETECTION FOR LOCALLY DEPENDENT DATA" Local dependence is common in multivariate and non-Euclidean data sequences, such as network data. We consider the testing and estimation of change-points in such sequences. A new way of permutation, circular block permutation with a randomized starting point, is proposed and studied for a scan statistic utilizing graphs representing the similarity between observations. The proposed permutation approach could correctly address for local dependence and make it possible the theoretical treatments for the non-parametric graph-based scan statistic for locally dependent data. We derive accurate analytic approximations to the significance of graph-based scan statistics under the circular block permutation framework, facilitating its application to locally dependent multivariate or object data sequences.
9:20 AM to 9:45 AM	Piotr Fryzlewicz London School of Economics	"MULTISCALE METHODS, RECURSION AND DATA-ADAPTIVE CHANGE-POINT DETECTION" The talk starts on a general note: we first attempt to define a "multiscale" method / algorithm as a recursive program acting on a dataset in a suitable way. Wavelet transformations, unbalanced wavelet transformations and binary segmentation are all examples of multiscale methods in this sense. Using the example of binary segmentation, we illustrate the benefits of the recursive formulation of multiscale algorithms from the software implementation and theoretical tractability viewpoints. We then turn more specific and study the canonical problem of a-posteriori detection of multiple change-points in the mean of a piecewise-constant signal observed with noise. Even in this simple set-up, many publicly available state-of-the-art methods struggle for certain classes of signals. In particular, this misperformance is observed in methods that work by minimising a "fit to the data plus a penalty" criterion, the reason being that it is challenging to think of a penalty that works well over a wide range of signal classes. To overcome this issue, we propose a new approach whereby methods learn from the data as they proceed, and, as a result, operate differently for different signal classes. As an example of this approach, we revisit our earlier change-point detection algorithm, Wild Binary Segmentation, and make it data-adaptive by equipping it with a recursive mechanism for deciding "on the fly" how many sub-samples of the input data to draw, and w here to draw them. This is in contrast to the original Wild Binary Segmentation, which is not recursive. We show that this significantly improves the algorithm particularly for signals with frequent change-points.
9:45 AM to 10:10 AM	Guenther Walther Stanford University	"THE ESSENTIAL HISTOGRAM" The aim of a histogram is to provide a simple visualization of the empirical distribution. One way to formalize this task is as follows: Construct the simplest histogram (i.e. the histogram with the fewest jumps) that provides a good estimate of the empirical distribution (i.e. probabilities are estimated optimally over all intervals). We present results towards the construction of such an `essential histogram'. The procedure is based on distribution-free likelihood ratios defined on intervals from a certain collection of intervals. This collection is constructed to allow efficient computation as well as statistically optimal inference. Joint work with Housen Li, Axel Munk, and Hannes Sieling.
IP12	**Infinite-Dimensional Stochastic Analysis and its Applications, Thursday, 28 June 2018, 10:30 AM to 12:10 PM, ⑧ MD9-0102GH**	
Org / Chair	Arnab Ganguly, Louisiana State University / Arnab Ganguly, Louisiana State University	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
10:30 AM to 10:55 AM	Arnab Ganguly Louisiana State University	"APPROXIMATION OF INVARIANT DISTRIBUTIONS OF ERGODIC DIFFUSIONS: CLT AND MODERATE DEVIATIONS" Estimation of functionals of invariant distribution of ergodic diffusions is computationally a hard problem. Solving the stationary Kolmogorov forward equation, even numerically, to get the invariant distribution is computationally expensive for most stochastic models. A probabilistic approach is to use ergodic averages together with an Euler-Maruyama numerical scheme. But choosing a fixed discretization step for such a scheme, as is often done in practice, could lead to infinite error for these approximation of invariant distribution problems! In this talk, we will consider an Euler discretization based numerical scheme for this problem with appropriately scaled discretization steps and discuss the corresponding error analysis. The main highlights will be a central limit theorem and moderate deviation asymptotics.
10:55 AM to 11:20 AM	Yaozhong Hu University of Alberta at Edmonton	#N/A
11:20 AM to 11:45 AM	Janna Lierl University of Connecticut	"NEUMANN HEAT FLOW AND GRADIENT FLOW FOR THE ENTROPY ON NON-CONVEX DOMAINS" For large classes of non-convex subsets Y in R^n or in Riemannian manifolds (M,g) or in RCD-spaces (X,d,m) we prove that the gradient flow for the Boltzmann entropy on the restricted metric measure space (Y,d_Y,m_Y) exists - despite the fact that the entropy is not semiconvex - and coincides with the heat flow on Y with Neumann boundary conditions.
11:45 AM to 12:10 PM	Ruoyu Wu University of North Carolina at Chapel Hill	"LOCAL CHARACTERIZATION OF DYNAMICS ON LARGE SPARSE GRAPHS" We study large systems of interacting particles in which each particle is associated with a vertex in a graph and interacts only with (the empirical measure of) its neighbors. The case of complete graph falls under the purview of classical mean-field limits, and it is well known that (under suitable assumptions) the dynamics of a typical particle is governed by a nonlinear Markov process. In this talk, we consider the complementary sparse case when the underlying graph converges in a suitable sense to a countably infinite locally finite graph G, and describe various limit results, both in the setting of diffusions and Markov chains. In particular, when G is a d-regular tree, we obtain an autonomous characterization of the local dynamics of the neighborhood of a typical node. We also obtain a local characterization for the annealed dynamics on a class of Galton-Watson trees. The proofs rely on a certain Markov random field structure of the dynamics on the countably infinite graph G, which may be of independent interest. This is joint work with Daniel Lacker and Kavita Ramanan.
IP13	**Analysis of Complex Data , Tuesday, 26 June 2018, 1:30 PM to 3:10 PM, ④ LT26**	
Org / Chair	Xuming He, University of Michigan / Xuming He, University of Michigan	
1:30 PM to 1:55 PM	Jian Huang University of Iowa	#N/A
1:55 PM to 2:20 PM	Weijie Su University of Pennsylvania	"STATISTICAL INFERENCE FOR ONLINE LEARNING AND STOCHASTIC APPROXIMATION VIA HIERARCHICAL INCREMENTAL" Stochastic gradient descent (SGD) is an immensely popular approach for online learning in settings where data arrives in a stream or data sizes are very large. However, despite an ever-increasing volume of work on SGD, much less is known about the statistical inferential properties of SGD-based predictions. Taking a fully inferential viewpoint, this paper introduces a novel procedure termed HiGrad to conduct statistical inference for online learning, without incurring additional computational cost compared with SGD. The HiGrad procedure begins by performing SGD updates for a while and then splits the single thread into several threads, and this procedure hierarchically operates in this fashion along each thread. With predictions provided by multiple threads in place, a t-based confidence interval is constructed by decorrelating predictions using covariance structures given by the Ruppert–Polyak averaging scheme. Under certain regularity conditions, the HiGrad confidence interval is shown to attain asymptotically exact coverage probability. Finally, the performance of HiGrad is evaluated through extensive simulation studies and a real data example. An R package higrad has been developed to implement the method.
2:20 PM to 2:45 PM	Ying Sun King Abdullah University of Science and Technology	"EFFICIENT ESTIMATION FOR NON-STATIONARY SPATIAL COVARIANCE FUNCTIONS" Spatial processes exhibit non-stationarity in many climate and environmental applications. Convolution-based approaches are often used to construct non-stationary covariance functions in Gaussian processes. Although convolution-based models are flexible, their computation is extremely expensive when the dataset is large. Most existing methods rely on fitting an anisotropic but stationary model locally and reconstructing the spatially varying parameters. In this study, we propose a new estimation procedure to approximate a class of non-stationary Matern covariance functions by the local-polynomial fitting of the covariance parameters. The proposed method allows for efficient estimation of a richer class of non-stationary covariance functions with the local stationary model as a special case. We also develop an approach for fast high-resolution simulation with non-stationary features on the small scale and apply it to precipitation data in the climate model.
2:45 PM to 3:10 PM	Shurong Zheng Northeast Normal University	"GLOBAL TESTING FOR HIGH-DIMENSIONAL CORRELATION MATRICES" Testing correlation structures has attracted extensive attention in the literature due to both its importance in real applications and several major theoretical challenges. The aim of this paper is to develop a general framework of testing correlation structures for the one-, two-, and multiple sample testing problems under a high dimensional setting when both the sample size and data dimension go to infinity. Our global test statistics are designed to deal with both the dense and sparse alternatives. We systematically investigate the asymptotic null distribution, power function, and unbiasedness of each global test statistic. Theoretically, we make great efforts to deal with the non-independency of all random matrices of the sample correlation matrices. We use simulation studies and real data analysis to illustrate the versatility and practicability of our global test statistics.
IP14	**Stochastic Partial Differential Equations, Wednesday, 27 June 2018, 10:30 AM to 12:10 PM, ⑧ MD9-0102GH**	
Org / Chair	Yaozhong Hu, University of Alberta at Edmonton / Yaozhong Hu, University of Alberta at Edmonton	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
10:30 AM to 10:55 AM	Jian Song Hong Kong University	"TO BE DONE" To be done
10:55 AM to 11:20 AM	Xiaoming Song Drexel University	"NONLINEAR FEYNMAN-KAC FORMULAE FOR SPDES WITH SPACE-TIME NOISE" We study a class of backward doubly stochastic differential equations (BDSDEs) involving martingales with spatial parameters, and show that they provide probabilistic interpretations (Feynman-Kac formulae) for certain semilinear stochastic partial differential equations (SPDEs) with space-time noise. As an application of the Feynman-Kac formulae, random periodic solutions and stationary solutions to certain SPDEs are obtained.
11:20 AM to 11:45 AM	Jie Xiong Southern University of Science and Technology	"LARGE DEVIATIONS FOR SPDES WITH LOCALLY MONOTONE COEFFICIENTS DRIVEN BY LÉVY NOISE" We establish a large deviation principle for a type of stochastic partial differential equations (SPDEs) with locally monotone coefficients driven by Lévy noise. The weak convergence method plays an important role. This talk is based on a joint paper with Jianliang Zhai.
11:45 AM to 12:10 PM	Xicheng Zhang Wuhan University	"STRONG UNIQUENESS OF DEGENERATE SDES WITH HÖLDER DIFFUSION COEFFICIENTS" In this paper we prove a new strong uniqueness result and a weak existence result for possibly degenerate multidimensional stochastic differential equations with Sobolev diffusion coefficients and rough drifts. In particular, examples with Holder diffusion coefficients are provided to show our results.
IP15	**Recent Advancements in Spatial and Spatio-Temporal Statistics, Thursday, 28 June 2018, 8:30 AM to 10:10 AM, ⑤ LT33**	
Org / Chair	Hsin Cheng Huang, Academia Sinica / Hsin Cheng Huang, Academia Sinica	
8:30 AM to 8:55 AM	Chun Shu Chen National Changhua University of Education	"ON FIXED EFFECTS ESTIMATION FOR SPATIAL REGRESSION WITH SPATIAL CONFOUNDING" Spatial regression models are often used to analyze the ecological and environmental data sets over a continuous spatial support. Issues of collinearity among covariates are often considered in modeling, but only rarely in discussing the relationship between covariates and unobserved spatial random processes. Past researches have shown that ignoring this relationship (or, spatial confounding) would have significant influences on the estimation of regression parameters. To improve this problem, an idea of restricted spatial regression is used to ensure that the unobserved spatial random process is orthogonal to covariates, but the related inferences are mainly based on Bayesian frameworks. In this paper, an adjusted generalized least squares estimation method is proposed to estimate regression coefficients, resulting in estimators that perform better than conventional methods. Statistical inferences of the proposed methodology are justified both theoretically and numerically.
8:55 AM to 9:20 AM	Mikyoung Jun Texas A&M University	"SPATIO-TEMPORAL SHORT-TERM WIND FORECAST: A PROACTIVE REGIME-SWITCHING METHOD" Accurate short-term forecasts are indispensable for harnessing wind energy reliably. In the past decade, models that produce regime-dependent forecasts have received considerable attention. Most of these methods are "reactive," meaning that they react to observed changes, rather than anticipate upcoming regime changes, thus potentially undermining their forecasting ability. In light of this, this paper advocates a more "proactive" regime-switching approach, as a transition from the reactive "change detection" to a proactive "change anticipation" paradigm. The essence of our approach is to calibrate forecasts obtained via a base model to safeguard against out-of-sample regime changes. We model this forecast calibration by means of two elements: the observed wind regime and the runlength, which is the time elapsed since the most recent regime change. Concurrently, a monitoring mechanism for the most recent history of wind data ensures that in-sample changes are equally accounted for. Our approach further allows the incorporation of well-known features of wind fields: spatio-temporal correlations, ow-dependent asymmetries and non-stationarity, in addition to regime-switching dynamics. Using one year of turbine-specic wind farm data, we show that the proactive regime-switching approach can offer a wide margin of improvement over existing forecasting methods in terms of both wind speed and power predictions. This is joint work with Ahmed Aziz Ezzat and Yu Ding.
9:20 AM to 9:45 AM	Haonan Wang Colorado State University	"LOCALLY STATIONARY SPATIO-TEMPORAL MODELING: ESTIMATION AND INFERENCE" We consider the problem of estimation and inference for spatiotemporal models under local stationarity. We develop a novel spatio-temporal distance expanding (STDE) asymptotic framework on a fixed domain, which can be used to study the properties of statistical inference procedures for spatio-temporal models. We further study several models with different parametric and semiparametric mean structures. Parameter estimation is carried out using maximum likelihood method or profile maximum likelihood method. Theoretical properties of the resulting estimates, including consistency and asymptotic normality, are established under our proposed asymptotic framework. In addition, for partially linear models, we use the bimodal kernel to mitigate the effect of correlated errors. We demonstrate the finite sample properties of our proposed estimation method through simulations and a real data example.
9:45 AM to 10:10 AM	Christopher Wikle University of Missouri	"MULTI TIME-SCALE SPATIO-TEMPORAL DYNAMIC STATISTICAL MODELS MOTIVATED BY MACHINE LEARNING" Spatio-temporal data are ubiquitous in engineering and the sciences, and their study is important for understanding and predicting a wide variety of processes. One of the chief difficulties in modeling spatial processes that change with time is the complexity of the dependence structures that must describe how such a process varies, and the presence of high-dimensional complex datasets and large prediction domains. It is particularly challenging to specify parameterizations for nonlinear dynamical spatio-temporal models that are simultaneously useful scientifically and efficient computationally. Statisticians have developed some "deep" mechanistically-motivated models that can accommodate process complexity as well as the uncertainties in the predictions and inference. However, these models can be expensive and are typically application specific. On the other hand, the science, engineering, and machine learning communities have developed alternative approaches for nonlinear spatio-temporal modeling, in some cases with fairly parsimonious parameterizations. These approaches can be quite flexible and sometimes can be implemented quite efficiently, but typically without formal uncertainty quantification. Here, we present a multi time-scale spatio-temporal dynamical model that places a special parsimonious class of recurrent neural networks in a statistical framework that can account for uncertainty. This is illustrated on a multi-scale process related to long lead-time forecasting of atmospheric events given ocean conditions.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
IP16	**Advances in Regression and Classification, Wednesday, 27 June 2018, 1:30 PM to 3:10 PM, ⑧ MD9-0102GH**	
Org / Chair	Binyan Jiang, Hong Kong Polytechnic University / Binyan Jiang, Hong Kong Polytechnic University	
1:30 PM to 1:55 PM	Xin Guo Hong Kong Polytechnic University	"CENTERED REPRODUCING KERNELS AND THEIR APPLICATIONS" We study the centering transformation of reproducing kernels, based on either a probability distribution, or the discrete uniform distribution on a sample of observations. The obtained kernel is also a reproducing kernel, with the associated reproducing kernel Hilbert space lying in the orthogonal complement of constant functions. We proved that the kernel space complexity in terms of effective dimensions is an asymptotic invariant to this transformation. We also obtained some relations between the integral operators before and after the kernel transformation. We applied the centered kernels to the regularized least squares scheme and the constant component is naturally separated from the regression function. We also explore the applications of centered kernels to the coordinate kernel polynomial models for variable selection.
1:55 PM to 2:20 PM	Limin Peng Emory University	"TRAJECTORY QUANTILE REGRESSION FOR LONGITUDINAL DATA" Quantile regression has demonstrated promising utility in longitudinal data analysis. Existing work is primarily focused on modeling cross-sectional outcomes, while outcome trajectories often carry more substantive information in practice. In this work, we develop a trajectory quantile regression framework that is designed to robustly and flexibly investigate how latent individual trajectory features are related to observed subject characteristics. The proposed models are built under multilevel modeling with usual parametric assumptions lifted or relaxed. We derive our estimation procedure by novelly transforming the problem at hand to quantile regression with perturbed responses and adapting the bias correction technique for handling covariate measurement errors. We establish desirable asymptotic properties of the proposed estimator, including uniform consistency and weak convergence. Extensive simulation studies confirm the validity of the proposed method as well as its robustness. An application to the DURABLE trial uncovers sensible scientific findings and illustrates the practical value of our proposals.
2:20 PM to 2:45 PM	Zhisheng Ye National University of Singapore	"SHAPE CONSTRAINTS ON THE COVARIATE EFFECTS IN STOCHASTIC DEGRADATION PROCESSES" Effective oxidation-based elimination of emerging contaminants (ECs) requires a good understanding of the effects of treatment conditions, such as the kinds and dosages of reagents, on the EC degradation rate. Due to limited knowledge on the complex reaction mechanism and the multiple covariates to represent the treatment conditions, it is generally hard to parametrically quantify the relation between these covariates and the degradation rate. On the other hand, qualitative analysis based on chemical mechanisms often provides shape information of the covariate-rate relation, such as monotonicity and local concavity in each coordinate. Based on the chemical kinetics, we use stationary stochastic processes for parametric modeling of log-transformed EC degradation under each combination of the treatment conditions. The tensor product Bernstein bases are then used to approximate the covariate-rate relation. The shape information is naturally translated to constraints on the coefficients of the bases functions. Likelihood-based inference procedures are developed for both point and interval estimation in the proposed models. Simulation results show that the use of shape information significantly improves the accuracy of estimates. The proposed method is successfully applied to EC degradation data collected from real experiments.
2:45 PM to 3:10 PM	Teng Zhang University of Central Florida	"FLEXIBLE EXPECTILE REGRESSION IN REPRODUCING KERNEL HILBERT SPACE" Expectile, first introduced by Newey and Powell (1987) in the econometrics literature, has recently become increasingly popular in risk management and capital allocation for financial institutions due to its desirable properties such as coherence and elicitability. The current standard tool for expectile regression analysis is the multiple linear expectile regression proposed by Newey and Powell in 1987. The growing applications of expectile regression motivate us to develop a much more flexible nonparametric multiple expectile regression in a reproducing kernel Hilbert space. The resulting estimator is called KERE which has multiple advantages over the classical multiple linear expectile regression by incorporating non-linearity, non-additivity and complex interactions in the final estimator. The kernel learning theory of KERE is established. We develop an efficient algorithm inspired by majorization-minimization principle for solving the entire solution path of KERE. It is shown that the algorithm converges at least at a linear rate. Extensive simulations are conducted to show the very competitive finite sample performance of KERE. We further demonstrate the application of KERE by using personal computer price data.
IP17	**Statistical Causal Inference, Tuesday, 26 June 2018, 3:30 PM to 5:10 PM, ⑥ LT34**	
Org / Chair	Manabu Kuroki, Yokohama National University / Manabu Iwasaki, Yokohama City University	
3:30 PM to 3:55 PM	Hei Chan National Institute of Informatics	"LATEST RESULTS ON THE CAUSAL IDENTIFICATION PROBLEM IN LINEAR SEMS" Structural Equation Models (SEM) is a widely used tool for causal analysis. In a linear SEM, the relationships between observed variables are expressed in linear equations. The structure of the equations is such that they not only express the linear relationships between the variables, together with a stochastic error term for unobserved factors, but also the causal dependence among the observed variables, which can be graphically represented by a causal diagram. In the causal diagram, which is a directed acyclic graph, directed edges represent direct causal effects between two variables, while bi-directed edges represent correlations of error terms between two variables. The causal identification problem in a linear SEM is to estimate the strength of a causal effect from one variable to another variable (direct effect, total effect, or effect conditional on other variables), from the combination of observed data and the causal diagram of the SEM. The two main methods for solving this problem are algebraic methods (such as Grobner basis computations) which attempt to solve the set of simultaneous equations based on Wright's method of path analysis, and graphical methods (such as back-door criterion, instrumental variable, and path-specific instrumental variable) which check if a causal effect is identifiable by testing for certain graphical criteria in the causal diagram. We discuss the latest results of solving the causal identification problem in linear SEMs, and introduce the Instrumental Variable Function, which is computed from the covariances and causal effects between variables. We then show the correspondence of the Instrumental Variable Function to both algebraic and graphical methods, and its usefulness in understanding the causal identification problem, and apply it to solve some difficult cases. This paper is joint research with Manabu Kuroki of Yokohama National University.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
3:55 PM to 4:20 PM	Shohei Shimizu Shiga University	"CAUSAL DISCOVERY, PREDICTION MECHANISMS, AND CONTROL" Causal structure learning is one of the most exciting topics in the fields of machine learning and statistics. Several existing methods have been shown to consistently estimate causal direction assuming linear or some form of nonlinear relationship and no hidden common causes. However, the estimation results could be distorted if either assumption is violated. We develop an approach to estimating the possible causal direction between two observed variables when hidden common causes are present. We propose a linear non-Gaussian acyclic structural equation model with individual-specific effects that are sometimes the source of confounding. Thus, modeling individual-specific effects as latent variables allows hidden common causes to be considered. We also consider connecting the underlying causal structure of a data generation process and the causal structure of a prediction mechanism by proposing a framework that estimates the necessary intervention of a feature such that a desired prediction is obtained.
4:20 PM to 4:45 PM	Miao Wang Peking University	#N/A
4:45 PM to 5:10 PM	Jiji Zhang Lingnan University	"OCCAM'S RAZORS IN THE STATISTICAL INFERENCE OF MARKOVIAN CAUSAL MODELS" The causal Faithfulness assumption is widely adopted in the inference of Markovian causal models and is often treated as expressing a methodological preference for simplicity. Viewed this way, it is useful to treat the assumption as a conjunction of a purely statistical assumption and a causal assumption of minimality. In this talk, I compare a few formulations of the minimality component of the causal Faithfulness assumption, and present a novel learning theoretic vindication of such Occam's razors for causal inference.
IP18	**New Developments in Nonparametric and High-Dimensional Bayesian Models, Wednesday, 27 June 2018, 3:30 PM to 5:10 PM, ③ LT29**	
Org / Chair	Jaeyong Lee, Seoul National University / Hee-Seok Oh, Seoul National University	
3:30 PM to 3:55 PM	Woncheol Jang Seoul National University	"ESTIMATIONS AND GENERALIZATIONS OF BIC" We present a new approach to Bayes factors based on Laplace expansions (as BIC) which we call GBIC (Generalized BIC), borrowing the name from the approximation derived in Berger, Ghosh and Mukhopadhyay (2003) with which it is related. In our approach, we do not include the prior in the Laplace expansion, but choose it appropriately so that it produces close-form expressions for the resulting GBIC. We explore both joint priors [to be done] and independent priors for the component parameters. To help choose the scale of the prior, we use a novel definition of selective sample size which allows for different selective sample sizes for the parameters. The new GBIC avoids many of the difficulties commonly associated with BIC, and can often be shown to be consistent. We also produce a modified GBIC which is more favorable to complex models while still retaining consistency.
3:55 PM to 4:20 PM	Cheolwoo Park University of Georgia	"REGULARIZED AGGREGATION OF STATISTICAL PARAMETRIC MAPS" Combining statistical parametric maps (SPM) from individual subjects is the goal in some types of group-level analyses of functional magnetic resonance imaging (fMRI) data. Brain maps are usually combined using a simple average across subjects, making them susceptible to subjects with outlying values. Furthermore, t tests are prone to false positives and false negatives when outlying values are observed. We propose a regularized unsupervised aggregation method for SPMs to find an optimal weight for aggregation, which aids in detecting and mitigating the effect of outlying subjects. We also present a bootstrap-based weighted t test using the optimal weights to construct an activation map robust to outlying subjects. We validate the performance of the proposed aggregation method and test using simulated and real data examples. Results show that the regularized aggregation approach can effectively detect outlying subjects, lower their weights, and produce robust SPMs.
4:20 PM to 4:45 PM	Shuhei Mano Institute of Statistical Mathematics	"SAMPLERS FROM NON-EXCHANGEABLE PRIOR PROCESSES AND HYPERGEOMETRIC SYSTEMS" A Prior process in Bayesian nonparametrics is characterized by an exchangeable partition probability function (EPPF), which is a sample from prior process. Construction of samplers from a prior process is an important block in an implementation of nonparametric Bayesian analyses. Familiar examples are Polya-like urn schemes and reversible jump MCMC. A Polya-like urn scheme achieves direct sampling, but demands infinite exchangeability of EPPF. In this talk, a novel direct sampler from general Gibbs-type partitions, where we do not assume infinite exchangeability, is presented. The sampler is based on the observation that Gibbs-type partitions have close relationship with an hypergeometric system.
4:45 PM to 5:10 PM		#N/A
IP20	**New Frontiers of Functional and High-Dimensional Data Analysis, Thursday, 28 June 2018, 10:30 AM to 12:10 PM, ⑥ LT34**	
Org / Chair	Yehua Li, University of California, Riverside / Yehua Li, University of California, Riverside	
10:30 AM to 10:55 AM	Lily Wang Iowa State University	"ESTIMATION AND INFERENCE FOR IMAGE-ON-SCALAR REGRESSION" Motivated by recent work analyzing imaging data in the neuroimaging studies, we consider a class of linear functional response regression models for imaging responses and scalar predictors. We propose to use bivariate splines over triangulation to handle the possibly irregular domain of the active regions, and introduce a regularization method for local-region sparse estimation. Our method can accurately identify the sparse active regions and produce a smooth estimator on the active regions. In addition, it can significantly reduce the variability of estimation and improve the interpretability of the relationship between the response and the predictors. The proposed spline estimators of the coefficient functions are proved to be asymptotically normal under some regularity conditions. Asymptotic and data-driven simultaneous confidence corridors for the coefficient functions are constructed. Highly efficient and scalable estimation algorithm is developed. Monte Carlo simulation studies are conducted to examine the finite-sample performance of the proposed method. The proposed method is applied to the spatially normalized FDG-PET data of Alzheimer's Disease Neuroimaging Initiative (ADNI).

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
10:55 AM to 11:20 AM	Raymond Wong Texas A&M University	"PARTIALLY LINEAR FUNCTIONAL ADDITIVE MODELS FOR MULTIVARIATE FUNCTIONAL DATA" We investigate a class of partially linear functional additive models (PLFAM) that predicts a scalar response by both parametric effects of a multivariate predictor and nonparametric effects of a multivariate functional predictor. We jointly model multiple functional predictors that are cross-correlated using multivariate functional principal component analysis (mFPCA), and model the nonparametric effects of the principal component scores as additive components in the PLFAM. To address the high dimensional nature of functional data, we let the number of mFPCA components diverge to infinity with the sample size, and adopt the COmponent Selection and Smoothing Operator (COSSO) penalty to select relevant components and regularize the fitting. A fundamental difference between our framework and the existing high dimensional additive models is that the mFPCA scores are estimated with error, and the magnitude of measurement error increases with the order of mFPCA. We establish the asymptotic convergence rate for our estimator, while allowing the number of components diverge. When the number of additive components is fixed, we also establish the asymptotic distribution for the partially linear coefficients. The practical performance of the proposed methods is illustrated via simulation studies and a crop yield prediction application.
11:20 AM to 11:45 AM	Fang Yao University of Toronto and Peking University	#N/A
11:45 AM to 12:10 PM		#N/A
IP21	**Integrative Analysis of Multi-Omics Data , Wednesday, 27 June 2018, 8:30 AM to 10:10 AM, ④ LT26**	
Org / Chair	Gen Li, Columbia University / Irina Gaynanova, Texas A&M University	
8:30 AM to 8:55 AM	Kuangnan Fang Xiamen University	"INTEGRATIVE SPARSE PRINCIPAL COMPONENT ANALYSIS" In the analysis of data with high-dimensional covariates and small sample sizes, dimension reduction techniques have been extensively employed. Principal component analysis (PCA) is perhaps the most popular dimension reduction techniques. To effectively remove noises and generate more interpretable results, the sparse PCA (SPCA) technique has been developed. With the high dimensionality, the analysis of a single dataset often generates unsatisfactory results. In a series of studies under the "regression analysis + variable selection" setting, it has been shown that integrative analysis provides an effective way of pooling information from multiple independent datasets and outperforms single-dataset analysis and many alternative multi-datasets analyses, especially including the classic meta-analysis. In this study, with multiple independent datasets, we propose conducting dimension reduction using a novel iSPCA (integrative SPCA) approach. Penalization is adopted for regularized estimation and selection of important loadings. Advancing from the existing integrative analysis studies, we further impose contrasted penalties, which may generate more accurate estimation/selection. Multiple settings on the similarity across datasets are comprehensively considered. Consistency properties of the proposed approach are established, and effective computational algorithms are developed. A wide spectrum of simulations demonstrate competitive performance of iSPCA over the alternatives. Two sets of data analysis further establish its practical applicability.
8:55 AM to 9:20 AM	Irina Gaynanova Texas A&M University	"STRUCTURAL LEARNING AND INTEGRATIVE DECOMPOSITION OF MULTI-VIEW DATA" The increased availability of the multi-view data (data on the same samples from multiple sources) has led to strong interest in models based on low-rank matrix factorizations. These models represent each data view via shared and individual components, and have been successfully applied for exploratory dimension reduction, association analysis between the views, and further learning tasks such as consensus clustering. Despite these advances, there remain significant challenges in modeling partially-shared components, and identifying the number of components of each type (shared/partially-shared/individual). In this work, we formulate a novel linked component model that directly incorporates partially-shared structures. We call this model SLIDE for Structural Learning and Integrative DEcomposition of multi-view data. We prove the existence of SLIDE decomposition and explicitly characterize the identifiability conditions. The proposed model fitting and selection techniques allow for joint identification of the number of components of each type, in contrast to existing sequential approaches. In our empirical studies, SLIDE demonstrates excellent performance in both signal estimation and component selection. We further illustrate the methodology on the breast cancer data from The Cancer Genome Atlas repository. This is joint work with Gen Li.
9:20 AM to 9:45 AM	Qi Long University of Pennsylvania	"GENERALIZED BI-CLUSTERING ANALYSIS FOR INTEGRATIVE Â€“OMICS ANALYSIS USING BIOLOGICAL INFORMATION" Advances in technology have enabled generation of multiple types of -omics data in many biomedical and clinical studies, and it is desirable to pool such data in order to improve the power of identifying important molecular signatures and patterns. However, such integrative analyses present new analytical and computational challenges. To address some of these challenges, we propose a sparse Bayesian generalized bi-clustering analysis (GBC) which enables integrating multiple omics modalities with incorporation of biological knowledge through the use of adaptive structured shrinkage priors. The proposed methods can accommodate both continuous and discrete data. MCMC and EM algorithms are developed for estimation. Numerical studies are conducted to demonstrate that our methods achieve improved feature selection and prediction in identifying disease subtypes and latent drivers, compared to existing methods.
9:45 AM to 10:10 AM		#N/A
IP22	**Frontiers in Financial Statistics, Wednesday, 27 June 2018, 1:30 PM to 3:10 PM, ④ LT26**	
Org / Chair	Yingying Li, Hong Kong University of Science and Technology / Mengmeng Ao, Xiamen University	
1:30 PM to 1:55 PM	Raymond Kan University of Toronto	#N/A

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
1:55 PM to 2:20 PM	Olivier Ledoit University of Zürich	"EFFICIENT SORTING: A MORE POWERFUL TEST FOR CROSS-SECTIONAL ANOMALIES" Many researchers seek factors that predict the cross-section of stock returns. The standard methodology sorts stocks according to their factor scores into quantiles and forms a corresponding long-short portfolio. Such a course of action ignores any information on the covariance matrix of stock returns. Historically, it has been difficult to estimate the covariance matrix for a large universe of stocks. We demonstrate that using the recent DCC-NL estimator of Engle et al. (2017) substantially enhances the power of tests for cross-sectional anomalies: On average, `Student' t-statistics more than double.
2:20 PM to 2:45 PM	Yuan Liao Rutgers University	#N/A
2:45 PM to 3:10 PM	Xinghua Zheng Hong Kong University of Science and Technology	"APPROACHING MEAN-VARIANCE EFFICIENCY FOR LARGE PORTFOLIOS" We study the large dimensional Markowitz optimization problem. Given any risk constraint level, we introduce a new approach for estimating the optimal portfolio. The approach relies on a novel unconstrained regression representation of the mean-variance optimization problem, combined with high-dimensional sparse regression methods. Our estimated portfolio, under a mild sparsity assumption, asymptotically achieves mean-variance efficiency and meanwhile effectively controls the risk. To the best of our knowledge, this is the first approach that can achieve these two goals simultaneously for large portfolios. The superior properties of our approach are demonstrated via comprehensive simulation and empirical analysis. Based on joint work with Mengmeng Ao and Xinghua Zheng.
IP23	**Advanced Statistical Methods for High-Dimensional Microbiome Data Analysis, Tuesday, 26 June 2018, 3:30 PM to 5:10 PM, ⑦ LT31**	
Org / Chair	Huilin Li, New York University / Huilin Li, New York University	
3:30 PM to 3:55 PM	Lingling An University of Arizona	"MEDIATION ANALYSIS IN INVESTIGATING THE ROLE OF MICROBIOTA IN HUMAN HEALTH" In this research, we propose to utilize new advances in information science to select mediation mechanisms in an integrated human gene expression and microbial abundance dataset to predict clinical outcomes of diseases. This method utilizes the concept of contributed information, information penalized for redundancy with other selected variables, and generates test statistics sequentially by feature using a greedy search algorithm. The efficacy of this approach is examined in a simulation study. The results of this preliminary simulation study show that this nonparametric approach has good performance for selecting important mediators.
3:55 PM to 4:20 PM	Richard Bonneau New York University	"MULTI-TASK APPROACHES TO LEARNING GENE-REGULATORY AND MICROBIAL INTERACTION NETWORKS FROM MULTIPLE L" Learning gene and microbial interaction networks both require large data-sets, yet much public data is distributed across many studies with unknown batch effect. Leveraging multiple sources of information, such as different publicly available gene expression datasets, could therefore be helpful when learning a network of interest. Integrating data across different studies, however, raises numerous technical concerns. Hence, a common approach in network inference, and broadly in genomics and microbiome research, is to separately learn models from each dataset and combine the results. Individual models, however, often suffer from under-sampling, poor generalization and limited network recovery. I will discuss a new multitask learning approach for joint network inference across several datasets. Our method initially estimates the activities of transcription factors, and subsequently, infers the relevant network topology. The method allows for adaptive penalties that can be used to favor models that include interactions derived from multiple sources of prior knowledge including orthogonal genomics experiments. Lastly we will discuss model formulations that allow our MTL-network inference approach to be adapted to learning microbial interaction networks from multiple data sets.
4:20 PM to 4:45 PM	Yijuan Hu Emory University	"TESTING PRESENCE-ABSENCE ASSOCIATION IN THE MICROBIOME USING THE LINEAR DECOMPOSITION MODEL (LDM)" There is now a proliferation of 16S studies thanks to precipitous drops in sequencing costs. Comparison of microbiome profiles with respect to differential clinical outcomes can reveal important knowledge towards understanding the effects of microbiome on diseases. The difference in microbiome profiles may be driven by the presence-absence status of certain microbes. Several distance measures were designed for quantifying presence-absence dissimilarities. While PERMANOVA can be used for testing the hypothesis that whether the global presence-absence pattern is associated with a covariate, it cannot be used for detecting individual OTUs whose presence-absence status is associated with the covariate. We recently developed a linear decomposition model, referred to as LDM, which can perform both global test of association and detection of associated OTUs within one framework. In this ongoing work, we extend LDM for testing presence-absence association in the microbiome. In particular, as the presence-absence association is strongly confounded by the variation in library size, we adopt rarefaction as the normalization method and exploit data through multiple rarefactions, and we show how to aggregate multiple rarefied datasets within the LDM framework. We evaluate the LDM via extensive simulation studies and application to a real microbiome dataset.
4:45 PM to 5:10 PM	Huilin Li New York University	"AN EFFICIENT TWO-STAGE MICROBIAL ASSOCIATION MAPPING FRAMEWORK" We develop four estimators, together with their implementation algorithms, for the FCGGM. We establish the consistency and the convergence rates of one of the estimators under different sets of sufficient conditions with varying strengths. We compare our FCGGM with the existing functional Gaussian graphical model by simulation, under both non-Gaussian and Gaussian graphical models, and apply our method to an EEG data set to construct brain networks.
IP24	**Advances in Methods and Applications for Dependent Data, Friday, 29 June 2018, 3:30 PM to 5:10 PM, ③ LT29**	
Org / Chair	Chae Young Lim, Seoul National University / Chae Young Lim, Seoul National University	
3:30 PM to 3:55 PM	Soutir Bandyopadhyay Colorado School of Mines	#N/A

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
3:55 PM to 4:20 PM	Chih Hao Chang National University of Kaohsiung	"ASYMPTOTIC THEORY OF CONDITIONAL GENERALIZED INFORMATION CRITERION FOR LINEAR MIXED EFFECTS MODEL SE" We consider selecting linear mixed-effects models for possibly unbalanced data, where both the fixed-effects and random-effects can be misspecified. We introduce a conditional generalized information criterion (CGIC) for linear mixed-effects models, which is extended from conditional Akaike's information criterion provided by Vaida and Blanchard (2005). We establish the consistency and the asymptotic loss efficiency (with respect to the Kullback-Leibler loss) of CGIC for the mixed-effects models, where the dimensions of both effects can go to infinity with the sample size. We also allow the number of clusters to be fixed, where the random-effects parameters cannot be consistently estimable, which further makes difficult to the development of asymptotic theorem of CGIC.
4:20 PM to 4:45 PM	Jungsoon Choi Hanyang University	"BAYESIAN SPATIALLY DEPENDENT VARIABLE SELECTION FOR SMALL AREA HEALTH MODELING" Statistical methods for spatial health data to identify the significant covariates associated with the health outcomes are of critical importance. Most studies have developed variable selection approaches in which the covariates included appear within the spatial domain and their effects are fixed across space. However, the impact of covariates on health outcomes may change across space and ignoring this behavior in spatial epidemiology may cause the wrong interpretation of the relations. Thus, the development of a statistical framework for spatial variable selection is important to allow for the estimation of the space-varying patterns of covariate effects as well as the early detection of disease over space. In this work, we develop flexible spatial variable selection approaches to find the spatially-varying subsets of covariates with significant effects. A Bayesian hierarchical latent model framework is applied to account for spatially-varying covariate effects. We present a simulation example to examine the performance of the proposed models with the competing models. We apply our models to a county-level low birth weight incidence dataset in Georgia.
4:45 PM to 5:10 PM	Ji Meng Loh New Jersey Institute of Technology	"SINGLE-INDEX MODEL FOR INHOMOGENEOUS SPATIAL POINT PROCESSES" I will introduce a single-index model for the intensity of an inhomogeneous spatial point process, relating the intensity function to an unknown function ρ of a linear combination of a p-dimensional spatial covariate process. Such a model extends and generalizes the commonly used log-linear model. I will describe an estimating procedure for ρ and for the coefficient parameters β. Consistency and asymptotic normality of estimates of β can be achievedunder some regularity assumptions. I will show results from a simulation studyshowing the effectiveness of the procedure and from fitting the model to a dataset of fast food restaurant locations in New York City.
IP25	**Advances in Statistical Machine Learning, Thursday, 28 June 2018, 8:30 AM to 10:10 AM, ⑥ LT34**	
Org / Chair	Jianqing Fan, Princeton University / ,	
8:30 AM to 8:55 AM	Yang Feng Columbia University	"A LIKELIHOOD-RATIO TYPE TEST FOR STOCHASTIC BLOCK MODELS WITH BOUNDED DEGREES" A fundamental problem in network data analysis is to test whether a network contains statistical significant communities. We study this problem in the stochastic block model context by testing H0: Erdos-Renyi model vs. H1: stochastic block model. This problem serves as the foundation for many other problems including the testing-based methods for determining the number of communities and community detection. Existing work has been focusing on growing-degree regime while leaving the bounded-degree case untreated. Here, we propose a likelihood ratio type procedure based on regularization to test stochastic block models with bounded degrees. We derive the limiting distributions as power Poisson laws under both null and alternative hypotheses, based on which the limiting power of the test is carefully analyzed. The joint impact of signal-to-noise ratio and the number of communities on the asymptotic results is also unveiled. The proposed procedures are examined by both simulated and real-world network datasets. Our proofs depend on the contiguity theory for random regular graphs developed by Janson (1995).
8:55 AM to 9:20 AM	Jiashun Jin Carnegie Mellon University	#N/A
9:20 AM to 9:45 AM	Zheng (Tracy) Ke University of Chicago	#N/A
9:45 AM to 10:10 AM		#N/A
IP26	**Statistical Frontier in the Era of Big Data with Applications in Genomics, Wednesday, 27 June 2018, 1:30 PM to 3:10 PM, ⑦ LT31**	
Org / Chair	Jin Liu, Duke-NUS Medical School / Jin Liu, Duke-NUS Medical School	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
1:30 PM to 1:55 PM	Mingwei Dai Hong Kong University of Science and Technology	"IGESS: A STATISTICAL APPROACH TO INTEGRATING INDIVIDUAL-LEVEL GENOTYPE DATA AND SUMMARY STATISTICS" Results from genome-wide association studies (GWAS) suggest that a complex phenotype is often affected by many variants with small effects, known as 'polygenicity'. Tens of thousands of samples are often required to ensure statistical power of identifying these variants with small effects. However, it is often the case that a research group can only get approval for the access to individual-level genotype data with a limited sample size (e.g. a few hundreds or thousands). Meanwhile, summary statistics generated using single-variant-based analysis are becoming publicly available. The sample sizes associated with the summary statistics datasets are usually quite large. How to make the most efficient use of existing abundant data resources largely remains an open question. Results: In this study, we propose a statistical approach, IGESS, to increasing statistical power of identifying risk variants and improving accuracy of risk prediction by integrating individual level genotype data and summary statistics. An efficient algorithm based on variational inference is developed to handle the genome-wide analysis. Through comprehensive simulation studies, we demonstrated the advantages of IGESS over the methods which take either individual-level data or summary statistics data as input. We applied IGESS to perform integrative analysis of Crohns Disease from WTCCC and summary statistics from other studies. IGESS was able to significantly increase the statistical power of identifying risk variants and improve the risk prediction accuracy from 63.2% (±0.4%) to 69.4% (±0.1%) using about 240 000 variants.
1:55 PM to 2:20 PM	Yuling Jiao Zhongnan University of Economics and Law	#N/A
2:20 PM to 2:45 PM	Heng Peng Hong Kong Baptist University	"BOLT-SSI: FULLY SCREENING INTERACTION EFFECTS FOR ULTRA-HIGH DIMENSIONAL DATA" Detecting interaction effects among predict variables to response variables is often an crucial step in regression modeling of real data for various applications. In this paper by marginal likelihood functions, we firstly introduce a simple sure screening procedure (SSI) to fully detect significant pure interaction between predict variables and the response variable in the high or ultra-high dimensional generalized linear regression models. Furthermore, we suggest to discretize continuous predict variables, and utilize the Boolean operation for the marginal likelihood estimates. The so called BOLT-SSI procedure is proposed to accelerate the sure screening speed of the procedure. We investigate the sure screening properties of SSI and BOLT-SSI. Our studies have several important features. First, to most ultra-high dimensional data in practice, the proposed sure screening methods can fully detect any pure interaction effects among ultra-high dimensional data. It is an impossible finished task from theoretical insight. Second, the proposed method efficiently takes the advantages of computer architecture to speedup the proposed algorithm and make trade-off between the computation burden and statistical modeling efficiency. Specially, regarding the interaction effect detecting study as a special example, by this study we show the limitation of theoretical investigation from the practical insight, and illustrate that how to make trade-off between engineering techniques and theoretical investigations.
2:45 PM to 3:10 PM		#N/A
2:45 PM to 3:10 PM		#N/A
IP27	**The Ergodicity of Stochastic Partial Differential Equations and Related Topics, Friday, 29 June 2018, 3:30 PM to 5:10 PM, ④ LT26**	
Org / Chair	Yong Liu, Peking University / Yong Liu, Peking University	
3:30 PM to 3:55 PM	Yuan Liu Chinese Academy of Sciences	"ERGODICITY AND ASYMPTOTIC STABILITY OF FELL SEMIGROUPS ON METRIC SPACES" In this talk, we will give a review of ergodic criteria for Feller semigroups on metric spaces.
3:55 PM to 4:20 PM	Bin Xie Shinshu University	"HYPERCONTRACTIVE PROPERTY FOR REFLECTED SPDE DRIVEN BY SPACE-TIME WHITE NOISE" The SPDE with reflection, one kind of random parabolic obstacle problems, is regarded as the infinite dimensional Skorohod equation and is very important in applications. In this talk, we will mainly discuss the hypercontractive property of the Markov semigroup associated with the reflected SPDE driven by an additive space-time white noise under very weak conditions. In particular, thanks to the hypercontractivity of the Markov semigroup, some important properties, such as, the compactness of the Markov semigroup, the exponential convergences of the Markov semigroup to its unique invariant measure are obtained.
4:20 PM to 4:45 PM	Jianliang Zhai University of Science and Technology of China	"WELL-POSED FOR 2-D STOCHASTIC NAVIER-STOKES EQUATIONS DRIVEN BY LÉVY NOISE" Under the classical Lipschitz and linear growth assumptions, we established the existence and uniqueness of strong (in probability sense and PDE sense) solutions for 2-D Stochastic Navier-Stokes equations driven by multiplicative Lévy noise.
4:45 PM to 5:10 PM	Deng Zhang Shanghai Jiao Tong University	"SCATTERING FOR STOCHASTIC NONLINEAR SCHRÖDINGER EQUATIONS" introducing a copula Gaussian random elements Hilbert spaces, leading to what we call the Functional Copula Gaussian Graphical Model (FCGGM). This model removes the marginal
IP28	**Information Theory and Statistics, Friday, 29 June 2018, 10:30 AM to 12:10 PM, ③ LT29**	
Org / Chair	Po Ling Loh, University of Wisconsin–Madison / Po Ling Loh, University of Wisconsin–Madison	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
10:30 AM to 10:55 AM	Yudong Chen Cornell University	"LEAVE-ONE-OUT ANALYSIS FOR CONVEX AND NONCONVEX MATRIX COMPLETION" We introduce a powerful technique, Leave-One-Out, to the analysis of matrix completion problems. This technique allows us to obtain entry-wise bounds for iterative stochastic procedures. We demonstrate its power in analyzing two quintessential algorithms for matrix completion: the non-convex approach Singular Value Projection (SVP), and the convex relaxation approach Nuclear Norm Minimization (NNM). For SVP, we prove for the first time that the unmodified form of this algorithm, without sample splitting, converges linearly in infinity norm. For NNM, we study its dual solution and establish the first sample complexity bound that depends optimally on the matrix dimension and condition number.
10:55 AM to 11:20 AM	Varun Jog University of Wisconsin–Madison	"INFORMATION THEORETIC PERSPECTIVES ON LEARNING ALGORITHMS" In statistical learning theory, generalization error is used to quantify the degree to which a supervised machine learning algorithm may overfit to training data. We overview some recent work [Xu and Raginsky (2017)] that bounds generalization error of empirical risk minimization based on the mutual information I(S;W) between the algorithm input S and the algorithm output W. We leverage these results to derive generalization error bounds for a broad class of iterative algorithms that are characterized by bounded, noisy updates with Markovian structure, such as stochastic gradient Langevin dynamics (SGLD). We describe certain shortcomings of mutual information-based bounds, and propose alternate bounds that employ the Wasserstein metric from optimal transport theory. We compare the Wasserstein metric-based bounds with the mutual information-based bounds and show that for a class of data generating distributions, the former leads to stronger bounds on the generalization error.
11:20 AM to 11:45 AM	Cynthia Rush Columbia University	"SPARSE REGRESSION CODES: COMMUNICATIONS VIA HIGH-DIMENSIONAL LINEAR REGRESSION" In this talk, I present a provably fast, reliable, and capacity-achieving coding scheme for the additive white Gaussian noise (AWGN) channel. Coding over the AWGN channel is an important and relevant communications problem for which the fundamental limits were originally established by Claude Shannon in his seminal 1948 paper. Practical high-rate schemes for such channels were originally developed in the 1990s, allowing for the prevalence of wireless communication today. However, the reliability of such schemes has only been demonstrated empirically or the schemes were designed for sources and channels with finite, discrete input. Sparse regression codes, presented here, are based on a high-dimensional linear regression model, and they have fast decoders using approximate message passing (AMP) algorithms with error probability that is exponentially small for every fixed communication rate below capacity. In this presentation we describe the framework for the sparse regression coding scheme with the AMP decoder. This work is joint with Adam Greig, Kuan Hsieh, and Ramji Venkataramanan.
11:45 AM to 12:10 PM	Vincent Tan National University of Singapore	"ASYMPTOTIC COUPLING AND ITS APPLICATIONS IN INFORMATION THEORY" A coupling of two distributions P_X and P_Y is a joint distribution P_{XY} with marginal distributions equal to P_X and P_Y. Given marginals P_X and P_Y and a real-valued function f(P_{XY}) of the joint distribution P_{XY}, what is its minimum over all couplings P_{XY} of P_X and P_Y? We study the asymptotics of such coupling problems with different f's. These include the maximal coupling, minimum distance coupling, maximal guessing coupling, and minimum entropy coupling problems. We characterize the limiting values of these coupling problems as the number of copies of X and Y tends to infinity. We show that they typically converge at least exponentially fast to their limits. Moreover, for the problems of maximal coupling and minimum excess-distance probability coupling, we also characterize (or bound) the optimal convergence rates (exponents). Furthermore, for the maximal guessing coupling problem we show that it is equivalent to the probability distribution approximation problem. Therefore, some existing results the latter problem can be used to derive the asymptotics of the maximal guessing coupling problem. We also study the asymptotics of the maximal guessing coupling problem for two \emph{general} sources and a generalization of this problem, named the \emph{maximal guessing coupling through a channel problem}. We apply the preceding results to several new information-theoretic problems, including exact intrinsic randomness, exact resolvability, channel capacity with input distribution constraint, and perfect stealth and secrecy communication.
IP29	**Recent Advances in Statistical Methods for Analysis of Genomics Data , Wednesday, 27 June 2018, 10:30 AM to 12:10 PM, ④ LT26**	
Org / Chair	Qi Long, University of Pennsylvania / Qi Long, University of Pennsylvania	
10:30 AM to 10:55 AM	Mengjie Chen University of Chicago	"VARIABILITY-PRESERVING IMPUTATION FOR ACCURATE GENE EXPRESSION RECOVERY IN SINGLE CELL RNASEQ" Single cell RNA sequencing (scRNAseq) is becoming increasingly popular for transcriptomic profiling in genomics studies. However, high level of noise and excessive number of zero values in single cell gene expression measurements pose important challenges for accurate expression quantification. Here, we develop a method to impute the zero values in scRNAseq to facilitate accurate transcriptome quantification at the single cell level. Our method is based on a sparse nonnegative regression model, examines one cell at a time, and through a progressive selection strategy, infers a sparse set of local neighborhood cells that are most predictive of expression levels of the cell of interest. The zero values in the cell of interest are then imputed by a weighted summation of the expression measurements of these local neighborhood cells, with their imputation weights estimated through a quadratic programming algorithm. A key feature of our method is its ability to preserve gene expression variability across cells after imputation. We refer to our method as the Variability-preserving ImPutation for Expression Recovery (VIPER). We apply our method and compare it with existing imputation methods in several well-designed real data-based analytical experiments. We show that, compared to existing approaches, VIPER achieves superior imputation accuracy, maintains gene expression heterogeneity across cells, recovers gene expression measurements that better resemble the bulk RNAseq measurements, and facilitates more reproducible differential expression analysis.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
10:55 AM to 11:20 AM	Hongkai Ji Johns Hopkins University	"SINGLE-CELL ATAC-SEQ SIGNAL EXTRACTION AND ENHANCEMENT" Single-cell assay of transposase-accessible chromatin followed by sequencing (scATAC-seq) is an emerging new technology for studying gene regulation. Unlike the conventional ChIP-seq, DNase-seq and ATAC-seq technologies which measure average behavior of a cell population, scATAC-seq measures regulatory element activities within each individual cell, thereby allowing one to examine the heterogeneity of a cell population. Analyzing scATAC-seq data is challenging because the data are highly sparse and discrete. We present a statistical model to effectively extract signals from the noisy scATAC-seq data. Our method leverages information in massive amounts of publicly available DNase-seq data to enhance the scATAC-seq signal. We demonstrate through real data analyses that this approach substantially improves the accuracy for reconstructing genome-wide regulatory element activities.
11:20 AM to 11:45 AM	Jin Liu Duke-NUS Medical School	"JOINT ANALYSIS OF INDIVIDUAL-LEVEL AND SUMMARY-LEVEL GWAS DATA BY LEVERAGING PLEIOTROPY" A large number of recent genome-wide association studies (GWASs) for complex phenotypes confirm the early conjecture for polygenicity, suggesting the presence of large number of variants with only tiny or moderate effects. However, due to the limited sample size of a single GWAS, many associated genetic variants are too weak to achieve the genome-wide significance. These undiscovered variants further limit the prediction capability of GWAS. Restricted access to the individual-level data and the increasing availability of the published GWAS results motivate the development of methods integrating both the individual-level and summary-level data. How to build the connection between the individual-level and summary-level data determines the efficiency of using the existing abundant summary-level resources with limited individual-level data, and this issue inspires more efforts in the existing area. In this study, we propose a novel statistical approach, LEP, which provides a novel way of modeling the connection between the individual-level data and summary-level data. LEP integrates both types of data by LEveraing Pleiotropy to increase the statistical power of risk variants identification and the accuracy of risk prediction. The algorithm for parameter estimations is based on variational inference and scalable to handle large-scale genome-wide data. Through comprehensive simulation studies, we demonstrated the advantages of LEP over the existing methods. We further applied LEP to perform integrative analysis of Crohn's disease from WTCCC and summary statistics from some other autoimmune diseases. LEP was able to significantly increase the statistical power of identifying risk variants and improve the risk prediction accuracy from 63.39% (± 0.58%) to 68.33% (± 0.32%) using about 195,000 variants.
11:45 AM to 12:10 PM	Chaolong Wang Genome Institute of Singapore	"GENE-BASED RARE VARIANT ASSOCIATION TESTS FOR ANCESTRY-MATCHED CASE-CONTROL DATA" With an increasingly large amount of human sequencing data available, analysis incorporating external controls becomes a popular and cost-effective approach to boost statistical power in disease association studies. To prevent spurious association due to population stratification, it is important to carefully match the ancestry backgrounds of cases and external controls. However, rare variant association tests based on a standard logistic regression model, including both the burden test and sequence kernel association test (SKAT), are conservative when all ancestry-matched strata have the same case-control ratio and might become anti-conservative when case-control ratio varies across strata. To account for the matching structure, we propose analogous gene-based tests based on a conditional logistic regression (CLR) model, namely CLR-burden and CLR-SKAT. We show that the CLR model coupled with ancestry matching is a general approach to remove confounding effects due to population stratification. Through extensive simulations of population stratification and matching schemes, we demonstrate that both CLR-burden and CLR-SKAT robustly control the type 1 error, and are more powerful than standard burden test and SKAT in ancestry-matched data. Furthermore, because CLR-based tests allow for different case-control ratios across strata, a full-matching scheme can be employed to fully utilize available cases and controls to accelerate the discovery of disease genes.
IP30	**Emerging Methods in Big Data Analytics, Wednesday, 27 June 2018, 10:30 AM to 12:10 PM, ⑤ LT33**	
Org / Chair	Ping Ma, University of Georgia / Yehua Li, University of California, Riverside	
10:30 AM to 10:55 AM	TN Sriram University of Georgia	"ONLINE SEQUENTIAL LEVERAGING SAMPLING METHOD FOR STREAMING TIME SERIES DATA" Advances in data acquisition technology pose challenges in analyzing large volumes of streaming data. Sampling is a natural yet powerful tool for analyzing such data sets due to their competent estimation accuracy and low computational cost. Unfortunately, sampling methods and their statistical properties for streaming data, especially streaming time series data, are not well studied in the literature. In this article, we propose an online leverage-based sequential sampling algorithm for streaming time series data, which is assumed to come from an autoregressive model of order p â‰¥ 1 (AR(p)). The proposed sequential leveraging sampling method samples only one consecutively recorded block from the data stream for inference. While the starting point of the sequential sampling scheme is chosen using a random mechanism based on leverage scores of the data, the subsample size is decided by a sequential sampling threshold. We show that an appropriately normalized sequential least squares estimator of the AR parameter vector is uniformly asymptotically normally distributed for non-explosive AR(p) model. Simulation studies and real data examples are presented to evaluate the empirical performance of the proposed sequential leveraging sampling method.
10:55 AM to 11:20 AM	Tao Wang Shanghai Jiao Tong University	"GRAPH-ASSISTED INVERSE REGRESSION FOR COUNT DATA AND ITS APPLICATION TO SEQUENCING DATA" Multivariate count data, such as sequencing reads in genomics, are often connected to a clinical phenotype of interest. We develop a flexible framework for dimension reduction in regression, with predictors that are correlated counts, by modelling the conditional distribution of the predictors, given the response, using a pairwise Poisson graphical model. This new framework, called Graphical Inverse Regression with node-conditional Log-linear models (GIRL), allows us to derive a sufficient reduction of the predictors, while adjusting for the dependence structure among them. We propose a regularized criterion for estimating both the reduction structure and the network structure. The estimation algorithm can be implemented efficiently on a parallel computer. We also introduce an adaptive version and a sparse variant of the proposed procedure. The methods are evaluated on simulated data, and are applied to a gut microbiome sequencing data set.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
11:20 AM to 11:45 AM	Jingfei Zhang University of Miami	"A BIVARIATE POINT PROCESS MODEL WITH APPLICATION TO SOCIAL MEDIA USER CONTENT GENERATION" In this talk, we propose a new class of bivariate point process models to model the activity patterns of social media users. The proposed class of models has the flexibility to accommodate the complex behaviors of modern social media users and to provide straightforward insight into users' online content generating behavior. A composite likelihood approach and a composite likelihood EM procedure are developed to overcome the challenges in parameter estimation. We show the consistency and asymptotic normality of the maximum composite likelihood estimator. We apply our proposed method to President Donald Trump's Twitter data and uncover changes in various aspects of his tweeting behavior during the presidential campaign and the presidency. Moreover, we apply our method to a large scale social media data and find interesting subgroups of users with distinct behaviors. Additionally, we discuss the effect of social ties on a user's online content generating behavior.
11:45 AM to 12:10 PM		#N/A
IP31	**Functional Data Classification: Perspectives from Statistics and Machine Learning, Friday, 29 June 2018, 1:30 PM to 3:10 PM, ④ LT26**	
Org / Chair	Taps Maiti, Michigan State University / Chae Young Lim, Seoul National University	
1:30 PM to 1:55 PM	Yunjin Choi National University of Singapore	"COMPUTING CONDITIONAL DENSITY OF EIGENVALUES IN HIGH-DIMENSION" We propose a novel method for testing whether the currently selected principal components capture all the statistically significant signals in the given high-dimensional data set. The proposed method shows almost-exact type 1 error control property and decent size of power in detecting signals at the same time, while existing approaches do not enjoy the exact type 1 error property and lose power under some scenarios. Central to our work is the post-selection inference framework and the weighted Tracy-Widom law. The post-selection framework yields exact type 1 error control property by reflecting the data-driven model selection procedure to the test statistics. We utilize the weighted Tracy-Widom law to achieve the asymptotic distribution of the test statistics so that it can be applied to large scale data without heavy computation.
1:55 PM to 2:20 PM	Subhra Sankar Dhar Indian Institute of Technology	#N/A
2:20 PM to 2:45 PM	Po Ling Loh University of Wisconsin–Madison	"TWO INFERENCE PROBLEMS FOR NETWORK CONTAGION" We present two problems involving statistical inference for mathematical models of contagion spreading over a fixed network. The first problem concerns hypothesis testing for the underlying graph over which the disease is spreading, when we only observe the infection states of individual nodes after a single epidemic outbreak. We present a permutation test that is valid under appropriate conditions on the homogeneity of the spreading parameters and assumptions regarding the symmetry groups of the graphs involved in the null and alternative hypotheses. The second problem concerns parameter estimation for a similar type of contagion model, which incorporates covariate information on each of the edges of the graph. In this setting, we assume the structure of the graph is known, and we also know the order in which nodes contract the disease from their infected neighbors. We derive consistency and asymptotic normality of the maximum likelihood estimator, which may be obtained via convex optimization. This is joint work with Justin Khim (UPenn).
2:45 PM to 3:10 PM	Taps Maiti Michigan State University	"FUNCTIONAL GRAPHICAL MODEL BASED CLASSIFICATION: APPLICATION TO NEUROIMAGING" The functional magnetic resonance imaging (fMRI) records signals coming from different areas in human brains, which show activities and states of brains. This measurements result in high-dimensional time series or functional data. In this paper, we propose a functional Gaussian graphical model to describe the distribution and the correlation structure of this kind of high-dimensional data, and a quadratic discriminant function for this graphical model. Our simulation study shows that this classification method outperforms two other existing methods. Also, we present two real data classification applications.
IP32	**Advances in Modeling and Analysis of High-Dimensional Data , Thursday, 28 June 2018, 10:30 AM to 12:10 PM, ⑦ LT31**	
Org / Chair	Naveen Narisetty, University of Illinois / Naveen Narisetty, University of Illinois	
10:30 AM to 10:55 AM	Arindam Chatterjee Indian Statistical Institude	"HIGHER ORDER ASYMPTOTICS FOR BOOTSTRAPPED POST-MODEL SELECTION ESTIMATORS IN HIGH DIMENSIONS" Chatterjee and Lahiri (2013) showed that under suitable conditions, the residual Bootstrap is second order correct for studentized pivots based on the Adaptive Lasso. In this talk, we show that the second order correctness property holds quite generally for a number of penalized regression methods satisfying a version of the Oracle property of Fan and Li (2001). In particular, we show that under some suitable conditions, the LASSO and some popular nonconvex penalization functions including the SCAD and the MCP also enjoy second order correctness. Joint work with S. N. Lahiri (NCSU) and Debraj Das (Wisconsin Madison).
10:55 AM to 11:20 AM	Zijian Guo Rutgers University	"SEMI-SUPERVISED INFERENCE FOR EXPLAINED VARIANCE IN HIGH-DIMENSIONAL LINEAR REGRESSION AND ITS APPLI" We consider statistical inference for the explained variance $\beta^{\intercal}\Sigma \beta$ under the high-dimensional linear model $Y=X\beta+\epsilon$ in the semi-supervised setting, where β is the regression vector and Σ is the design covariance matrix. A calibrated estimator, which efficiently integrates both labelled and unlabelled data, is proposed. It is shown that the estimator achieves the minimax optimal rate of convergence in the general semi-supervised framework. The optimality result characterizes how the unlabelled data affects the minimax optimal rate. Moreover, the limiting distribution for the proposed estimator is established and data-driven confidence intervals for the explained variance are constructed. We further develop a randomized calibration technique for statistical inference in the presence of weak signals and apply the obtained inference results to a range of important statistical problems, including signal detection and global testing, prediction accuracy evaluation, and confidence ball construction. The numerical performance of the proposed methodology is demonstrated in simulation studies and an analysis of estimating heritability for a yeast segregant data set with multiple traits.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
11:20 AM to 11:45 AM	Jialiang Li National University of Singapore	"MULTI-THRESHOLD ACCELERATED FAILURE TIME MODEL" A two-stage procedure for simultaneously detecting multiple thresholds and achieving model selection in the segmented accelerated failure time (AFT) model is developed in this paper. In the first stage, we formulate the threshold problem as a group model selection problem so that a concave 2-norm group selection method can be applied. In the second stage, the thresholds are finalized via a refining method.\nWe establish the strong consistency of the threshold estimates and regression coefficient estimates under some mild technical conditions. The proposed procedure performs satisfactorily in our simulation studies. Its real world applicability is demonstrated via analyzing a follicular lymphoma data.
11:45 AM to 12:10 PM	Ying Wei Columbia University	#N/A
IP33	**Recent Advances in Machine Learning, Wednesday, 27 June 2018, 3:30 PM to 5:10 PM, ⑤ LT33**	
Org / Chair	Yang Ning, Cornell University / Cheng Li, National University of Singapore	
3:30 PM to 3:55 PM	Ian McKeague Columbia University	#N/A
3:55 PM to 4:20 PM	Kean Ming Tan Unniversity of Minnesota	"A NONCONVEX APPROACH TO SPARSE GENERALIZED EIGENVALUE PROBLEM AND ITS APPLICATION TO STATISTICS" Sparse generalized eigenvalue problem (GEP) plays a pivotal role in a large family of high-dimensional learning tasks, including sparse Fisher's discriminant analysis, canonical correlation analysis, and sufficient dimension reduction. Sparse GEP involves solving a non-convex optimization problem. Most of the existing methods and theory in the context of specific statistical models that are special cases of the GEP require restrictive structural assumptions on the input matrices. In this paper, we propose a two-stage computational framework to solve the sparse GEP. At the first stage, we solve a convex relaxation of the sparse GEP. Taking the solution as an initial value, we then exploit a nonconvex optimization perspective and propose the truncated Rayleigh flow method (Rifle) to estimate the leading generalized eigenvector and show that it converges linearly to a solution with the optimal statistical rate of convergence. Theoretically, our method significantly improves upon the existing literature by eliminating structural assumptions on the input matrices. To achieve this, our analysis involves two key ingredients: (i) a new analysis of the gradient based method on nonconvex objective functions, as well as (ii) a fine-grained characterization of the evolution of sparsity patterns along the solution path. Thorough numerical studies are provided to validate the theoretical results.
4:20 PM to 4:45 PM	Qing Cheng Shanghai University of Finance and Economics	"MARGINAL SCREENING FOR INTERACTION DETECTION VIA MIXED DERIVATIVES IN HIGH DIMENSIONS" Detecting interactions in high dimensions is a very challenging task in the literature. In this article we suggest mixed derivatives, the expectations of the off-diagonal elements of the Hessian matrix, to detect interactions for both linear and nonlinear models. We suggest two versions of our proposal: one is based on the centralized response and the other is based on the residual obtained from linear regression. We show that, our proposal possesses the desired sure screening and the ranking consistency properties under mild regularity conditions. Numerical studies through comprehensive simulations and an application confirm that our proposal is superior to existing approaches in terms of the minimal model size and the false positive and negative rates, especially when there exist quite many important interactions.
4:45 PM to 5:10 PM	Cheng Li National University of Singapore	"A DIVIDE-AND-CONQUER BAYESIAN APPROACH TO LARGE-SCALE KRIGING" Flexible hierarchical Bayesian modeling of massive data is challenging due to poorly scaling computations in large sample size settings. The motivation comes from spatial process models for analyzing geostatistical data, which typically entail computations that become prohibitive as the number of spatial locations becomes large. We propose a three-step divide-and-conquer strategy within the Bayesian paradigm to achieve massive scalability for any spatial process model. We partition the data into a large number of subsets, apply a readily available spatial process model on every subset in parallel, and optimally combine the posterior distributions estimated on all the subsets into a pseudo posterior distribution that is used for predictive and parametric inference and residual surface interpolation. We call this approach ``Distributed Kriging'' (DISK). The Bayes risk of estimating the true residual spatial surface using the DISK posterior distribution decays to zero at a nearly optimal rate under mild assumptions. While DISK is a general approach to divide-and-conquer Bayesian nonparametric regression, we focus on its applications in spatial statistics and demonstrate its empirical performance using models based on stationary full-rank and nonstationary low-rank Gaussian process priors. A variety of simulations and a geostatistical analysis of the Pacific Ocean sea surface temperature data validate our theoretical results.
IP34	**Analysis of Data with Dependent Structure , Friday, 29 June 2018, 8:30 AM to 10:10 AM, ⑥ LT34**	
Org / Chair	Zhen Pang, Hong Kong Polytechnic University / Zhen Pang, Hong Kong Polytechnic University	
8:30 AM to 8:55 AM	Guodong Li University of Hong Kong	"MULTILINEAR LOW-RANK VECTOR AUTOREGRESSIVE MODELING VIA TENSOR DECOMPOSITION" The classical Vector Autoregressive (VAR) model is the most commonly used multivariate time series model. The VAR model involves a large number of parameters so it can suffer from the curse of dimensionality for high-dimensional time series data. The reduced-rank coefficient model can alleviate the problem but the low-rank structure along the time direction for time series models has never been considered. In this paper, we rearrange the parameters in the VAR model to a tensor form, and propose a multilinear low-rank VAR model via tensor decomposition that effectively exploits the temporal and cross-sectional low-rank structure. Under this framework, an alternating least squares algorithm is developed for maximum likelihood estimation in the low-dimensional case and its asymptotic properties are studied. For the high-dimensional and large-scaled time series data, we develop a Sparse Higher-Order Reduced Rank (SHORR) estimator to further reduce the number of parameters and perform variable selection. The non-asymptotic error bound for the SHORR estimator is established and an ADMM-based algorithm is proposed. Effectiveness of the methods is demonstrated on simulated and real data.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:55 AM to 9:20 AM	Chun-Ling, Catherine Liu Hong Kong Polytechnic University	"SEMIPARAMETRIC BAYESIAN ANALYSIS FOR LONGITUDINAL MIXED EFFECTS MODELS WITH NON-NORMAL AR(1) ERRORS" In this talk, we present nonparametric Bayesian methodology for statistical inference under the setting of the longitudinal mixed effects model with non-normal AR(1) errors. We model the nonparametric zero-mean noise in the autoregression residual by the Dirichlet process (DP) mixture model. Applying the empirical likelihood tool, an adjusted sampler based on the Polya urn representation of DP is proposed to incorporate information of the moment constraints of the mixing distribution. Gibbs sampling algorithm based on the adjusted sampler is proposed to approximate the posterior distributions under DP priors. The proposed method can be easily extended to deal with other moment constraints owing to the wide application background of empirical likelihood. Simulation studies evaluate the performance of the proposed method. Our method is illustrated in analysis of a longitudinal data set from a psychiatric study.
9:20 AM to 9:45 AM	Mengjiao Peng Nanyang Technological University	"JOINT REGRESSION ANALYSIS FOR SURVIVAL DATA IN THE PRESENCE OF TWO SEMI-COMPETING RISKS" Unlike the usual semi-competing risks setting that includes two types of events: the nonterminal and terminal events, in some clinical studies with complex disease process individuals may be at risk of several different types of clinical events. We consider a situation in which three events: two nonterminal events and one terminal event are of interest. This is the case of our motivating bone marrow transplant study in which individuals may experience the acute Graft-Versus-host disease (GvHD), relapse and death after an allogeneic transplant, where the GvHD is associated with the relapse free survival, both the GVHD and relapse are intermediate nonterminal events subject to dependent censoring by the informative terminal event death, but not vice versa. We propose a novel statistical approach that jointly models times to these three types of events using a pair of copulas to account for dependence structures, while each marginal distribution of the event times is formulated by a Cox proportional hazards model. We develop an estimation procedure based on pseudo likelihood and carry out simulation studies to examine the performance of the proposed method in finite samples. The practical utility of the proposed methodology is further illustrated with data from the motivating example on bone marrow transplant.
9:45 AM to 10:10 AM	Liming Xiang Nanyang Technological University	"GENERALIZED LINEAR MIXED MODELS WITH MISSING COVARIATES AND DISTRIBUTION-FREE RANDOM EFFECTS" We consider generalized linear mixed models in which random effects are free of parametric distributions and missing at random data are present in some covariates. To overcome the problem of missing data, we propose two novel methods relying on auxiliary variables: a penalized conditional likelihood method when covariates are independent of random effects, and a two-step procedure consisting of a pairwise likelihood for estimating fixed effects in the first step and a penalized conditional likelihood for estimating random effects in the second step while covariates can be related to random effects. Our methods require no distribution assumption for random effects and allow a nonparametric error structure, thus providing great flexibility in capturing a board range of behaviors of both the error term and random effects. We show that the proposed estimators enjoy desirable properties such as consistency and asymptotically normality, and assess their finite sample performance through extensive simulation studies. The proposed methods are further illustrated using a longitudinal data set on forest health monitoring.
IP35	**Challenges and Advances in High-Dimensional Data Analysis, Wednesday, 27 June 2018, 8:30 AM to 10:10 AM, ⑤ LT33**	
Org / Chair	Cheolwoo Park, University of Georgia / Cheolwoo Park, University of Georgia	
8:30 AM to 8:55 AM	Jeongyoun Ahn University of Georgia	"GENERALIZED RAYLEIGH RATIO OPTIMIZATION FOR HIGH-DIMENSIONAL DISCRIMINATION" In multi-class discrimination with high dimensional data, identifying a lower-dimensional subspace with maximum class separation is crucial. We focus on a less conventional optimization criterion for finding such a discriminant subspace, which is the ratio of two traces: the trace of between-class scatter matrix and the trace of within-class scatter matrix. Since this problem is not well-defined for high-dimensional data, we propose to regularize the within trace while maximizing the between trace. A careful investigation reveals that this optimization has an innate connection to the eigenvalue decomposition of a symmetric indefinite matrix that is a function of the traces and a tuning parameter. For the sake of better interpretability for high dimensional problems, we add a sparse estimation component to the algorithm that relies on an iterative soft-thresholding. Interesting relationships between the proposed method and some classical methods such as Fisher's linear discriminant analysis are discussed. Empirical examples on simulated and real data sets suggest that the proposed method works competitively with or better than existing approaches in a wide range of problems, with respect to both variable selectivity and classification accuracy.
8:55 AM to 9:20 AM	Won Chang University of Cincinnati	"CALIBRATING AN ICE SHEET MODEL USING HIGH-DIMENSIONAL BINARY SPATIAL DATA" Rapid retreat of ice in the Amundsen Sea sector of West Antarctica may cause drastic sea level rise, posing significant risks to populations in low-lying coastal regions. Calibration of computer models representing the behavior of the West Antarctic Ice Sheet is key for informative projections of future sea level rise. However, both the relevant observations and the model output are high-dimensional binary spatial data; existing computer model calibration methods are unable to handle such data. Here we present a novel calibration method for computer models whose output is in the form of binary spatial data. To mitigate the computational and inferential challenges posed by our approach, we apply a generalized principal component based dimension reduction method. To demonstrate the utility of our method, we calibrate the PSU3D-ICE model by comparing the output from a 499-member perturbed-parameter ensemble with observations from the Amundsen Sea sector of the ice sheet. Our methods help rigorously characterize the parameter uncertainty even in the presence of systematic data-model discrepancies and dependence in the errors. Our method also helps inform environmental risk analyses by contributing to improved projections of sea level rise from the ice sheets.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
9:20 AM to 9:45 AM	Changwon Lim Chung-Ang University	"ANALYSIS OF THE YEARBOOK FROM THE KOREA METEOROLOGICAL ADMINISTRATION USING A TEXT-MINING ALGORITHM" Many people have recently posted about personal interests on social media. The development of the Internet and computer technology has enabled the storage of digital forms of documents that has resulted in an explosion of the amount of textual data generated; subsequently there is an increased demand for technology to create valuable information from a large number of documents. A text mining technique is often used since text-based data is mostly composed of unstructured forms that are not suitable for the application of statistical analysis or data mining techniques. This study analyzed the Meteorological Yearbook data of the Korea Meteorological Administration (KMA) with a text mining technique. First, a term dictionary was constructed through preprocessing and a term-document matrix was generated. This term dictionary was then used to calculate the annual frequency of term, and observe the change in relative frequency for frequently appearing words. We also used regression analysis to identify terms with increasing and decreasing trends. We analyzed the trends in the Meteorological Yearbook of the KMA and analyzed trends of weather related news, weather status, and status of work trends that the KMA focused on. This study is to provide useful information that can help analyze and improve the meteorological services and reflect meteorological policy.
9:45 AM to 10:10 AM	Miles Lopes University of California, Davis	"NORMAL DISTRIBUTION WITH A SHRINKAGE INVERSE WISHART PRIORS" Recently published functional graphical models rely on the assumption that the random functions are Hilbert-space-valued Gaussian random elements.
IP36	**Applications of Random Matrix Theory to High-Dimensional Inference, Friday, 29 June 2018, 1:30 PM to 3:10 PM, ⑤ LT33**	
Org / Chair	Debashis Paul, University of California, Davis / Debashis Paul, University of California, Davis	
1:30 PM to 1:55 PM	Gourab Mukherjee University of Southern California	"IMPROVED SHRINKAGE PREDICTION UNDER A SPIKED COVARIANCE STRUCTURE" We develop a novel shrinkage rule for prediction in a high dimensional non-exchangeable hierarchical Gaussian model with an unknown spiked covariance structure. We propose a family of commutative priors which, governed by a power hyper-parameter, ranges from perfect independence to highly dependent scenarios. It induces a wide class of predictors whose evaluation involves quadratic forms of smooth functions of the unknown covariance. Our proposed adaptive prediction procedure outperforms factor model based plug-in predictors by using uniformly consistent estimators of the quadratic forms involved in the coordinate-wise shrinkage strategies. We further improve our predictor by introducing possible reduction in its variability through a novel coordinate-wise shrinkage policy that only uses covariance level information and can be adaptively tuned using the sample eigen structure of the high dimensional spiked covariance model. Simulation studies are conducted to show that in many settings the proposed method substantially improves the performance of traditional plug-in based shrinkage procedures which first estimate the covariance and thereafter optimize over the hyper-parameters.
1:55 PM to 2:20 PM	Guangming Pan Nanyang Technological University	"LIMITING LAWS FOR DIVERGENT SPIKED EIGENVALUES AND LARGEST NON-SPIKED EIGENVALUE OF SAMPLE MATRICES" This talk is about the spiked eigenvalues of sample covariance matrices when the dimension and sample size both tend to infinity with certain rate. We show that the asymptotic distribution of the largest eigenvalue is the Tracy-Widow law and the largest non spike eigenvalue is Gaussian distributed. We also explore its application in the factor model
2:20 PM to 2:45 PM	Debashis Paul University of California, Davis	"HIGH-DIMENSIONAL TESTS FOR GENERAL LINEAR HYPOTHESES THROUGH SPECTRAL SHRINKAGE" We consider the problem of testing linear hypotheses under a multivariate linear regression model when the dimension of the observations is comparable to the sample size. This framework also includes tests for Multivariate Analysis of Variance (MANOVA). We propose a family of rotation-equivariant tests with a class of analytic functions used as spectral regularizers for the residual sample covariance matrix. We derive asymptotic normality of the test statistic under some moment conditions on the observations. We also study asymptotic power of the test under various probabilistic local alternatives. We carry out simulation studies to examine the numerical performance of the test in the MANOVA set-up, and compare with several competing tests for high-dimensional MANOVA problems. (This is a joint work with Haoran Li and Alexander Aue)
2:45 PM to 3:10 PM		#N/A
IP37	**Statistical Inverse Problems in Neuromaging, Friday, 29 June 2018, 10:30 AM to 12:10 PM, ④ LT26**	
Org / Chair	Debashis Paul, University of California, Davis / Debashis Paul, University of California, Davis	
10:30 AM to 10:55 AM	Hernando Ombao King Abdullah University of Science and Technology	"HERITABILITY OF EEG COHERENCE DURING A VISUAL WORKING MEMORY TASK" Authors: Dustin Pluta* (UC Irvine and KAUST), Zhaoxia Yu (UC Irvine) and Hernando Ombao (KAUST) A number of recent studies have found evidence that characteristics of functional brain connectivity are significantly associated with various genetic markers, however, the majority of work in this area has been restricted to resting state fMRI data. We here present the results from a novel study of EEG coherence during a working memory task from 350 healthy university students. Using recently developed statistical methods for testing associations between high-dimensional feature sets (which improves upon existing statistical methods through the use of non-Euclidean metrics), we identify specific sets of channels for which the coherence measures in the delta, theta, alpha, and beta frequency bands are significantly associated with a set of genetic markers previously implicated as risk factors for Alzheimer's disease. Additionally, we compare these heritability estimates with genome-wide heritability and with estimates from a set of neurotransmitter genes related to dopamine regulation. These results suggest that some genetic factors linked to Alzheimer's disease may also play a role in working memory performance in healthy individuals.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
10:55 AM to 11:20 AM	Armin Schwartzman University of California, San Diego	"DO NOT TEST FOR ACTIVATION IN FMRI BUT ESTIMATE THE REGIONS OF ACTIVATION" Null hypothesis testing lies at the foundation of human brain mapping as the core method for fMRI inference. However, recent studies have shown that under optimal conditions the null hypothesis is never true, and brain activity related to a task can be found everywhere in the brain. Rather than testing for significance, we propose to directly estimate the spatial extent of interesting brain activity, defined as excursion sets of the percentage BOLD signal change above a pre-defined threshold. The uncertainty in the estimates is then captured by a nested pair of spatial confidence regions (CRs) called inner and outer sets. These spatial CRs are defined in such a way that the true excursion sets include the inner set and are included in the outer set with a given confidence. Asymptotic coverage probabilities may be determined using the Gaussian kinematic formula or via a multiplier bootstrap. The method is illustrated in task fMRI data from the Human Connectome Project.
11:20 AM to 11:45 AM	Hang Yu Nanyang Technological University	"TUNING-FREE BAYESIAN INFERENCE OF FUNCTIONAL BRAIN NETWORKS" We present a novel Bayesian method to learn function brain networks in an automated fashion. Specifically, we employ Gaussian graphical models (GMM) to describe the functional networks. As a result, the network structure is characterized by the zero pattern of the precision matrix corresponding to the GGM. From a Bayesian perspective, we regard the off-diagonal entries in the precision matrix as random variables and impose sparse-promoting priors on them, resulting in automatic sparsity determination. With the help of stochastic gradients, an efficient variational Bayes algorithm is derived to learn the model. The time complexity of the algorithm is only quadratic in dimension, whereas the complexity of the state-of-the-art methods is typically cubic in dimension. Thus, the proposed approach is more applicable to inferring large-scale networks. As an application, we learn functional networks from real-life fMRI data and find that resulting networks yield high classification accuracy for the mind-state prediction problem.
11:45 AM to 12:10 PM	Hongtu Zhu MD Anderson Cancer Center	#N/A
IP38	**Recent Advances and Applications of Wavelet Methods in Nonparametric Statistics, Friday, 29 June 2018, 3:30 PM to 5:10 PM, ⑤ LT33**	
Org / Chair	Spiridon Penev, University of New South Wales / Gery Geenens, University of New South Wales	
3:30 PM to 3:55 PM	Hassan Doosti Macquarie University	"NONPARAMETRIC ESTIMATION OF A QUANTILE DENSITY FUNCTION BY WAVELET METHODS" In this talk three different nonparametric wavelet estimators of the quantile density function are proposed. Consistency of the wavelet estimators is established under the L_p risk. Numerical study illustrates the good performance of our estimators.
3:55 PM to 4:20 PM	Gery Geenens University of New South Wales	"SHAPE-PRESERVING WAVELET-BASED MULTIVARIATE DENSITY ESTIMATION" Wavelet estimators for a probability density f enjoy many good properties, however they are not `shape-preserving' in the sense that the final estimate may not be non-negative or integrate to unity. A solution to negativity issues may be to estimate first the square-root of f and then square this estimate up. This paper proposes and investigates such an estimation scheme, generalising to higher dimensions some previous constructions which are valid only in one dimension. The estimation is mainly based on nearest-neighbour-balls. The theoretical properties of the proposed estimator are obtained, and it is shown to reach the optimal rate of convergence uniformly over large classes of densities under mild conditions. Simulations show that the new estimator performs as well in terms of Mean Integrated Squared Error as the classical wavelet estimator and better than it in terms of Mean Squared Hellinger Distance between the estimator and the truth, while automatically producing estimates which are bona fide densities.
4:20 PM to 4:45 PM	Fabien Navarro ENSAI, Rennes	"WAVELET ESTIMATION WITH ADDITIVE AND MULTIPLICATIVE NOISE" We consider the estimation of an unknown regression function from a nonparametric regression model having the feature to have multiplicative noise and additive noise. We focus our attention on wavelet methods; we develop a linear wavelet estimator and a nonlinear wavelet estimator based on a thresholding of the wavelet coefficients estimators. We prove that they attains fast rates of convergence under the mean integrated square error over over Besov spaces. The obtained rates are fast and are obtained under some weak conditions on the model. A numerical study supports the theory.
4:45 PM to 5:10 PM		#N/A
IP39	**Recent Advances in Design and Analysis of Experiments, Wednesday, 27 June 2018, 1:30 PM to 3:10 PM, ⑤ LT33**	
Org / Chair	Frederick Kin Hing Phoa, Academia Sinica / Frederick Kin Hing Phoa, Academia Sinica	
1:30 PM to 1:55 PM	Yasmeen Akhtar Academia Sinica	"A CONSTRUCTION OF COST-EFFICIENT DESIGNS WITH GUARANTEED REPEATED MEASUREMENTS ON INTERACTION EFFECT" This talk introduces a useful class of cost-efficient designs for multi-factor experiments, namely (supersaturated) repeated coverage designs or (SS)RCD in short. By running a small number of runs, (SS)RCD possesses the desired property that the measurements of all t-tuples from any t factors are guaranteed to be repeated at least λ times, where λ is an adjustable parameter that depends on the experimental resource. It is then compared to two famous classes of designs: Covering Arrays and Orthogonal Arrays. The former class lacks the resistance of outlying measurements as all t-tuples are guaranteed to be measured at least once only. The latter class utilizes too many experimental resources to guarantee all t-tuples are measured equally often. We develop a systematic method to construct families of (SS)RCD with small run sizes under the different number of factors and number of repetitions (λ).

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
1:55 PM to 2:20 PM	Ming Hung Kao Arizona State University	"SHAPE-PRESERVING WAVELET-BASED MULTIVARIATE DENSITY ESTIMATION" Wavelet estimators for a probability density f enjoy many good properties, however they are not `shape-preserving' in the sense that the final estimate may not be non-negative or integrate to unity. A solution to negativity issues may be to estimate first the square-root of f and then square this estimate up. This paper proposes and investigates such an estimation scheme, generalising to higher dimensions some previous constructions which are valid only in one dimension. The estimation is mainly based on nearest-neighbour-balls. The theoretical properties of the proposed estimator are obtained, and it is shown to reach the optimal rate of convergence uniformly over large classes of densities under mild conditions. Simulations show that the new estimator performs as well in terms of Mean Integrated Squared Error as the classical wavelet estimator and better than it in terms of Mean Squared Hellinger Distance between the estimator and the truth, while automatically producing estimates which are bona fide densities.
2:20 PM to 2:45 PM	Frederick Kin Hing Phoa Academia Sinica	"DESIGNING EXPERIMENTS WITH UNSTRUCTURED TREATMENTS FOR GENERAL NETWORK STRUCTURES" Experiments on connected units are commonly conducted in various fields, such as agriculture trials, medical experiments and social networks. In these applications, an experimental unit connects to one another, and the treatment applied to a unit has an effect, called a \emph{network effect}, on the responses of the neighboring units. Designing such experiments was rarely discussed in the literature. Parker, Gilmour, and Schormans (2017) initiated a study of A_s-optimal designs on connected experimental units with unstructured treatments, assuming that the network effects were unknown constants. In this work, we studied a similar design problem under an assumption that the network effects were random effects. It led to a property that the responses of two units were correlated if some neighbors of one unit and those of the other received the same treatment. Alphabetical optimality criteria were considered for selecting good designs with high efficiency of estimating the treatment effects and/or high accuracy of predicting the network effects. We provided theoretical conditions for designs to be optimal and illustrate our theory with some numerical examples.
2:45 PM to 3:10 PM	John Stufken Arizona State University	"OPTIMAL DESIGN FOR MIXED EFFECTS MODELS" Identifying optimal designs for correlated data is a difficult problem. Many classical results for independent data have no obvious generalization to correlated data. We propose a method to identify locally optimal designs for classes of linear, generalized linear, and nonlinear mixed effects models under commonly used optimality criteria by extending results for independent data. We demonstrate the method through a real life study, and investigate robustness of design efficiency to mis-specification of the covariance matrix for the random effects.
IP40	**High-Dimensional Functional Data Analysis, Wednesday, 27 June 2018, 10:30 AM to 12:10 PM, ⑥ LT34**	
Org / Chair	Xinghao Qiao, London School of Economics / Xinghao Qiao, London School of Economics	
10:30 AM to 10:55 AM	Bing Li Pennsylvania State University	"COPULA GAUSSIAN GRAPHICAL MODELS FOR FUNCTIONAL DATA" We consider the problem of constructing statistical graphical models for functional data; that is, the observations on the vertices are random functions. This types of data are common in medical applications such as EEG and fMRI. Recently published functional graphical models rely on the assumption that the random functions are Hilbert-space-valued Gaussian random elements. We relax this assumption by introducing a copula Gaussian random elements Hilbert spaces, leading to what we call the Functional Copula Gaussian Graphical Model (FCGGM). This model removes the marginal Gaussian assumption but retains the simplicity of the Gaussian dependence structure, which is particularly attractive for large data. We develop four estimators, together with their implementation algorithms, for the FCGGM. We establish the consistency and the convergence rates of one of the estimators under different sets of sufficient conditions with varying strengths. We compare our FCGGM with the existing functional Gaussian graphical model by simulation, under both non-Gaussian and Gaussian graphical models, and apply our method to an EEG data set to construct brain networks.
10:55 AM to 11:20 AM	Matthew Reimherr Pennsylvania State University	"THE SMOOTH ELASTIC NET FOR FUNCTION-ON-SCALAR REGRESSION" In this talk I will present new work concerning high-dimensional function-on-scalar regression. This new framework allows for simultaneous variable selection and parameter estimation, while carefully controlling the level of smoothing in the resulting parameter estimates. Our approach is based on using an RKHS penalty within an Elastic Net. We are able to obtain a fast coordinate descent algorithm as well as theoretical guarantees including strong oracle properties in multiple topologies. Numerical experiments comparing with previous approaches will also be presented.
11:20 AM to 11:45 AM	Xinghao Qiao London School of Economics	"REGRESSION WITH FUNCTIONAL ERRORS-IN-PREDICTORS:A GENERALIZED METHOD-OF-MOMENTS APPROACH" Functional regression is an important topic in functional data analysis. Traditionally, one often assumes that samples of the functional predictor are independent realizations of an underlying stochastic process, and are observed over a grid of points contaminated by independent and identically distributed measurement errors. In practice, however, the dynamic dependence across different curves may exist and the parametric assumption on the measurement error covariance structure could be unrealistic. In this paper, we consider functional linear regression with serially dependent functional predictors, when the contamination of predictors by measurement error is ``genuinely functional" with fully nonparametric covariance structure. Inspired by the fact that the autocovariance operator of observed functional predictors automatically filters out the impact from the unobservable measurement error, we propose a novel autocovariance-based generalized method-of-moments estimate of the slope parameter. The asymptotic properties of the resulting estimators under different functional scenarios are established. We also demonstrate that our proposed method significantly outperforms possible competitors through intensive simulation studies. Finally, the proposed method is applied to a public financial dataset, revealing some interesting findings.
11:45 AM to 12:10 PM		#N/A
IP41	**Recent Advances and Challenges in Big Data Inference, Wednesday, 27 June 2018, 8:30 AM to 10:10 AM, ⑧ MD9-0102GH**	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
Org / Chair	Zhao Ren, University of Pittsburgh / Fengnan Gao, Fudan University	
8:30 AM to 8:55 AM	Jinyuan Chang Southwestern University of Finance and Economics	"HIGH-DIMENSIONAL STATISTICAL INFERENCES WITH OVER-IDENTIFICATION" Over-identification is a signature feature of the influential Generalized Method of Moments (Hansen, 1982) that flexibly allows more moment conditions than the model parameters. Investigating over-identification together with high-dimensional statistical problems is challenging and remains less explored. In this paper, we study two high-dimensional statistical problems with over-identification. The first one concerns statistical inferences associated with multiple components of the high-dimensional model parameters, and the second one is on developing a specification test for assessing the validity of the over-identified moment conditions. For the first problem, we propose to construct a new set of estimating functions such that the impact from estimating the nuisance parameters becomes asymptotically negligible. Based on the new construction, a confidence set is estimated using empirical likelihood (EL) for the specified components of the model parameters. For the second problem, we propose a test statistic as the maximum of the marginal EL ratios respectively calculated from individual components of the high-dimensional moment conditions. Our theoretical analysis establishes the validity of the proposed procedures, accommodating exponentially growing data dimensionality, and our numerical examples demonstrate good performance and potential practical benefits of our proposed methods with high-dimensional problems.
8:55 AM to 9:20 AM	Andreas Elsener ETH Zürich	"SHARP ORACLE INEQUALITIES FOR STATIONARY POINTS OF NONCONVEX PENALIZED M-ESTIMATORS" A new framework to derive oracle inequalities for stationary points of nonconvex norm penalized M-estimators is proposed. One main novelty is that the oracle inequalities are sharp. The sharpness of the inequality is referred to the constant "1" in front of the approximation error term in the upper bound of the inequalities. The loss functions are assumed to be differentiable. The penalty is assumed to be a weakly decomposable norm. Applications of the general framework to sparse errors-in-variables, sparse PCA and sparse robust regression are presented. Joint work with Sara van de Geer (ETH Zurich).
9:20 AM to 9:45 AM	Lingzhou Xue Pennsylvania State University	#N/A
9:45 AM to 10:10 AM	Wenxin Zhou University of California, San Diego	"ADAPTIVE HUBER REGRESSION: ESTIMATION AND INFERENCE" This paper is devoted to investigate the theoretical underpinnings of both classical and contemporary statistical inference problems, where heavy-tailed data are present. In particular, we focus on inference for uncertainty quantification, including the construction of confidence sets and large-scale simultaneous hypothesis testing. In the presence of heavy-tailed data, finite sample properties of the least squares-based methods, typified by the sample mean, are suboptimal both theoretically and empirically. In this paper, we demonstrate that the adaptive Huber regression, combined with the multiplier bootstrap procedure, provides a useful robust alternative to the method of least squares. Our theoretical and empirical results reveal the effectiveness of the proposed method, and highlight the importance of having inference methods that are robust to heavy-tailed data.
IP42	**Asymptotic Theory in Probability and Statistics, Tuesday, 26 June 2018, 3:30 PM to 5:10 PM, ⑤ LT33**	
Org / Chair	Qi-Man Shao, Chinese University of Hong Kong / Qi-Man Shao, Chinese University of Hong Kong	
3:30 PM to 3:55 PM	Xia Chen University of Tennessee	#N/A
3:55 PM to 4:20 PM	Xiequan Fan Tianjin University	"SELF-NORMALIZED CRAMER TYPE LARGE DEVIATIONS FOR MARTINGALES" Self-normalized Cramer type large deviations for independent random variables has been well studied in recent year. One of the most interesting work is due to Jing, Shao and Wang (2003, Ann. Probab.). They proved that self-normalized Cramer type large deviation results hold only under a finite $(2+\rho)$th moment, $0eng Liu and Qi-Man Shao)
4:20 PM to 4:45 PM	Lan Gao Chinese University of Hong Kong	"CRAMER TYPE MODERATE DEVIATION FOR STUDENTIZED TRIMMED MEAN" Trimmed mean is a commonly used robust estimator of a location parameter. In practice, one can obtain the studentized trimmed mean via estimating the variance of trimmed mean by the Winsorized sample variance. It is well known that the studentized trimmed mean has asymptotic normality under mild conditions. There are many results on the absolute error of normal approximation and in contrast few results on relative error. In our talk, we will show the Cramer type moderate deviation for studentized trimmed mean. This is a joint work with Qiman Shao and Jiasheng Shi.
4:45 PM to 5:10 PM	Jiasheng Shi Chinese University of Hong Kong	"A CRAMÉR MODERATE DEVIATION THEOREM FOR GENERAL SELF-NORMALIZED SUMS" Asymptotic theory for self-normalized sums has been well studied in the past two decades. In this talk we focus on a general self-normalized sums \sum X_i / \sqrt{\sum Y_i^2}, where (X_i,Y_i), 1\leq i \leq n are independent random vectors. The Cramer type moderate deviation theorem is obtained under optimal moment condition. Applications to self-normalized dependent random variables will also be discussed. This is a joint work with Prof.Qi-Man SHAO and Lan Gao.
IP43	**Statistical Analysis of Data Heterogeneity, Thursday, 28 June 2018, 8:30 AM to 10:10 AM, ⑦ LT31**	
Org / Chair	Juan Shen, Fudan University / Ying Wei, Columbia University	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:30 AM to 8:55 AM	Marc Genton King Abdullah University of Science and Technology	"A STOCHASTIC GENERATOR OF GLOBAL MONTHLY WIND ENERGY WITH TUKEY G-AND-H AUTOREGRESSIVE PROCESSES" Quantifying the uncertainty of wind energy potential from climate models is a very time-consuming task and requires a considerable amount of computational resources. A statistical model trained on a small set of runs can act as a stochastic approximation of the original climate model, and be used to assess the uncertainty considerably faster than by resorting to the original climate model for additional runs. While Gaussian models have been widely employed as means to approximate climate models, the Gaussianity assumption is not suitable for winds at policy-relevant time scales, i.e., sub-annual. We propose a trans-Gaussian model for monthly wind speed that relies on an autoregressive structure with Tukey g-and-h innovations, a flexible new class that can separately model skewness and tail behavior. This temporal structure is integrated into a multi-step spectral framework that is able to account for global nonstationarities across land/ocean boundaries, as well as across mountain ranges. Inference can be achieved by balancing memory storage and distributed computation for a data set of 220 million points.
8:55 AM to 9:20 AM	Bo Li University of Illinois at Urbana-Champaign	#N/A
9:20 AM to 9:45 AM	Naveen Narisetty University of Illinois at Urbana-Champaign	"BAYESIAN REGULARIZATION FOR GRAPHICAL MODELS WITH UNEQUAL SHRINKAGE" We consider a Bayesian framework for estimating a high-dimensional sparse precision matrix, in which adaptive shrinkage and sparsity are induced by a mixture of Laplace priors. Besides discussing our formulation from the Bayesian standpoint, we investigate the MAP (maximum a posteriori) estimator from a penalized likelihood perspective that gives rise to a new non-convex penalty approximating the L0 penalty. Optimal error rates for estimation consistency in terms of various matrix norms along with selection consistency for sparse structure recovery are shown for the unique MAP estimator under mild conditions. For fast and efficient computation, an EM algorithm is proposed to compute the MAP estimator of the precision matrix and (approximate) posterior probabilities on the edges of the underlying sparse structure. Simulation studies and a real application are used to demonstrate the performance of our method compared. (This is joint work with Lingrui Gan and Feng Liang).
9:45 AM to 10:10 AM	Xiaolei Xun Fudan University	"OPTIMAL TEST PROCEDURES FOR MULTIPLE HYPOTHESES CONTROLLING THE FAMILYWISE EXPECTED LOSS" We consider the problem of testing multiple null hypotheses where a decision to reject or retain is to be made for each individual hypothesis. Based on the decision-theoretic framework, we propose to control the familywise expected loss instead of the conventional familywise error rate (FWER). Various loss functions can be adopted and the FWER is seen to result as a particular choice of the loss function. We search for decision rules that satisfy certain optimality criteria within a broad class of rules for which the expected loss is bounded by a pre-specified threshold under any parameter configuration. This approach is different from the canonical decision theory of maximizing a single utility function, but in analogy to classical hypothesis testing. We illustrate the methods with the problem of establishing efficacy of a new medicinal treatment in non-overlapping subgroups of patients.
IP44	**Recent Advances in Complex Data Analysis, Wednesday, 27 June 2018, 3:30 PM to 5:10 PM, ⑥ LT34**	
Org / Chair	Weining Shen, University of California, Irvine / Weixuan Zhu, Xiamen University	
3:30 PM to 3:55 PM	Yen Chi Chen University of Washington	"DENSITY TREE AND DENSITY RANKING IN SINGULAR MEASURES" A density tree (also known as a cluster tree of a probability density function) is a tool in topological data analysis that uses a tree structure to represent the shape of a density function. Even if the density function is multivariate, a density tree can always be displayed on a two-dimensional plane, making it an ideal tool for visualizing the shape of a multivariate dataset. However, in complex datasets such as GPS data, the underlying distribution function is singular so the usual density function and density tree no longer exist. To analyze this type of data and generalize the density tree, we introduce the concept of density ranking and ranking tree (also called an α-tree). We then show that one can consistently estimate the density ranking and the ranking tree using a kernel density estimator. Based on the density ranking, we introduce several geometric and topological summary curves for analyzing GPS datasets.
3:55 PM to 4:20 PM	Raphael Huser King Abdullah University of Science and Technology	"MODELING SPATIAL PROCESSES WITH UNKNOWN EXTREMAL DEPENDENCE CLASS" Many environmental processes exhibit weakening spatial dependence as events become more extreme. Well-known limiting models, such as max-stable or generalized Pareto processes, cannot capture this, which can lead to a preference for models that exhibit a property known as asymptotic independence. However, weakening dependence does not automatically imply asymptotic independence, and whether the process is truly asymptotically (in)dependent is usually far from clear. The distinction is key as it can have a large impact upon extrapolation, i.e., the estimated probabilities of events more extreme than those observed. In this work, we present a single spatial model that is able to capture both dependence classes in a parsimonious manner, and with a smooth transition between the two cases. The model covers a wide range of possibilities from asymptotic independence through to complete dependence, and permits weakening dependence of extremes even under asymptotic dependence. Censored likelihood-based inference for the implied copula is feasible in moderate dimensions due to closed-form margins. The model is applied to oceanographic datasets with ambiguous true limiting dependence structure. If time allows, we also show how a bivariate restriction of this copula model may be used to dynamically capture the extremal dependence structure of financial data over time.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
4:20 PM to 4:45 PM	Chee Ming Ting Universiti Teknologi Malaysia	"COMMUNITY DETECTION IN TIME-VARYING BRAIN NETWORKS USING DYNAMIC STOCHASTIC BLOCK MODELS" Brain networks exhibit the property of modular community structure with highly inter-connected nodes within a same module, but sparsely connected between different modules. Recent neuroimaging studies also suggest dynamic changes in brain connectivity over time. We present a statistical approach based on dynamic stochastic block models (SBM) to characterize changes in community structure of the brain functional networks inferred from neuroimaging time series data. The dynamic SBM is a non-stationary extension combining a static SBM with a Markov process to allow for temporal evolution of the community membership of nodes and the network connectivity. The model is formulated into a state-space form with sequential estimation of the time-varying parameters by Kalman filtering. We further partition the time-evolving community structure into a finite number of repeating regimes or states using K-means clustering and hidden Markov models. The method is applied to resting-state and task-based functional magnetic resonance imaging (fMRI) data to detect dynamic reconfiguration of network structure of the brain.
4:45 PM to 5:10 PM	Weixuan Zhu Xiamen University	"A BAYESIAN NONPARAMETRIC SPIKED PROCESS PRIOR FOR DYNAMIC MODEL SELECTION" In many applications, investigators consider processes that vary in space and time, with the goal of identifying temporally persistent and spatially localized departures of those processes from a baseline or "normal" behavior. In this manuscript, we propose a Bayesian nonparametric model selection approach for the analysis of spatio-temporal data, which takes into account the non-exchangeable nature of measurements collected over time and space. More specifically, a zero-inflated conditionally identically distributed (CID) species sampling prior is used to model temporal dependence in the selection, by borrowing information across time and assigning data to clusters associated to either a null or an alternate process. Spatial dependences are accounted for by means of a Markov random field (MRF) prior, which allows to inform the selection based on inferences conducted at nearby locations. We investigate the performances of our model by means of a simulation study and an application to a disease surveillance problem, for detecting outbreaks of pneumonia and influenza (P&I) mortality in the continental United States. We show how the proposed modeling framework compares favorably with respect to commonly adopted threshold methods for detecting outbreaks over time and also to recent proposals modeling more complex Markov switching dependences.
IP45	**Recent Advances in Random Matrix and Related Fields, Friday, 29 June 2018, 10:30 AM to 12:10 PM, ② LT28**	
Org / Chair	Zhonggen Su, Zhejiang University / Zhonggen Su, Zhejiang University	
10:30 AM to 10:55 AM	Zhigang Bao Hong Kong University of Science and Technology	#N/A
10:55 AM to 11:20 AM	Dangzheng Liu University of Science and Technology of China	"SINGULAR VALUE STATISTICS FOR THE SPIKED ELLIPTIC GINIBRE ENSEMBLE" The complex elliptic Ginibre ensemble is a random matrix model interpolating between the Gaussian unitary ensemble and the Ginibre ensemble. Its eigenvalues form a determinantal point process in the complex plane, however, until recently its singular values had been proved to build a Pfaffian point process (Kanazawa and Kieburg, arXiv:1804.03985). In this talk we consider an extended interpolating ensemble with singular values changing from the elliptic ensemble to the spiked Wishart ensemble. We prove that the singular values still build a Pfaffian point process and further derive a double contour integral for the correlation kernel. This is based on joint work with Yanhui Wang, Henan University.
11:20 AM to 11:45 AM	Dong Wang National University of Singapore	#N/A
11:45 AM to 12:10 PM	Lun Zhang Fudan University	#N/A
IP46	**Recent Advances in Statistical Machine Learning, Wednesday, 27 June 2018, 8:30 AM to 10:10 AM, ⑥ LT34**	
Org / Chair	Will Wei Sun, University of Miami / Jingfei Zhang, University of Miami	
8:30 AM to 8:55 AM	Lilun Du Hong Kong University of Science and Technology	"DYNAMIC TRACKING AND SCREENING IN MASSIVE DATASTREAMS" Technological advances have led to the emergence of an increasing number of applications requiring analysis of large-scale datasreams, consisting of multiple indefinitely long and time-evolving sequences. Consequently, there is a growing need for statistical methodologies to allow for performing inferential tasks in an online manner, continuously revising the model to reflect the current status of the underlying process. In particular, we are interested in constructing a large-scale dynamic tracking and screening (DTS) system capable of rapidly identifying irregular individual streams whose behavioral patterns deviate from that of majority. By fully exploiting the sequential feature of datastreams, we first develop a robust estimation approach under a framework of varying coefficient model. The scheme naturally accommodates unequally-spaced design points and updates estimates as new data arrive without the need to store an ever-increasing data history. A data-driven choice of optimal tuning parameter is accordingly proposed. Then, we suggest a new model-specification test tailored to streaming environment. The resulting DTS scheme is able to adapt time-varying structures appropriately, track changes in the underlying models, and hence maintain high identification accuracy in detecting irregular individuals. Moreover, we derive the asymptotic properties of the procedure and investigate its finite sample performance by means of a simulation study and a real data example.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:55 AM to 9:20 AM	Jiashi Feng National University of Singapore	"EMPIRICAL RISK LANDSCAPE ANALYSIS FOR UNDERSTANDING DEEP NEURAL NETWORKS" In this talk, I will introduce our recent work that aims to provide comprehensive landscape analysis of empirical risk in deep neural networks (DNNs), including the convergence behavior of its gradient, its stationary points and the empirical risk itself to their corresponding population counterparts, which reveals how various network parameters determine the convergence performance. Moreover, we prove the one-to-one correspondence of the non-degenerate stationary points between the empirical and population risks and provide convergence guarantee for each pair. We also establish the uniform convergence of the empirical risk to its population counterpart and further derive the stability and generalization bounds for the empirical risk. In addition, we analyze these properties for deep nonlinear neural networks with sigmoid activation functions. We prove similar results for convergence behavior of their empirical risk gradients, non-degenerate stationary points as well as the empirical risk itself.
9:20 AM to 9:45 AM	Xin He City University of Hong Kong	"SCALABLE KERNEL-BASED VARIABLE SELECTION WITH SPARSISTENCY" Variable selection is central to high-dimensional data analysis, and various algorithms have been developed. Ideally, a variable selection algorithm shall be flexible, scalable, and with theoretical guarantee, yet most existing algorithms cannot attain these properties at the same time. In this article, a three-step variable selection algorithm is developed, involving kernel-based estimation of the regression function and its gradient functions as well as a hard thresholding. Its key advantage is that it assumes no distributional model, admits general predictor effects, allows for scalable computation, and attains desirable asymptotic sparsistency. The proposed algorithm can be adapted to any reproducing kernel Hilbert space (RKHS) with different kernel functions, and can be extended to interaction selection with slight modification. Its computational cost is only linear in the data dimension, and can be further improved through parallel computing. The sparsistency of the proposed algorithm is established for general RKHS under mild conditions, including linear and Gaussian kernels as special cases. Its effectiveness is also supported by a variety of simulated and real examples.
9:45 AM to 10:10 AM		#N/A
IP47	**Bayesian Modelling and Computational Methods, Wednesday, 27 June 2018, 1:30 PM to 3:10 PM, ⑥ LT34**	
Org / Chair	Linda Tan, National University of Singapore / Linda Tan, National University of Singapore	
1:30 PM to 1:55 PM	Nial Friel University College Dublin	"INFORMED SUB-SAMPLING MCMC: APPROXIMATE BAYESIAN INFERENCE FOR LARGE DATASETS" This paper introduces a framework for speeding up Bayesian inference conducted in presence of large datasets. We design a Markov chain whose transition kernel uses an unknown fraction of fixed size of \nthe available data that is randomly refreshed throughout the algorithm. Inspired by the Approximate Bayesian Computation (ABC) literature, the subsampling process is guided by the fidelity to the observed data, as measured by summary statistics. The resulting algorithm, Informed Sub-Sampling MCMC (ISS-MCMC), is a generic and flexible approach which, contrary to existing scalable methodologies, preserves the simplicity of the \nMetropolis-Hastings algorithm. Even though exactness is lost, ie, the chain distribution approximates the posterior, we study and quantify theoretically this bias and show on a diverse set of examples that it yields excellent performances when the computational budget is limited. If available and cheap to compute, we show that setting the summary statistics as the maximum likelihood estimator is supported by theoretical arguments.
1:55 PM to 2:20 PM	David Nott National University of Singapore	"GAUSSIAN VARIATIONAL APPROXIMATION WITH FACTOR COVARIANCE STRUCTURE" Variational approximations have the potential to scale Bayesian computations to large datasets and highly parameterized models. Gaussian variational approximations are popular but can be computationally burdensome when the dimension of the model parameter is high. To circumvent this problem we consider a factor covariance structure. Stochastic gradient ascent methods are described for efficient implementation with gradient estimates obtained using the "reparametrization trick". We illustrate the methodology for robust P-spline regression and some high-dimensional logistic regression models. This is joint work with Victor Ong and Michael Smith.
2:20 PM to 2:45 PM	Michael Smith University of Melbourne	"IMPLICIT COPULAS FROM BAYESIAN REGULARIZED REGRESSION SMOOTHERS" We show how to extract the implicit copula of a response vector from a Bayesian regularized regression smoother with Gaussian disturbances. The copula can be used to compare smoothers that employ different shrinkage priors and function bases. We illustrate with three popular choices of shrinkage priors --- a pairwise prior, the horseshoe prior and a g prior augmented with a point mass as employed for Bayesian variable selection --- and both univariate and multivariate function bases. The implicit copulas are high-dimensional, have flexible dependence structures that are far from that of a Gaussian copula, and are unavailable in closed form. However, we show how they can be evaluated by first constructing a Gaussian copula conditional on the regularization parameters, and then integrating over these. Combined with non-parametric margins the regularized smoothers can be used to model the distribution of non-Gaussian univariate responses conditional on the covariates. Efficient Markov chain Monte Carlo schemes for evaluating the copula are given for this case. Using both simulated and real data, we show how such copula smoothing models can improve the quality of resulting function estimates and predictive distributions.
2:45 PM to 3:10 PM	Linda Tan National University of Singapore	"DYNAMIC DEGREE-CORRECTED BLOCKMODELS FOR SOCIAL NETWORKS: A NONPARAMETRIC APPROACH" We propose a nonparametric approach to model social networks using degree-corrected stochastic blockmodels. The static model consists of a stochastic blockmodel formulated using a probit regression and popularity parameters are incorporated to account for degree heterogeneity. Dirichlet processes are used to detect community structure and induce clustering in the popularity parameters. We further extend the static model to dynamic networks. We derive Gibbs samplers for posterior inference under a Bayesian approach. The models are illustrated using real social networks.
IP48	**Statistical Inference for Random Processes: Theory and Applications , Friday, 29 June 2018, 8:30 AM to 10:10 AM, ⑦ LT31**	
Org / Chair	Masayuki Uchida, Osaka University / Masayuki Uchida, Osaka University	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:30 AM to 8:55 AM	Yuta Koike University of Tokyo	"ASYMPTOTIC MIXED NORMALITY OF REALIZED COVARIANCE IN HIGH-DIMENSIONS" The asymptotic mixed normality of the realized covariance matrix for a multi-dimensional continuous semimartingale observed at a high-frequency is established, where the dimension may be much larger than the sample size. More precisely, a mixed-normal approximation of the error distribution in terms of the Kolmogorov distance is shown. The proof is based on a variant of the Chernozhukov-Chetverikov-Kato theory on high-dimensional central limit theorems for sums of independent random vectors, where the theory accommodates for random asymptotic covariance matrices. Application to testing the residual sparsity of a continuous-time factor model is presented.
8:55 AM to 9:20 AM	Hiroki Masuda Kyushu University	"OPTIMAL STABLE REGRESSION" We consider parametric inference for a regression model driven by a symmetric non-Gaussian stable Levy process, when the defining parameters: the stability index (activity index) and the scale in the noise term, and the (auto)regression coefficients in the trend term are all unknown. It is assumed that the process is observed at high frequency over a fixed time period. Then, we present a new asymptotically optimal result which properly takes the asymptotic degeneracy of the model into account and cannot be shared with cases of Gaussian driving noise. The result is valid in a unified manner for a wide variety of covariate processes; for example, they may be periodic and/or non-stationary. Also discussed are model selection issues and some possible model extensions.
9:20 AM to 9:45 AM	Masayuki Uchida Osaka University	"HYBRID MULTI-STEP ESTIMATORS FOR NON-ERGODIC DIFFUSION TYPE PROCESSES FROM REDUCED DATA" We deal with hybrid multi-step estimation of the unknown parameter for non-ergodic diffusion type processes from discrete observations. For both ergodic and non-ergodic diffusion type processes, there are many researches on asymptotic theory for the quasi-maximum likelihood estimator (QMLE) and the Bayes type estimator (BE), see, for example, Yoshida (1992, JMA; 2011, AISM), Genon-Catalot and Jacod (1993, AIHP), Kessler (1997, SJS), Uchida and Yoshida (2012, SPA; 2013, SPA; 2014, SISP). For ergodic diffusion processes, Kamatani and Uchida (2015, SISP) proposed the multi-step ML type estimator with the initial Bayes type estimator, which is called the hybrid multi-step (HMS) estimator with the initial Bayes type estimator. Kamatani, Nogita and Uchida (2016, BIC) considered the HMS estimator with the initial Bayes type estimator for non-ergodic diffusion type processes. Although the HMS estimator has good performance, there is a disadvantage that we spend much time computing the initial BE based on high frequency big data. Recently, Kaino, Uchida and Yoshida (2017, BIC) studied the HMS estimator with the initial Bayes type estimator based on the reduced data obtained from full sampled data of ergodic diffusion processes from the viewpoint computational cost. In this talk, applying the result of Kaino, Uchida and Uchida (2017, BIC) to non-ergodic diffusion type processes, we propose the HMS estimators with the initial Bayes type estimators based on reduced data obtained from full high frequency data of non-ergodic diffusion processes and show that the HMS estimator has asymptotic mixed normality and convergence of moments. This is a joint work with Yusuke Kaino.
9:45 AM to 10:10 AM	Nakahiro Yoshida University of Tokyo	"PARTIAL QUASI LIKELIHOOD ANALYSIS" The quasi likelihood analysis (QLA) is a framework of statistical inference for stochastic processes (Yoshida 2011). Through QLA, it is possible to systematically obtain limit theorems and precise tail probability estimates of the associated QLA estimators such as quasi maximum likelihood estimator (QMLE), quasi Bayesian estimator (QBE) and various adaptive estimators. Thanks to the universality, the QLA theory is rapidly expanding the range of its applications. Once a polynomial type large deviation (PLD) inequality, that is an essential ingredient for QLA, is established, we can obtain a very strong mode of convergence of the random field and the associated estimators. However, boundedness of high order of moments of functionals can be an obstacle to application. Easy to understand is a situation where there are two components (L,U) of stochastic processes and U has a fast mixing rate but L has a slow mixing rate. Suppose that the Rosenthal inequality can control the moments of a functional of U but cannot control the moments of a functional of L. In this situation, we cannot directly apply the present QLA theory. This problem is solved by introducing the partial quasi likelihood analysis (PQLA, Yoshida 2017). If there is a partial mixing structure (Yoshida 2004) in that U possesses a very good mixing rate conditionally on L, then we can apply a conditional version of QLA for given L to recover control of estimators.
IP49	**Nonlinear Regression with Non-Stationary Time Series, Thursday, 28 June 2018, 10:30 AM to 12:10 PM, ⑨ S16-06118**	
Org / Chair	Qiying Wang, University of Sydney / Qiying Wang, University of Sydney	
10:30 AM to 10:55 AM	Sanjaya Dissanayake University of Sydney	"NONPARAMETRIC ESTIMATION OF A COINTEGRATING REGRESSION MODEL WITH ENDOGENEITY AND LONG MEMORY" This paper explores the implications of adopting a feasible estimation method for linear and nonlinear cointegration models with multiple integrated time series comprising endogeneity and long memory. Theoretical concepts linked with model specification testing and parametric consistency of the observed nonlinear process are provided as a major contribution. The methodology comprises of utilizing nonlinear least squares as an initial step to estimate the parameters. It is followed by employing an optimal kernel method to nonparametrically estimate the integrable component of the non- parametric part from the residuals generated through the first step. Model testing and parametric consistency is followed by a simulation study of finite samples proving that the introduced nonlinear model outperforms the corresponding linear member in terms of bias as another contribution of the paper. For empirical illustrations the assessment technique is applied to inflation and gross domestic product per capita income data of Australia, Brazil, India and Russia.
10:55 AM to 11:20 AM	Ke Zhu University of Hong Kong	"MODEL CHECKS FOR NONLINEAR COINTEGRATING REGRESSION" Using the marked empirical processes, this paper develops a test of parametric specification in a nonlinear cointegrating regression model. Unlike the kernel-smoothed U-statistic considered in Gao et al. (2009) and Wang and Phillips (2012), our new test statistic avoids the use of bandwidth, which has some advantages for practitioners. Simulations and a real data example show that our new test has a good finite sample performance. Other contributions of this paper are to provide a rigorous proof on weak convergence for a class of martingales and construct a simulated estimator of the limiting null distribution, which are interesting in their own rights.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
11:20 AM to 11:45 AM	Jiangyan Peng University of Electronic Science and Technology of China	"WEAK CONVERGENCE TO STOCHASTIC INTEGRALS UNDER PRIMITIVE CONDITIONS IN NONLINEAR ECONOMETRIC MODELS" Limit theory with stochastic integrals plays a major role in time series econometrics. In earlier contributions on weak convergence to stochastic integrals, the literature commonly uses martingale and semi-martingale structures. Liang, Phillips, Wang, andWang (2016) (see alsoWang (2015), Chap. 4.5) currently extended weak convergence to stochastic integrals by allowing for a linear process or a Î±-mixing sequence in innovations. While these martingale, linear process and Î±-mixing structures have wide relevance, they are not sufficiently general to cover many econometric applications that have endogeneity and nonlinearity. This paper provides new conditions for weak convergence to stochastic integrals. Our frameworks allow for long memory processes, causal processes, and near-epoch dependence in innovations, which have applications in a wide range of econometric areas such as TAR, bilinear, and other nonlinear models.
11:45 AM to 12:10 PM	Hao Wu University of Sydney	#N/A
IP50	**Statistical Testing in Non- and Semiparametric Models, Friday, 29 June 2018, 10:30 AM to 12:10 PM, ⑤ LT33**	
Org / Chair	Jing Wang, University of Illinois at Chicago / Jing Wang, University of Illinois at Chicago	
10:30 AM to 10:55 AM	Jing Wang University of Illinois at Chicago	"TESTING LOGISTIC REGRESSION MODELS WITH FREE-KNOT SPLINES" A computational study of bootstrap confidence bands based on a free-knot spline regression is explored for the Generalized Linear Model in this paper. In free-knot spline regression, the knot locations as additional parameters offers greater flexibility and the potential to better account for rapid shifts in slope and other important structures in the target function. However, in ``freeing'' up the knots, the search for optimal solutions becomes very complicated. In particular, the ``lethargy'' property in the objective function results in many local optima with replicate knot solutions. To prevent solutions with identical knots, a penalized Quasi-likelihood estimating equation is proposed that relies on both a Jupp transformation of knot locations and an added penalty on solutions with small minimal distances between knots. Focusing on logistic regression for binary outcome data, a parametric bootstrap is used to study the variability of the proposed estimator and to construct confidence and for the unknown form of the logistic regression link function.
10:55 AM to 11:20 AM	Lijian Yang Tsinghua University	#N/A
11:20 AM to 11:45 AM	Yuanyuan Zhang Tsinghua University	"ORACALLY EFFICIENT INFERENCE FOR TIME VARYING ARCH MODEL" We relax this assumption by
11:45 AM to 12:10 PM	Roger Zoh Texas A&M University	"A POWERFUL BAYESIAN TEST FOR EQUALITY OF MEANS IN HIGH DIMENSIONS" We develop a Bayes factor based testing procedure for comparing two population means in high dimensional settings. In `large-p-small-n' settings, Bayes factors based on proper priors require eliciting a large and complex $p \times p$ covariance matrix, whereas Bayes factors based on Jeffrey's prior suffer the same impediment as the classical Hotelling T^{2} test statistic as they involve inversion of ill-formed sample covariance matrices. To circumvent this limitation, we propose that the Bayes factor be based on lower dimensional random projections of the high dimensional data vectors. We choose the prior under the alternative to maximize the power of the test for a fixed threshold level, yielding a restricted most powerful Bayesian test (RMPBT). The final test statistic is based on the ensemble of Bayes factors corresponding to multiple replications of randomly projected data. We show that the test is unbiased and, under mild conditions, is also locally consistent. We demonstrate the efficacy of the approach through simulated and real data examples.
IP51	**New Developments of High-Dimenisional Complex Data Analysis, Tuesday, 26 June 2018, 1:30 PM to 3:10 PM, ⑤ LT33**	
Org / Chair	Lan Wang, University of Minnesota / Kean Ming Tan, Unniversity of Minnesota	
1:30 PM to 1:55 PM	Xingdong Feng Shanghai University of Finance and Economics	#N/A
1:55 PM to 2:20 PM	Kean Ming Tan Unniversity of Minnesota	"A NONCONVEX APPROACH TO SPARSE GENERALIZED EIGENVALUE PROBLEM AND ITS APPLICATION TO STATISTICS" Sparse generalized eigenvalue problem (GEP) plays a pivotal role in a large family of high-dimensional learning tasks, including sparse Fisher's discriminant analysis, canonical correlation analysis, and sufficient dimension reduction. Sparse GEP involves solving a non-convex optimization problem. Most of the existing methods and theory in the context of specific statistical models that are special cases of the GEP require restrictive structural assumptions on the input matrices. In this paper, we propose a two-stage computational framework to solve the sparse GEP. At the first stage, we solve a convex relaxation of the sparse GEP. Taking the solution as an initial value, we then exploit a nonconvex optimization perspective and propose the truncated Rayleigh flow method (Rifle) to estimate the leading generalized eigenvector and show that it converges linearly to a solution with the optimal statistical rate of convergence. Theoretically, our method significantly improves upon the existing literature by eliminating structural assumptions on the input matrices. To achieve this, our analysis involves two key ingredients: (i) a new analysis of the gradient based method on nonconvex objective functions, as well as (ii) a fine-grained characterization of the evolution of sparsity patterns along the solution path. Thorough numerical studies are provided to validate the theoretical results.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
2:20 PM to 2:45 PM	Runze Li Pennsylvania State University	"LINEAR HYPOTHESIS TESTING FOR HIGH DIMENSIONAL GENERALIZED LINEAR MODELS" This paper is concerned with testing linear hypotheses in high-dimensional generalized linear models. To deal with linear hypotheses, we first propose constrained partial regularization method and study its statistical properties. We further introduce an algorithm for solving regularization problems with folded-concave penalty functions and linear constraints. To test linear hypotheses, we propose a partial penalized likelihood ratio test, a partial penalized score test and a partial penalized Wald test. We show that the limiting null distributions of these three test statistics are chi-square distribution with the same degrees of freedom, and under local alternatives, they asymptotically follow non-central chi-square distributions with the same degrees of freedom and noncentral parameter, provided the number of parameters involved in the test hypothesis grows to infinity at a certain rate. Simulation studies are conducted to examine the finite sample performance of the proposed tests. Empirical analysis of a real data example is used to illustrate the proposed testing procedures.
2:45 PM to 3:10 PM	Zehua Chen National University of Singapore	"PRINCIPLE OF CORRELATION AND FEATURE SELECTION IN HIGH-DIMENSIONAL MODELS WITH COMPLEX STRUCTURES" The intrinsic mechanism of feature selection for linear models is correlation. For instance, in various sequential methods, the features are selected according to their Pearson's correlation coefficients with the residual of a current model, in regularized least squares methods, at a fixed value of the penalty parameter, the active set is indeed the set of features whose absolute Pearson's correlation coefficients with the response exceed a certain threshold. We refer to this mechnism as the principle of correlation. In this talk, we explore sequential feature selectoin procedures in high-dimensional models with complex structures based on the principle of correlation. Specifically, we consider two such models: (i) a multi-response model where both the response variables and covariates have group structures, and (ii) an uni-response interaction model. For the first model, we develop a sequential canonical correlation search method. For the second model, we develop a sequential interaction groupselection method. We provide the asymptotic properties of these methods as well as the simulation studies comparing these methods with other existing approaches. The sequential methods based on the principle of correlation can achieve selection consistency under meld conditions. The simulation studies demonstrate that they have an edge over the other methods across a comprehensive simulation settings.
IP52	**Selective Inference with Knockoffs, Friday, 29 June 2018, 1:30 PM to 3:10 PM, ⑥ LT34**	
Org / Chair	Asaf Weinstein, Stanford University / Asaf Weinstein, Stanford University	
1:30 PM to 1:55 PM	Lucas Janson Harvard University	"SHOULD WE MODEL X IN HIGH-DIMENSIONAL INFERENCE?" For answering questions about the relationship between a response variable Y and a set of explanatory variables X, most statistical methods focus their assumptions on the conditional distribution of Y given X (or Y | X for short). I will describe some benefits of shifting those assumptions from the conditional distribution Y | X to the joint distribution of X, especially for high-dimensional data. First, modeling X can lead to assumptions that are more realistic and verifiable. Second, there are substantial methodological payoffs in terms of much greater flexibility in the tools an analyst can bring to bear on their data while also being guaranteed exact (non-asymptotic) inference. I will briefly mention some of my recent and ongoing work on methods for high-dimensional inference that model X instead of Y, as well as some challenges and interesting directions for the future.
1:55 PM to 2:20 PM	Yoshimasa Uematsu University of Southern California	"IPAD: A FACTOR APPROACH TO HIGH-DIMENSIONAL KNOCKOFFS INFERENCE" We consider the problem of false discovery rate (FDR) control for variable selection in high-dimensional factor models, where the association structure of covariates is modeled using a latent factor model popularly exploited in economics and finance. To achieve the FDR control, we adapt the general framework of model-X knockoffs in Candes et al. (2017) and suggest the new method of intertwined probabilistic factors decoupling (IPAD) for knockoffs inference in high-dimensional factor models. Our new method and work differ from the existing literature in at least four aspects: 1) when constructing the knockoff variables, we estimate the covariate distribution from data while Candes et al. (2017) assumed that it is known; 2) our procedure does not need any sample splitting; 3) we provide theoretical justifications on the asymptotic FDR control; and 4) we establish theory for the power analysis. Our simulation examples and real data analysis also demonstrate that our newly suggested method has appealing finite-sample performance with desired interpretability compared to some widely used methods such as the Lasso and factor autoregressive models.
2:20 PM to 2:45 PM	Asaf Weinstein Stanford University	#N/A
2:45 PM to 3:10 PM	Jia Zhou University of Science and Technlogy of China	"UNIFORM KNOCKOFF FILTER FOR HIGH-DIMENSIONAL CONTROLLED GRAPH RECOVERY" Learning the dependence structures in high-dimensional graphical models is of fundamental importance in many contemporary applications. Despite the fast growing literature, procedures with both guaranteed false discovery rate (FDR) control and high power for recovering the graphical structures remain largely unexplored. We develop a new method called uniform knockoff filter that controls the overall FDR in graph recovery based on control variables. Instead of controlling the FDR in a nodewise way, the proposed procedure utilizes a uniform threshold for the statistics based on a large-scale mixture of regression models associated with the graph, which enjoys not only theoretical guarantees of FDR control but also significantly higher power. Furthermore, a scalable implementation approach is developed for the uniform knockoff filter such that all control variables can be generated through a single estimation of the overall graphical structure. Numerical studies verify that our method outperforms existing approaches in power with FDR control.
IP53	**Stochastic Processes and their Applications, Tuesday, 26 June 2018, 1:30 PM to 3:10 PM, ⑦ LT31**	
Org / Chair	Dongsheng Wu, University of Alabama in Huntsville / Dongsheng Wu, University of Alabama in Huntsville	
1:30 PM to 1:55 PM	Cheng Ouyang University of Illinois at Chicago	#N/A

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
1:55 PM to 2:20 PM	Haihua Shi Nanjing University of Science and Technology	#N/A
2:20 PM to 2:45 PM	Yinghui Shi Jiangsu Normal University	"INTRINSIC ULTRACONTRACTIVITY OF LAPLACIAN OR FRACTIONAL LAPLACIAN PERTURBED BY NON-LOCAL OPERATORS" The intrinsic ultracontractivity for the subprocesses of two kinds of Markovian processes upon leaving any bounded open set in Rd will be presented in this talk, where the processes are associated with operators that can be seen as Laplacian or fractional Laplacian operators with lower order perturbations. (Joint work with Bingji Yi and Renming Song.)
2:45 PM to 3:10 PM	Dongsheng Wu University of Alabama in Huntsville	"SOME DIMENSION RESULTS OF GAUSSIAN RANDOM FIELDS" In this talk, we present some Hausdorff and packing dimension results for image and graph sets of Gaussian random fields. In particular, for anisotropic Gaussian fields, we extend the notion of Hausdorff dimension, packing dimension and packing dimension profile to (anisotropic) metric spaces. We believe that these extensions are of independent interest. This talk is based on joint works with Y. Du, A. Estrade. J. Miao and Y. Xiao.
IP54	**Random Fields on Manifolds and Statistical Inference, Friday, 29 June 2018, 3:30 PM to 5:10 PM, ⑥ LT34**	
Org / Chair	Yimin Xiao, Michigan State Univeristy / Dongsheng Wu, University of Alabama in Huntsville	
3:30 PM to 3:55 PM	Xiaohong Lan University of Science and Technology of China	"SAMPLE PATH PROPERTIES OF THE SOLUTION TO A STOCHASTIC HEAT EQUATION DRIVEN BY A SPACE-TIME NOISE" Consider a stochastic heat equation driven by a fractional noise in time with an isotropic spherical structure. We show that its solution exists in some space and gives an explicit form in terms of spectral representation. Moreover, we establish results on the variogram and strong local nondeterminism properties of this solution, which yield its exact modulus of continuity.
3:55 PM to 4:20 PM	Anatoliy Malyarenko Mälardalen University	#N/A
4:20 PM to 4:45 PM	Wei Ying Wu National Dong Hwa University	"TAIL ESTIMATION FOR THE CROSS-SPECTRAL DENSITY OF A BIVARIATE STATIONARY GAUSSIAN RANDOM FIELD" In this work, we consider the bivariate stationary Gaussian random field model. Under some assumptions on high-frequency behavior of (cross-)spectral densities, a modified multivariate Whittle likelihood are proposed to estimate parameters, which dominate tail behaviors of the (cross-) spectral densities. The related asymptotic properties of the proposed estimators are discussed under the regular infill sampling. The simulation studies are also presented for the finite sample situation.
4:45 PM to 5:10 PM		#N/A
IP55	**Emerging Statistical Inferences in Complex Data Analysis , Wednesday, 27 June 2018, 10:30 AM to 12:10 PM, ⑦ LT31**	
Org / Chair	Minge Xie, Rutgers University / Zijian Guo, Rutgers University	
10:30 AM to 10:55 AM	Jan Hannig University of North Carolina at Chapel Hill	"DEEP FIDUCIAL INFERENCE" R. A. Fisher, the father of modern statistics, developed the idea of fiducial inference during the first half of the 20th century. While his proposal led to interesting methods for quantifying uncertainty, other prominent statisticians of the time did not accept Fisher's approach as it became apparent that some of Fisher's bold claims about the properties of fiducial distribution did not hold up for multi-parameter problems. Beginning around the year 2000, the authors and collaborators started to re-investigate the idea of fiducial inference and discovered that Fisher's approach, when properly generalized, would open doors to solve many important and difficult inference problems. They termed their generalization of Fisher's idea as generalized fiducial inference (GFI). The main idea of GFI is to carefully transfer randomness from the data to the parameter space using an inverse of a data generating equation without the use of Bayes theorem. The resulting generalized fiducial distribution (GFD) can then be used for inference. After more than a decade of investigations, the authors and collaborators have developed a unifying theory for GFI, and provided GFI solutions to many challenging practical problems in different fields of science and industry. Overall, they have demonstrated that GFI is a valid, useful, and promising approach for conducting statistical inference. In this talk we discuss how certain computations within generalized fiducial framework can be made using a deep neural network. The resulting approximation to the fiducial distribution is termed deep fiducial distribution (DFD). We conclude by summarizing several difficult open problems related to this approach. (Joint work with Gang Li)
10:55 AM to 11:20 AM	George Michailidis University of Florida	"AN INTEGRATIVE STATISTICAL FRAMEWORK FOR MULTI-MODAL OMICS DATA" It is becoming increasingly common for patients to be profiled across multiple molecular compartments -genomic, transcriptomic, proteomic, metabolomic, etc. We develop a framework that leverages recent developments in the estimation of high-dimensional multi-layered graphical models that provide insights on regulatory mechanisms across molecular compartments (layers), as well as on molecular interactions within each layer and are also capable of accommodating outcome variables such as disease risk, or patient survival times. We discuss algorithmic issues, establish theoretical properties of the estimates, address testing issues for comparing two disease conditions and apply them to real data from a study focusing on hyperurecimia patients.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
11:20 AM to 11:45 AM	Dongchu Sun University of Missouri and East China Normal University	"BAYESIAN ANALYSIS OF COVARIANCE MATRIX OF MULTIVARIATE" Objective Bayesian analysis for unknown covariance matrix of multivariate normal has received a lot of attention in the last two decades. Discovering an unconstrained and statistically interpretable covariance matrix is still an open problem in statistics In this talk, we present a new class of the shrinkage inverse Wishart priors including both inverse Wishart and reference priors as special cases. It is showed that for a modified reference prior, the unknown covariance matrix can be estimated even when there are only three observations. A new MCMC algorithm is implemented for the class of priors and is shown to be more efficient comparing with the existing hit-and-run and Metropolis-Hasting algorithms.
11:45 AM to 12:10 PM	Dan Yang Rutgers University	"FACTOR MODEL FOR HIGH-DIMENSIONAL TENSOR TIME SERIES" Modern data collection capabilities have lead to massive quantity of time series. Large tensor (or multi-dimensional array) data are now routinely collected in a wide range of applications, and often such observations are taken over time, forming tensor time series. Although it is natural to turn the tensor observations into a long vector then use standard vector time series models, it is often the case that the fibers of a tensor represent different sets of information that are closely interplayed. We propose novel factor models for tensor time series that maintain and utilize the tensor structure to achieve greater dimensional reduction as well as more interpretable results. Estimation procedures and their theoretical properties are further investigated and demonstrated with simulated and real examples.
IP56	**Recent Advances in Functional Data and Complex Structures, Thursday, 28 June 2018, 8:30 AM to 10:10 AM, ⑧ MD9-0102GH**	
Org / Chair	Fang Yao, Peking University and University of Toronto / Fang Yao, Peking University and University of Toronto	
8:30 AM to 8:55 AM	Huang Hui Peking University	#N/A
8:55 AM to 9:20 AM	Yehua Li University of California, Riverside	"COVARIANCE ESTIMATION AND PRINCIPAL COMPONENT ANALYSIS FOR SPATIALLY DEPENDENT FUNCTIONAL DATA" We consider spatially dependent functional data collected under a geostatistics setting, where locations are sampled from a spatial point process and a random function is observed at each location. Observations on each function are made on discrete time points and contaminated with measurement errors. The error process at each location is modeled as a non-stationary temporal process rather than white noise. Under the assumption of spatial isotropy, we propose a tensor product spline estimator for the spatio-temporal covariance function. If a coregionalization covariance structure is further assumed, we propose a new functional principal component analysis method that borrow information from neighboring functions. Under a unified framework for both sparse and dense functional data, where the number of observations per curve is allowed to be of any rate relative to the number of functions, we develop the asymptotic convergence rates for the proposed estimators. The proposed methods are illustrated by simulation studies and a motivating example of the home price-rent ratio data in the New York metropolitan area.
9:20 AM to 9:45 AM	Yufeng Liu University of North Carolina at Chapel Hill	#N/A
9:45 AM to 10:10 AM	Yichao Wu University of Illinois at Chicago	"FUNCTIONAL ENVELOPE FOR MODEL-FREE SUFFICIENT DIMENSION REDUCTION" In this article, we introduce the functional envelope for sufficient dimension reduction and regression with functional and longitudinal data. Functional sufficient dimension reduction methods, especially the inverse regression estimation family of methods, usually involve solving generalized eigenvalue problems and inverting the infinite dimensional covariance operator. With the notion of functional envelope, essentially a special type of sufficient dimension reduction subspace, we develop a generic method to circumvent the difficulties in solving the generalized eigenvalue problems and inverting the covariance directly. We derive the geometric characteristics of the functional envelope and establish the asymptotic properties of related functional envelope estimators under mild conditions. The functional envelope estimators have shown promising performance in extensive simulation studies and real data analysis.
IP57	**Recent Progress of Statistical Inference for Economics and Social Science, Thursday, 28 June 2018, 10:30 AM to 12:10 PM, ④ LT26**	
Org / Chair	Kyusang Yu, KonKuk University / Kyusang Yu, KonKuk University	
10:30 AM to 10:55 AM	Eun Ryung Lee Sungkyunkwan University	"PARTIES IN POLITICAL SPECTRUM: STATISTICAL MODELING AND ESTIMATION OF POSITIONS FROM TEXTS" One central task in comparative politics is to locate party positions in a certain political space. For this purpose, several empirical methods have been proposed using text as data sources. In general, the analysis of texts to extract information is a difficult task. Its data structure is very complex and political texts usually contain a large number of words such that a simultaneous analysis of word counts becomes challenging. In this paper, we consider Poisson models for each word count simultaneously and provide a statistical analysis suitable for political text data. In particular, we allow for multi-dimensional party positions and develop a data-driven way of determining the dimension of positions. Allowing for multi-dimensional political positions gives new insights in the evolution of party positions and helps our understanding of a political system. Additionally, we consider a novel model which allows the political lexicon to change over time and develop an estimation procedure based on LASSO and fused LASSO penalization techniques to address high-dimensionality via significant dimension reduction. The latter model extension gives more insights into the potentially changing use of words by left and right-wing parties over time. Furthermore, the procedure is capable to identify automatically words having a discriminating effect between party positions. To address the potential dependence structure of the word counts over time, we included integer-valued time series processes into our modeling approach and we implemented a suitable bootstrap method to construct confidence intervals for the model parameters. We apply our approach to party manifesto data from German parties over all seven federal elections after German reunification. The data studies confirm that our procedure is robust, runs stable and leads to meaningful and interpretable results.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
10:55 AM to 11:20 AM	Youngkyung Lee Kangwon National University	"GENERALISED ADDITIVE DEPENDENCY INFLATED MODELS INCLUDING AGGREGATED COVARIATE" Let us assume that X, Y and U are observed and that the conditional mean of U given X and Y can be expressed via an additive dependency of X, lambda(X) Y and X + Y for some unspecified function lambda. This structured regression model can be transferred to a hazard model or a density model when applied on some appropriate grid, and has important forecasting applications via structured marker dependent hazards models or structured density models including age-period-cohort relationships. The structured regression model is also important when the severity of the dependent variable has a complicated dependency on waiting times X, Y and the total waiting time X + Y. In case the conditional mean of U approximates a density, the regression model can be used to analyse the age-period-cohort model, also when exposure data are not available. In case the conditional mean of U approximates a marker dependent hazard, the regression model introduces new relevant age-period-cohort time scale interdependencies in understanding longevity. A direct use of the regression relationship introduced in this paper is the estimation of the severity of outstanding liabilities in non-life insurance companies. The technical approach taken is to use B-splines to capture the underlying one-dimensional unspecified functions. It is shown via finite sample simulation studies and an application for forecasting future asbestos related deaths in the UK that the B-spline approach works well in practice. Special consideration has been given to ensure identifiability of all models considered.
11:20 AM to 11:45 AM	Christoph Rothe University of Mannheim	"ESTIMATING DERIVATIVES OF FUNCTION-VALUED PARAMETERS IN A CLASS OF MOMENT CONDITION MODELS" We develop a general approach to estimating the derivative of a function-valued parameter $\theta_o(u)$ that is identified for every value of u as the solution to a moment condition. This setup in particular covers interesting models for conditional distributions, such as quantile regression or distribution regression. Exploiting that $\theta_o(u)$ solves a moment condition, we obtain an explicit expression for its derivative from the Implicit Function Theorem, and then estimate the components of this expression by suitable sample analogues. The last step generally involves (local linear) smoothing. Our estimator can then be used for a variety of purposes, including the estimation of conditional density functions, quantile partial effects, and the distribution of bidders' valuations in structural auction models.
11:45 AM to 12:10 PM	Melanie Schienle Karlsruhe Institute of Technology	"DETERMINATION OF VECTOR ERROR CORRECTION MODELS IN HIGH DIMENSIONS" We provide a shrinkage type methodology which allows for simultaneous model selection and estimation of vector error correction models (VECM) when the dimension is large and can increase with sample size. Model determination is treated as a joint selection problem of cointegrating rank and autoregressive lags under respective practically valid sparsity assumptions. We show consistency of the selection mechanism by the resulting Lasso-VECM estimator under very general assumptions on dimension, rank and error terms. Moreover, with computational complexity of a linear programming problem only, the procedure remains computationally tractable in high dimensions. We demonstrate the effectiveness of the proposed approach by a simulation study and an empirical application to recent CDS data after the financial crisis.
IP58	**Stein's Method, Friday, 29 June 2018, 10:30 AM to 12:10 PM, ⑥ LT34**	
Org / Chair	Zhuosong Zhang, University of Melbourne / Zhuosong Zhang, University of Melbourne	
10:30 AM to 10:55 AM	Xiao Fang Chinese University of Hong Kong	"LIMIT THEOREMS WITH RATE OF CONVERGENCE UNDER SUBLINEAR EXPECTATIONS" Under the sublinear expectation $\mathbbm{E}[\cdot]:=\sup_{\theta\in \Theta} E_\theta[\cdot]$ for a given set of linear expectations $\{E_\theta: \theta\in \Theta\}$, we establish a new law of large numbers and a new central limit theorem with rate of convergence. We present some interesting special cases and discuss a related statistical inference problem. We also give an approximation and a representation of the G-normal distribution, which was used as the limit in Peng (2007)'s central limit theorem, in a probability space. This is joint work with Shige Peng, Qi-Man Shao and Yongsheng Song.
10:55 AM to 11:20 AM	Adrian Roellin National University of Singapore	"ERROR BOUNDS IN LOCAL LIMIT THEOREMS USING STEIN'S METHOD" We provide a general result for bounding the difference between point probabilities of integer supported distributions and the translated Poisson distribution, a convenient alternative to the discretized normal. We illustrate our theorem in the context of the Hoeffding combinatorial central limit theorem with integer valued summands, of the number of isolated vertices in an Erdos-Renyi random graph, and of the Curie-Weiss model of magnetism, where we provide optimal or near optimal rates of convergence in the local limit metric. In the Hoeffding example, even the discrete normal approximation bounds seem to be new. The general result follows from Stein's method, and requires a new bound on the Stein solution for the Poisson distribution, which is of general interest. This is joint work with A. B. Barbour and N. Ross; to appear in Bernoulli.
11:20 AM to 11:45 AM	Aihua Xia University of Melbourne	#N/A
11:45 AM to 12:10 PM	Zhuosong Zhang University of Melbourne	"CRAMER-TYPE MODERATE DEVIATION FOR UNBOUNDED EXCHANGEABLE PAIRS" Moderate deviation reflects the relative approximation error and plays an important role in probability theory. As a new method, Stein's method was connected with the moderate deviation by Chen, Fang and Shao (2013), who proved a Cram\'er-type moderate deviation under some boundedness assumptions. In this talk, we establish a Cram\'er-type moderate deviation of normal approximation for unbounded exchangeable pairs using Stein's method. We also consider some applications to the combinatorial central limit theorem and the Curie-Weiss model. This is a joint work with Professor Qi-Man Shao.
IP59	**New Developments in High-Dimensional and Complex Data Analysis, Wednesday, 27 June 2018, 3:30 PM to 5:10 PM, ⑦ LT31**	
Org / Chair	Yichuan Zhao, Georgia State University / Yichuan Zhao, Georgia State University	

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
3:30 PM to 3:55 PM	Lei Huang Southwestern Jiaotong University	"WEIGHTED VOLUME UNDER THE THREE-WAY RECEIVER OPERATING CHARACTERISTIC SURFACE" In medical researches, three-way ROC analysis has been widely concerned. This paper generalizes the volume under the surface(VUS) of three-way ROC analysis to weighted volume under the surface(WVUS) by introducing a weight function into the integration formula. This generalization is far from trivial, it practically allows researchers to calculate the diagnostic accuracy of some tests while some high probabilities of correct classification for certain classes are ensured. Theoretically, the asymptotic properties of the corresponding nonparametric and parametric estimators of WVUS have been well derived in this paper, which could provide some statistical inferences when comparing different diagnostic accuracies. Substantial simulations have been conducted to show the usefulness of the proposed estimators and also to show the necessity of considering WVUS. The applicability and feasibility of the proposed method has also been verified by a real data analysis about liver cancer .
3:55 PM to 4:20 PM	Binyan Jiang Hong Kong Polytechnic University	"PENALIZED INTERACTION ESTIMATION FOR ULTRAHIGH DIMENSIONAL QUADRATIC REGRESSION" Quadratic regression goes beyond linear model by simultaneously including main effects and interactions between the covariates. The problem of interaction estimation in high dimensional quadratic regression has received extensive attention in the past decade. In this article we introduce a novel method which allows us to estimate the main effects and interactions separately. Unlike existing methods for ultrahigh dimensional quadratic regressions, our proposal does not require the widely used heredity assumption. In addition, our proposed estimates have explicit formulas and obey the invariance principle at the population level. We estimate the interactions of matrix form under penalized convex loss function. The resulting estimates are shown to be consistent even when the covariate dimension is an exponential order of the sample size. We develop an efficient ADMM algorithm to implement the penalized estimation. This ADMM algorithm fully explores the cheap computational cost of matrix multiplication and hence is much more efficient than existing penalized meth- ods under heredity constraints. We demonstrate the promising performance of our proposal through extensive numerical studies.
4:20 PM to 4:45 PM	Bin Nan University of California, Irvine	"LARGE MATRIX ESTIMATION FROM HIGH-DIMENSIONAL TEMPORALLY DEPENDENT DATA" We consider the estimation of large covariance and precision matrices from high-dimensional sub-Gaussian or heavier-tailed observations with slowly decaying temporal dependence. The temporal dependence is allowed to be long-range so with longer memory than those considered in the current literature. We show that several commonly used methods for independent observations can be applied to the temporally dependent data. The rates of convergence are obtained, and the properties of sparsistency and sign-consistency are also established. A gap-block cross-validation method is proposed for the tuning parameter selection, which performs well in simulations. As a motivating example, we study the brain functional connectivity using resting-state fMRI time series data with long-range temporal dependence. This is a joint work with Hai Shu.
4:45 PM to 5:10 PM	Yichuan Zhao Georgia State University	"EMPIRICAL LIKELIHOOD FOR THE BIVARIATE SURVIVAL FUNCTION UNDER UNIVARIATE CENSORING" The bivariate survival function plays an important role in multivariate survival analysis. Using the idea of influence functions, we develop empirical likelihood confidence intervals for the bivariate survival function in the presence of univariate censoring. It is shown that the empirical log-likelihood ratio has an asymptotic standard chi-squared distribution with one degree of freedom. A comprehensive simulation study shows that the proposed method outperforms both the traditional normal approximation method and the adjusted empirical likelihood method in most cases. The Diabetic Retinopathy Data are analyzed for illustration of the proposed procedure. The talk is based on joint work with Haitao Huang.
IP60	**Recent Advances in Statistical Methods for Big and Complex Data, Friday, 29 June 2018, 1:30 PM to 3:10 PM, ⑦ LT31**	
Org / Chair	Lihui Zhao, Northwestern University / Zhigang Li, Dartmouth College	
1:30 PM to 1:55 PM	Zhigang Li Dartmouth College	"CONDITIONAL REGRESSION FOR MICORBIOME DATA" The human microbiome plays critical roles in human health and has been linked to many diseases. While advanced sequencing technologies can characterize the composition of the microbiome in unprecedented detail, it remains challenging to disentangle the complex interplay between human microbiome and disease risk factors due to the complicated nature of microbiome data. Excessive numbers of zero values, high dimensionality, the hierarchical phylogenetic tree and compositional structure are compounded and consequently make existing methods inadequate to appropriately address these issues. We propose a multivariate two-part zero-inflated logistic normal (MZILN) model to analyze the association of disease risk factors with individual microbial taxa and overall microbial community composition. This approach can naturally handle excessive numbers of zeros and the compositional data structure with the discrete part and the logistic-normal part of the model. For parameter estimation, an estimating equations approach is employed that enables us to address the complex inter-taxa correlation structure induced by the hierarchical phylogenetic tree structure and the compositional data structure. This model is able to incorporate standard regularization approaches to deal with high dimensionality. Simulation shows that our model outperforms existing methods. Our approach is also compared to others using the analysis of real microbiome data.
1:55 PM to 2:20 PM	Chunlin Wang Xiamen University	"SEMIPARAMETRIC INFERENCE FOR MULTIPLE SAMPLES WITH EXCESS ZERO OBSERVATIONS" A non-standard, but not uncommon, situation is to observe multiple samples of nonnegative data which have a high proportion of zeros. This talk will focus on some important, and fundamental, statistical inference problems for such data structure. A unique feature of the target populations is that the distribution of each sample is characterized by a non-standard mixture of a singular distribution at zero and a skewed nonnegative component. We propose modelling the nonnegative components using a semiparametric, multiple-sample, density ratio model. Under this semiparametric setup, we first propose an empirical likelihood ratio (ELR) test for homogeneity and show that this ELR has a χ^2-type limiting distribution under the homogeneous null hypothesis. A conststent nonparametric bootstrap procedure is further proposed to calibrate the finite sample distribution of the ELR. Then we also develop a new ELR statistic for making inferences on the means of multiple nonnegative distributions, and show that this ELR has a χ^2-type limiting distribution under a general null hypothesis. Some simulation and real data analysis results will also be presented.

INVITED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
2:20 PM to 2:45 PM	Rui Wang Harvard Medical School	"A STOCHASTIC SECOND-ORDER GENERALIZED ESTIMATING EQUATIONS APPROACH FOR ESTIMATING ICCS" Cluster randomized trials are well-suited to evaluate the community-level effectiveness of intervention strategies against infectious diseases. Design and analysis of cluster randomized trials must account for the intraclass correlation coefficient, the degree to which individuals within a community are more similar to one another than to individuals in other communities. This quantity can be estimated via second-order generalized estimating equations (GEE2). However, its use has been hindered by the computational difficulties. We develop stochastic algorithms for solving GEE2, which alleviates reliance on parameter starting values and provides substantially faster and higher convergence rates that the widely used deterministic root-solving methods. In addition, we build upon GEE2 in the setting of informative missing data, for which we incorporate a second-order inverse-probability weighting scheme and doubly robust estimating equations that guard against partial model misspecification. We highlight the need to model correlation among missing indicators in cluster data settings.
2:45 PM to 3:10 PM	Tao Yu National University of Singapore	"MSL COMPONENT DENSITY ESTIMATION IN MIXTURE MODELS WITH KNOWN MIXING PROPORTIONS" Mixture models appear in many research areas. In genetic and epidemiology applications, sometimes the mixture proportions may vary but are known. For such data, the existing methods for the underlying component density estimation may produce undesirable results: negative values in the density estimates. In this paper, we propose a maximum smoothed likelihood method to estimate these component density functions. The proposed estimates maximize a smoothed log likelihood function which can inherit all the important properties of probability density functions. A majorization-minimization algorithm is suggested to compute the proposed estimates numerically. We show that, starting from any initial value, the algorithm converges. Furthermore, we establish the asymptotic convergence rate of the {L_1 errors} of our proposed estimators. Our method provides a general framework for dealing with many similar mixture model problems. An adaptive procedure is suggested for choosing the bandwidths in our estimation procedure. Simulation studies show that the proposed method is very promising and can be much more efficient than the existing method in terms of {the L_1 errors}. A malaria data application shows the advantages of our method over others.
IP61	**High-Dimensional Statistical Theory, Wednesday, 27 June 2018, 3:30 PM to 5:10 PM, ⑧ MD9-0102GH**	
Org / Chair	Wen Xin Zhou, University of California, San Diego / Wen Xin Zhou, University of California, San Diego	
3:30 PM to 3:55 PM	Heather Battey Imperial College London	"LARGE NUMBERS OF EXPLANATORY VARIABLES" The lasso and its variants are powerful methods for regression analysis when there are a small number of study individuals and a large number of potential explanatory variables. There results a single model, while there may be several simple representations equally compatible with the data. I will outline a different approach, whose aim is essentially a confidence set of models. A probabilistic assessment of the method will be given. The talk is based on joint work with David R Cox.
3:55 PM to 4:20 PM	Guang Cheng Purdue University	#N/A
4:20 PM to 4:45 PM	Fengnan Gao Fudan University	"MAXIMUM LIKELIHOOD ESTIMATION OF SUBLINEAR PREFERENTIAL ATTACHMENT MODELS AND JANSON'S URNS" The preferential attachment (PA) network is a popular way of modeling the social networks, the collaboration networks and etc. The PA network model is an evolving network model where new nodes keep coming in. When a new node comes in, it establishes only one connection with an existing node. The random choice on the existing node is via a multinomial distribution with probability weights based on a preferential function f on the degrees. f maps the natural numbers to the positive real line and is assumed apriori non-decreasing, which means the nodes with high degrees are more likely to get new connections, i.e. "the rich get richer". Under sublinear parametric assumptions on the PA function, we proposed the maximum likelihood estimator on f. We show that the MLE yields optimal performance with the asymptotic normality results. Despite the optimal property of the MLE, it depends on the history of the network evolution, which is often difficult to obtain in practice. To avoid such shortcomings of the MLE, we propose the quasi maximum likelihood estimator (QMLE), a history-free remedy of the MLE. To prove the asymptotic normality of the QMLE, a connection between the PA model and Svante Janson's urn models is exploited. This is partially joint work with Aad van der Vaart.
4:45 PM to 5:10 PM	Kengo Kato Tokyo University	"RANDOMIZED INCOMPLETE U-STATISTICS IN HIGH DIMENSIONS" This paper studies inference for the mean vector of a high-dimensional U-statistic. In the era of Big Data, the dimension d of the U-statistic and the sample size n of the observations tend to be both large, and the computation of the U-statistic is prohibitively demanding. Data-dependent inferential procedures such as the empirical bootstrap for U-statistics is even more computationally expensive. To overcome such computational bottleneck, incomplete U-statistics obtained by sampling fewer terms of the U-statistic are attractive alternatives. In this paper, we introduce randomized incomplete U-statistics with sparse weights whose computational cost can be made independent of the order of the U-statistic. We derive non-asymptotic Gaussian approximation error bounds for the randomized incomplete U-statistics in high dimensions, namely in cases where the dimension d is possibly much larger than the sample size n, for both non-degenerate and degenerate kernels. In addition, we propose novel and generic bootstrap methods for the incomplete U-statistics that are computationally much less-demanding than existing bootstrap methods, and establish finite sample validity of the proposed bootstrap methods. The proposed bootstrap methods are illustrated on the application to nonparametric testing for the pairwise independence of a high-dimensional random vector under weaker assumptions than those appearing in the literature.

TOPIC CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
TCP02	**Nonparametric and Semiparametric Methods for Complex Data, Friday, 29 June 2018, 3:30 PM to 5:10 PM, ⑦ LT31**	
Org / Chair	Toshio Honda, Hitotsubashi University / Toshio Honda, Hitotsubashi University	
3:30 PM to 3:50 PM:	Rizky Reza Fauzi Kyushu University	"NEW TYPE OF GAMMA KERNEL DENSITY ESTIMATOR" We discuss a new kernel type estimator for nonnegatively supported density function f(x), using the pdf of gamma distribution. Chen (2000) introduced two gamma kernels which, even though free of boundary bias, have an order of convergence of variances O(1/(nâˆšh)) in the interior and O(1/(nh)) near boundary. Hence, under some conditions for x and h, Chen (2000) showed his estimators having O(n^(-4/5)) for the optimal mean squared error. Here we use another gamma density as a kernel function, and modify it with the expansions of exponential and logarithmic functions. Our modified gamma kernel density estimator is not only free of boundary bias, but the variance also does not depend on 1/âˆšx, with order O(n^(-1)h^(-1/4)) in the interior and O(n^(-1)h^(-3/4)) in the boundary region. Furthermore, the optimal mean squared error has order O(n^(-8/9)) in the interior and O(n^(-8/11)) in the boundary. Some simulation results to show the method's performance will be presented in the last part.
3:50 PM to 4:10 PM:	Toshio Honda Hitotsubashi University	"ADAPTIVELY WEIGHTED GROUP LASSO FOR SEMIPARAMETRIC QUANTILE REGRESSION MODELS" We propose an adaptively weighted group Lasso procedure for simultaneous variable selection and structure identification for varying coefficient quantile regression models and additive quantile regression models with ultra-high dimensional covariates. Under a strong sparsity condition, we establish selection consistency of the proposed Lasso procedure when the weights therein satisfy a set of general conditions. This consistency result, however, is reliant on a suitable choice of the tuning parameter for the Lasso penalty, which can be hard to make in practice. To alleviate this difficulty, we suggest a BIC-type criterion, which we call high-dimensional information criterion (HDIC), and show that the proposed Lasso procedure with the tuning parameter determined by HDIC still achieves selection consistency. Our simulation studies support strongly our theoretical findings. This is joint work with Ching-Kang Ing and Wei-Ying Wu.
4:10 PM to 4:30 PM:	Kanta Naito Shimane University	"REGRESSION WITH STAGEWISE MINIMIZATION ON RISK FUNCTION" This paper is concerned with a curve estimation based on empirical risk minimization. The estimator is composed as a convex combination of words (learners) in a dictionary. A word is selected in each step of the proposed stagewise algorithm, which minimizes a certain divergence measure. A non-asymptotic error bound of the estimator is developed, and it is shown that the error bound becomes sharp as the number of iterations of the algorithm increases. A simulation study and real data example confirm the performance of the estimator.
4:30 PM to 4:50 PM:	Hidetoshi Matsui Shiga University	"QUADRATIC REGRESSION FOR FUNCTIONAL-ON-FUNCTION MODELS" We consider the problem of constructing a regression model where both a predictor and a response are given as functions. We extend the functional linear model to the quadratic one, where the quadratic term also takes into account the interaction between the argument of the functional data. We assume that the predictor and the coefficient functions are expressed by basis expansions, and then parameters included in the model are estimated by the penalized likelihood method by assuming that the error function follows a Gaussian process. In order to select a regularization parameter included in the penalized likelihood method, we derive some model selection criteria by applying the result of information theoretic and Bayesian approach. We also apply the proposed method to the analysis of weather data, and then investigate what the results provides.
4:50 PM to 5:10 PM:	Ryota Yabe Shinshu University	"UNIFORM CONVERGENCE RATE OF KERNEL FUNCTIONS WITH TRANSFORMED NONSTATIONARY REGRESSOR" We provide the results of uniform convergence rate of a kernel function with a transformed nonstationary regressor to apply kernel estimation theory in semiparametric time series model. In semiparametric model, to develop parameter estimation theory based on kernel estimation, we need the results of uniform convergence rate with respect to parameters. (For example, See Andrews(1995) for details.) In time series literature, Hansen(2008) and Kristensen(2009) has derived uniform convergence rates of the kernel function with stationary data. In this talk, we consider the uniform convergence rate of the kernel function in a nonstationary process such as unit root or long memory. In econometrics, this process has been considered as representation of a stochastic trend that macroeconomic data possesses such as GDP and interest rate, so that the co-integrated model, which is a kind of a linear model based on nonstationary processes, has been studied and applied for various directions. Since the beginning of study for the nonlinear co-integrated model by Park and Phillips(2001), linearity of co-integrated model is considered to be restrictive to capture movement and properties of economic data such as concavity of economic correlation and discrete data. In these days, the nonlinear co-integrated model is often non-parametrically treated and there are a lot of studies. Chan and Wang (2014) has derived uniform convergence rate of the kernel function to treat the kernel-based estimator for the covariate function. We extend their work to nonlinear transformed nonstationary data to deal with nonlinear times series data.
TCP03	**Stochastic Partial Differential Equations II, Tuesday, 26 June 2018, 1:30 PM to 3:10 PM, ⑧ MD9-0102GH**	
Org / Chair	Yaozhong Hu, University of Alberta at Edmonton / Yaozhong Hu, University of Alberta at Edmonton	
1:30 PM to 1:50 PM:	Yiming Jiang Nankai University	"STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS WITH SPACE-TIME FRACTIONAL NOISES" We consider a class of stochastic partial differential equations (SPDEs) driven by space-time fractional noises. We mainly study the existence and uniqueness of solutions to the SPDEs, and the regularity of the solutions.
1:50 PM to 2:10 PM:	Litan Yan Donghua University	#N/A
2:10 PM to 2:30 PM:		#N/A

TOPIC CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
2:30 PM to 2:50 PM:		#N/A
2:50 PM to 3:10 PM:		#N/A
TCP04	**Computational Statistics for High-Frequency Data, Wednesday, 27 June 2018, 1:30 PM to 3:10 PM, ⑨ S16-06118**	
Org / Chair	Hiroki Masuda, Kyushu University / Hiroki Masuda, Kyushu University	
1:30 PM to 1:50 PM:	Shoichi Eguchi Kyushu University, Osaka University	"IMPROVEMENT OF MODEL SELECTION FUNCTION IN YUIMA PACKAGE" There are several studies of model selection for stochastic differential equations (SDEs), which includes the contrast-based information criterion for ergodic diffusion processes (CIC) and the Schwarz type information criterion for locally asymptotically quadratic models (BIC, quasi-BIC). Based on these studies, we implemented the model selection function for SDEs in R package yuima. However, this function has some improvements. In this talk, we will first overview the model selection methods for SDEs and explain the improvements of model selection function.
1:50 PM to 2:10 PM:	Akitoshi Kimura University of Tokyo	"THE ASYMPTOTIC VARIANCE ESTIMATORS OF THE CORRELATION ESTIMATOR BETWEEN LATENT PROCESSES" In this talk, we treat a model in which the finite variation part of a two-dimensional semi-martingale is expressed by time-integration of latent processes. We propose a correlation estimator between the latent processes and show its consistency and asymptotic mixed normality. Moreover, we propose two types of estimators for asymptotic variance of the correlation estimator and show their consistency in a high frequency setting. Our model includes doubly stochastic Poisson processes whose intensity processes are correlated It\^o processes. We compare our asymptotic variance estimators based on the simulation of the doubly stochastic Poisson processes.
2:10 PM to 2:30 PM:	Daisuke Kurisu Tokyo Institute of Technology	"NONPARAMETRIC INFERENCE ON LEVY-DRIVEN ORNSTEIN-UHLENBECK PROCESSES UNDER DISCRETE OBSERVATIONS" In this paper, we study nonparametric inference for a stationary L\'evy-driven Ornstein-Uhlenbeck (OU) process $X = (X_{t})_{t \geq 0}$ with a compound Poisson subordinator. We propose a new spectral estimator for the L\'evy measure of the L\'evy-driven OU process X under macroscopic observations. We derive multivariate central limit theorems for the estimator over a finite number of design points. We also derive high-dimensional central limit theorems for the estimator in the case that the number of design points increases as the sample size increases. Building upon these asymptotic results, we develop methods to construct confidence bands for the L\'evy measure and propose a practical method for bandwidth selection.
2:30 PM to 2:50 PM:	Teppei Ogihara Institute of Statistical Mathematics	"LOCAL ASYMPTOTIC MIXED NORMALITY FOR INTEGRATED DIFFUSION PROCESSES" We study statistical inference for integrated diffusion processes and consider asymptotic properties of this model in a high-frequency limit. This model arises when we observe a process after passage through an electric filter, and is also related to modeling of the stochastic volatility in finance. Gloter and Gobet (2008) studied this model and showed the local asymptotic mixed normality (LAMN) when the latent diffusion process is one-dimensional. The LAMN property is important in asymptotic statistical theory and enables us to discuss the asymptotic efficiency of estimators. We extend their results of the LAMN property to multi-dimensional diffusion processes which may have a feedback from the integral process. Then we can apply these results to a Langevin equation which is a model for molecular motion. We also consider the construction of an efficient estimator. This is joint work with Masaaki Fukasawa in Osaka University.
2:50 PM to 3:10 PM:	Yuma Uehara Institute of Statistical Mathematics	"GAUSSIAN QUASI-LIKELIHOOD ESTIMATION FOR ERGODIC LEVY DRIVEN SDE UNDER MODEL MISSPECIFICATION" We consider the Gaussian quasi-likelihood estimation for ergodic Levy driven SDE under model misspecification. To handle the misspecification effect, we utilize the notion of "extended Poisson equation". Based on the tool, we present the tail probability estimates and asymptotic normality of Gaussian quasi maximum likelihood estimator.
TCP05	**Tensor Theory and its Applications, Tuesday, 26 June 2018, 3:30 PM to 5:10 PM, ⑧ MD9-0102GH**	
Org / Chair	Mitsuhiro Miyazaki, Kyoto University of Education / Mitsuhiro Miyazaki, Kyoto University of Education	
3:30 PM to 3:50 PM:	Deniz Akdemir StatGen Consulting	#N/A

TOPIC CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
3:50 PM to 4:10 PM:	Yoshitaka Nakajima Kyushu University	"ENGLISH PHONEMES IN A SPACE EXTRACTED FROM SPECTRAL CHANGES OF SPEECH IN TIME" Spectral changes of speech in time can be represented, preserving almost perfect intelligibility, by power fluctuations in narrow frequency bands simulating the function of the auditory periphery. Since 20 such narrow bands cover the frequency range 50-7000 Hz, speech signals can be expressed as movements in a 20-dimensional space. A 3-dimensional subspace can be extracted from this space by factor analysis (Ueda & Nakajima, 2017, Sci. Rep.). Here, the spectra of English phonemes, on average, make a configuration related closely to English phonology: Vowels (e.g., /Ã¦/, /ɪ/, or /aɪ/), sonorant consonants (e.g., /w/, /l/, or /n/), and obstruents (e.g., /v/, /s/, or /p/) were separated rather clearly (Nakajima, Ueda et al., 2017, Sci. Rep.). The phonemes formed a curve corresponding to the â€˜sonority scaleâ€™ as described in phonology. Sonority means the likeliness of a phoneme to be, or to be closer to, syllable nuclei. \n\nTwo questions arose. First, this sonority scale projected onto a plane has a peculiar shape like a dragon, and is different for each individual speaker. The curve has a head, the high end of sonority as of /É‘:/ or /ÊŒ/, and a tail, ending up with /Ê’/, /dÊ’/, /Êƒ/, and /tÊƒ/. We would like to know whether there is a way to express this shape mathematically. The mathematical function to be used should show linguistic reality. For example, the separation of vowels and consonants, observed in any language, should be reflected somehow in the mathematical expression. Second, the 3-dimensional configuration occupies only places very close to one of two orthogonal planes. We wonder whether it is possible to represent the English phonemes and the sonority scale in a way easier to grasp intuitively, constructing a new type of subspace like a folded plane if necessary. Additionally, we are interested in how English syllables and phrases are shaped as movements in this, or in a newly constructed, subspace. (Collaboration with Kazuo Ueda. Supported by JSPS.)\n
4:10 PM to 4:30 PM:	Yi Shen Beihang University	"THE TENSOR RANK OF TENSOR PRODUCT OF TWO THREE-QUBIT W STATES IS EIGHT" Tensor rank is related to various tasks in quantum information, such as state transformation, characterizing tripartite pure state and fermionic system. The relation is heavily decided by the multiplication of tensors. It is known that the tensor rank is not multiplicative under the tensor Kronecker product. Very recently it has been proved that the tensor rank is also not multiplicative under the tensor product. In particular it has been shown that the tensor rank of tensor product of two three-qubit W states is at most eight. We show that it is exactly eight in our paper. We also construct the upper bound of the tensor rank of tensor product of many three-qubit W states.
4:30 PM to 4:50 PM:	Taiji Suzuki University of Tokyo	"ESTIMATING NONLINEAR TENSOR PRODUCT IN INFINITE DIMENSIONAL SPACE BY KERNEL AND NEURAL NETWORK MODEL" In this talk, we deal with a nonparametric estimation problem in which the target function has a tensor decomposition structure. To estimate such a function, we consider two approached such as kernel method and deep learning. As for the kernel method, we consider the alternating minimization procedure. It is shown that the alternating minimization achieves the minimax optimal rate. As for the deep learning approach, we consider a fully connected deep neural network with ReLU activation function. By assuming the sparsity of the weights, it is shown that the deep learning approach can avoid curse of dimensionality even if the tensor rank of the target function is infinite.
4:50 PM to 5:10 PM:	Jun Tsuchida Tokyo University of Science	"TENSOR CANONICAL COVARIANCE ANALYSIS WITH QUANTIFICATION METHOD" In recent year, many multivariate analysis methods are extended from data matrix to data tensor. This extension is not only multivariate analysis for one data-set such as PCA, but multivariate analysis for many data-sets such as canonical correlation analysis. However, many extended methods assume that the data has no categorical variables. That is, data has only numerical data. Therefore, there is the case that it is not suitable to apply such tensor method to data that has categorical valuables, such as longitude questionnaire data. In this presentation, to overcome this problem, we extend canonical covariance analysis for tensor data with quantification method. In addition, even if data has categorical and numerical variable at the same time, the extended method is applied to tensor data. Although canonical covariance analysis is one of special case of canonical correlation analysis, canonical covariance analysis includes many multivariate analysis methods because the model of canonical covariance analysis is represented like canonical correlation analysis when given many data-sets.
TCP06	**Recent Advances in High-Dimensional Data Analysis, Friday, 29 June 2018, 8:30 AM to 10:10 AM, ⑧ MD9-0102GH**	
Org / Chair	Chi Tim Ng, Chonnam National University / Chi Tim Ng, Chonnam National University	
8:30 AM to 8:50 AM:	Young-Geun Choi Fred Hutchinson Cancer Research Center	#N/A
8:50 AM to 9:10 AM:	Woojoo Lee Inha University	"LIKELIHOOD-BASED INFERENCE FOR BOUNDS OF CAUSAL PARAMETERS" It is a common causal-inference problem that, even with theoretically infinite samples, we might be able to only provide bounds for the parameters of interest. This problem occurs naturally, for example, in estimating causal interaction between two risk factors and in estimating the average causal effect using the instrumental-variable or Mendelian randomization method. Current procedures include linear programming to get the estimated bounds, plus bootstrapping to get confidence intervals. We describe a likelihood-based procedure that automatically yields the interval estimate from the flat likelihood region, and show some theory that allows us to construct confidence intervals from this non-regular likelihood. Finally, we illustrate the procedure with examples from the estimation of causal interaction between two risk factors, and treatment effect under partial compliance.

TOPIC CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
9:10 AM to 9:30 AM:	Chi Tim Ng Chonnam National University	"EXTERIOR ALGORITHM FOR CHANGE POINT DETECTION AND ESTIMATION OF FINANCIAL TIME SERIES DATA" An efficient exterior point algorithm is proposed for smoothing and change point detection of financial time series data under the penalized likelihood approach. The proposed method has O(n) computational complexity and is applicable to a broad class of time series model proposed in Bardet and Wintenberger (2009) that encompasses ARMA-GARCH as a special case. Under certain conditions, the estimated model has piecewise constant coefficients. Asymptotic properties of the penalized likelihood estimators are established. The possibility of real-time forecasting that update the prediction within O(1) time upon arrival of new signal is also discussed.
9:30 AM to 9:50 AM:	Van Cuong Nguyen Chonnam National University	"TRUNCATING A PENALTY WITHOUT INTRODUCING NEW TUNING PARAMETERS" By introducing the idea of matching thresholding rule, it is illustrated that both bridge penalty and log penalty can be truncated so as to circumvent certain difficulties in numerical computation and ambiguities in the definition of local minimality due to the fact that these penalties have derivatives going to negative infinity at zero. Such difficulties hinder the applications of these penalties in statistics although it is reported in the literature that they\nallow recovery of sparse structure in the data under some conditions. The proposed truncation method does not involve any extra tuning parameters. Additionally, we show in the simulation studies that the penalties obtained by the proposed truncation method outperform many other state-of-art penalties, particularly in the variable selection problems with strongly correlated covariates. The one-one correspondence between the truncated penalties and their thresholding rules are also established.
9:50 AM to 10:10 AM:	Kaimeng Zhang Chonnam National University	"PRINCIPAL COMPONENT ANALYSIS BASED ON MAXIMUM SEPARATION OF CLUSTERS" In a questionnaire survey involving multiple sensitive attributes, both raw responses and statistically-inferred knowledge of a respondent should be protected. In this paper, we raise the point of view that the ability of classifying respondents is a breach of privacy. To evaluate such an ability, we propose new concepts of principal component analysis by looking for a linear combination of the raw data (or rotated data) that allows researchers to separate the respondents clearly into two groups. The inferred types of the respondents should be considered privacy too if they are closely related to certain sensitive attributes. Misclassification rate can therefore be used as a measure of privacy protection. An efficient dynamic programming algorithm with complexity O(n^2\max{n,p\}) is described, where n is the number of respondents and p is the number of attributes.
TCP07	**New Developments in Nonparametric Methods , Friday, 29 June 2018, 10:30 AM to 12:10 PM, ⑦ LT31**	
Org / Chair	Lan Xue, Oregon State University / Lan Xue, Oregon State University	
10:30 AM to 10:50 AM:	Yan Fang Shanghai University of Internaional Business and Economics	"ROBUST ESTIMATION OF ADDITIVE FRONTIER FUNCTION WITH QUANTILE REGRESSION AND SHAPE CONSTRAINTS" When modeling the production frontier function, we are often concerned with high dimensionality and function shape constraints (e.g. monotonicity and convexity). However, most of the existing methods failed to address both of the concerns. In this paper, we propose a nonparametric quantile regression to estimate the flexible additive frontier, which deals with high dimensional data and shape constraints efficiently. With our two-step constrained polynomial B-spline method based on L1 optimization, we are able to scale up to very high dimension with satisfactory efficiency; also our method can work with various constraints and shows good resistance to outliers as well. Both the simulation study and a real data application show that the proposed estimator outperforms the existing ones, especially with the presence of outliers.
10:50 AM to 11:10 AM:	Takeru Matsuda University of Tokyo	"MINIMAX ADAPTIVE REDUCED-RANK REGRESSION" In this study, we consider simultaneous estimation of mean parameters in several Gaussian sequence models. This problem corresponds to nonparametric regression with several response variables. For mean parameters in one Gaussian sequence model, the blockwise James--Stein estimator is known to be minimax adaptive over Sobolev ellipsoids. On the other hand, as a natural extension of the James--Stein estimator, Efron and Morris proposed an empirical Bayes estimator of a normal mean matrix and proved its minimaxity. This estimator provides large risk reduction when the true value of the mean parameter has low rank. We introduce a multivariate version of Sobolev ellipsoids and show that the blockwise Efron--Morris estimator is minimax adaptive over them. In nonparametric regression with several response variables, the blockwise Efron--Morris estimator provides large risk reduction when the correlation between response variables is large. Such situation often occurs in real data like array signals, e.g., neuroimaging data and seismic data. Thus, the blockwise Efron--Morris estimator attains reduced-rank nonparametric regression.
11:10 AM to 11:30 AM:	Carmen Tekwe Texas A&M University	#N/A
11:30 AM to 11:50 AM:	Zhaoliang Wang Beijing University of Technology	"SPLINE ESTIMATOR FOR ULTRA-HIGH DIMENSIONAL PARTIALLY LINEAR VARYING COEFFICIENT MODELS" In this paper, we simultaneously study variable selection and estimation problems for sparse ultra-high dimensional partially linear varying coefficient models, where the number of variables in linear part can grow much faster than the sample size while many coefficients are zeros and the dimension of nonparametric part is fixed. We apply the B-spline basis to approximate each coefficient function. First, we demonstrate the convergence rates as well as asymptotic normality of the linear coefficients for the oracle estimator when the nonzero components are known in advance. Then we propose a nonconvex penalized estimator and derive its oracle property under mild conditions. Furthermore, we address issues of numerical implementation and of data adaptive choice of the tuning parameters. Some Monte Carlo simulations and an application to a breast cancer dataset are provided to corroborate our theoretical findings in finite samples.

TOPIC CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
11:50 AM to 12:10 PM:	Nan Zhang Fudan University	"ADAPTIVE BASIS SAMPLING FOR EXPONENTIAL FAMILY SMOOTHING SPLINES" Second-generation sequencing technologies have become the default method for genomics and epigenomics analysis. Second-generation sequencing technologies sequence tens of millions of DNA/cDNA fragments in parallel and one gets a sequence of short read counts along the genome. Effective extraction of signals in these short read counts is the key to the success of sequencing technologies. Nonparametric methods, in particular, smoothing splines, have been used extensively for modeling and processing single sequencing sample. However, nonparametric joint modeling of multiple second-generation sequencing samples is still lacking due to expensive computational cost. To achieve a scalable computation for large data sets with multivariate covariates, we develop an adaptive sampling method to select basis functions and construct a low-dimensional approximation of the estimates. Our asymptotic analysis shows such approximation converges to the true function at the same rate as full basis smoothing spline estimator. The empirical performance is demonstrated through both simulation studies and second-generation sequencing data examples.
TCP08	**Risk Predictions and Validations for Complex Diseases, Friday, 29 June 2018, 10:30 AM to 12:10 PM, ⑧ MD9-0102GH**	
Org / Chair	Hua Zhong, NYU School of Medicine / Hua Zhong, NYU School of Medicine	
10:30 AM to 10:50 AM:	Mengling Liu NYU School of Medicine	#N/A
10:50 AM to 11:10 AM:	Li-Xuan Qin Memorial Sloan Kettering Cancer Center	"ON THE USE OF CROSS VALIDATION FOR MOLECULAR CLASSIFICATION" Reproducibility of scientific experimentation has become a major concern, due to the perception that many published biomedical studies cannot be replicated. In this talk we draw attention to the connection between inflated over-optimistic findings and the use of cross-validation for error estimation in molecular classification studies, in the presence of confounding handling effects in the data. We demonstrate this important yet over-looked complication of cross validation using a unique pair of datasets on the same set of tumor samples. One dataset was collected with uniform handling to prevent handling effects; the other dataset was collected without uniform handling and exhibited handling effects. The paired datasets were used to estimate the biological effects of the samples and the handling effects of the arrays in the latter dataset, which were then used to simulate data using virtual re-hybridization following various array-to-sample assignment schemes. Our study showed that (1) cross-validation tended to under-estimate the error rate when the data possessed confounding handling effects, (2) depending on the relative amount of handling effects, normalization may further worsen the under-estimation of the error rate, (3) balanced assignment of arrays to comparison groups allowed cross-validation to provide an unbiased error estimate. Our study demonstrates the benefits of balanced array assignment for reproducible molecular classification and calls for caution on the routine use of data normalization and cross-validation in such analysis.
11:10 AM to 11:30 AM:	Pei Wang Icahn School of Medicine at Mount Sinai	#N/A
11:30 AM to 11:50 AM:	Hua Zhong NYU School of Medicine	#N/A
11:50 AM to 12:10 PM:		#N/A
TCP09	**Data Mining Technology for Big Data, Thursday, 28 June 2018, 8:30 AM to 10:10 AM, ⑩ S17-0405**	
Org / Chair	Jianwei Chen, San Diego State University and Huaqiao University / Wuliu Zhang, Huaqiao University	
8:30 AM to 8:50 AM:	Yueer Gao Huaqiao University	"RESEARCH ON THE CONVENTIONAL BUS LINE ADJUSTMENT OF RAIL TRANSIT CO-ROUTE SECTION BASED ON THE TRAFFIC CAPACITY CONSTRAINT OF BUS LANE"
8:50 AM to 9:10 AM:	Cheng Wang Huaqiao University	"AN ESSAY BASED ON INTEGRATION LEARNING"
9:10 AM to 9:30 AM:	Wenhua Wang Huaqiao University	"HOT RESEARCH TOPICS ON “ONE BELT AND ONE ROAD”: A KEY-WORD BASED STATISTICAL METHOD "
9:30 AM to 9:50 AM:	Wuliu Zhang Huaqiao University	"DATA MINING ANALYSIS OF CHINESE INTERNET LENDING MARKET INFORMATION"
9:50 AM to 10:10 AM:	Yongguang Zou Huaqiao University	"STUDY ON THE TIME-SPACE CHARACTERISTICS OF TOURISM SECURITY IN SOUTHEAST ASIA BY CHINESE CITIZENS -- BASED ON BIG DATA PERSPECTIVE."
TCP10	**Statistical Methods for Big Data and Their Applications, Thursday, 28 June 2018, 10:30 AM to 12:10 PM, ⑩ S17-0405**	
Org / Chair	Jianwei Chen, San Diego State University and Huaqiao University / Xindong Zhao, Huaqiao University	
10:30 AM to 10:50 AM:	Jianwei Chen San Diego State University and Huaqiao University	"SEMI-PARAMETRIC SPACE-TIME MODEL FOR FUNCTION DATA"

TOPIC CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
10:50 AM to 11:10 AM:	Ridong Hu Huaqiao University	"AN EMPIRICAL STUDY ON THE HETEROGENEITY OF CHINESE LISTED COMPANIES, INCENTIVE AND R&D INVESTMENT OF SENIOR EXECUTIVES."
11:10 AM to 11:30 AM:	Yonghong Xu Xiamen University	"MISMEASURE OF CONSUMER'S CONFIDENCE INDEX" I found the values of most of Consumer's Confidence Index were below critical point when economy was booming, and I defined CCI Bias as the gap between CCI I measured and the unknowable real CCI. Using consumption multiplier effect and ratchet effect, I proved the existence of CCI bias, and the empirical study also showed CCI bias when American was in an economic boom but Consumer Sentiment Index of University of Michigan was below 100 in most period. Then I suggested three models to adjust CCI bias, and using it to adjust CCI bias, the further empirical research found the adjusted Consumer Sentiment University of Michigan fitted American economy better than it did before.
11:30 AM to 11:50 AM:	Xindong Zhao Huaqiao University	"MODEL SELECTION INVOLVING MANY CANDIDATE VAR MODELS"
11:50 AM to 12:10 PM:		#N/A

Last updated
7/2/2018

CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
CP01	**Tuesday, 26 June 2018, 1:30 PM to 3:10 PM, ⑨ S16-06118**	
Chair	Miao Yu Tsai, National Changhua University of Education	
1:30 PM to 1:50 PM	Miao Yu Tsai National Changhua University of Education	"CONCORDANCE CORRELATION COEFFICIENTS ESTIMATED BY VARIANCE COMPONENTS FOR LONGITUDINAL DATA" The concordance correlation coefficient (CCC) is widely used to assess agreement between two observers for continuous responses. Further, the CCC is extended for measuring agreement with discrete data. This paper proposes a variance components (VC) approach that allows dependency between repeated measurements over time to assess intra-agreement for each observer and inter- and total agreement among multiple observers simultaneously under extended three-way generalized linear mixed-effects models (GLMMs) for longitudinal normal and Poisson data. Furthermore, we propose a weight matrix to compare with existing weight matrices. Simulation studies are conducted to compare the performance of the VC and generalized estimating equations approaches with different weight matrices for repeated measurements from longitudinal normal and Poisson data. In conclusion, the VC approach with consideration of the correlation structure of longitudinal repeated measurements gives satisfactory results with small mean square errors and nominal 95% coverage rates for all sample sizes.
1:50 PM to 2:10 PM	Ryoya Oda Hiroshima University	"A CONSISTENT VARIABLE SELECTION METHOD IN THE HIGH-DIMENSIONAL MULTIPLE RESPONSES LINEAR REGRESSION" We deal with the variable selection problem for selecting explanatory variables in the high-dimensional multiple responses linear regression with a normality assumption. As a such usual selection method, the method by using the values of the model selection criterion for all the combinations of explanatory variables is known. However, in recent years, it is necessary to analyze high-dimensional data such that the dimensions of response variables and explanatory variables are large. That is a problem that we cannot use the usual selection method because the method has enormous calculation cost for such data. Then, in this presentation, we propose the high-speed variable selection method even when the dimensions are large, and show the selection method has consistency.
2:10 PM to 2:30 PM	Shin Fu Tsai National Taiwan University	"COMPARING COEFFICIENTS IN GAUSSIAN MIXTURE REGRESSION MODELS" When observed data are collected from a population with several sub-populations and there is no information regarding the membership of each observation, Gaussian mixture regression models are often used for characterizing the relationship between response and some covariates of interest. In practice, it is often required to compare the regression coefficients between different subpopulations. In this talk, I will introduce a new method to address this practical issue. Based on the concept of generalized inference, a numerical algorithm will be presented for drawing realizations from the distribution of parameters, then the inference procedure can be implemented easily. Some real-life examples will be given for illustrating the proposed method, and computational issues will also be discussed.
2:30 PM to 2:50 PM	Inge Koch University of Adelaide	"FEATURE MAPS FOR HIGH-DIMENSIONAL BINARY BASED ON A DIFFERENCE IN PROPORTIONS OF OCCURRENCE " We consider high-dimensional data which are binary or can be transformed to binary data. Of special interest are data with a spatial distribution. For such data we propose a method for extracting and ranking variables based on their rate of occurrence and the difference in proportions of occurrence (DIPPS) of two non-overlapping set of observations and determine the number of top-ranked variables in a data-driven way. We summarise the 'best' extracted variables in a single DIPPS number. The method can be used on its own or to enhance clustered data as it provides more information than simple cluster membership. For spatial data, the DIPPS number provides a useful way of summarising the best variables in a single feature map which has a natural interpretation. We illustrate the performance of the DIPPS approach with applications to proteomics imaging mass spectrometry data from patients with ovarian cancer. We indicate how the method can be extended to high-dimensional spatial data with continuous variables.
2:50 PM to 3:10 PM	Jiancang Zhuang Institute of Statistical Mathematics	"DETECTION AND REPLENISHMENT OF MISSING DATA IN MARKED POINT PROCESSES" Records of geophysical events, such as earthquakes and volcanic eruptions, are usually modeled as marked point processes. However, these records often have missing data, resulting in underestimations of the corresponding hazards. This study proposes a fast approach for replenishing missing data in the records of temporal point processes with time-separable marks. The basis of this method is that, if such a point process is completely observed, it can be transformed into a homogeneous Poisson process on the unit square $[0,1]^2$ by a biscale empirical transformation. This approach includes three key steps: (1) Transform the process onto $[0,1]^2$ using the empirical transformation, and find a time-mark range that likely contains missing events; (2) Estimate a new empirical distribution function based on the data in the time-mark range in which the events are supposed to be completely observed; (3) Generate events in the missing region. In addition, we show the consistency and asymptotic normality of the empirical distribution estimated in the second step. Then, we test this method on a synthetic dataset, and apply it to the records of volcanic eruptions of the Hakone Volcano in Japan and to a dataset of the aftershock sequence following the 2008 Wenchuan Ms7.9 earthquake in Southwest China. The results show that this algorithm provides a useful way to estimate missing data and to replenish incomplete records of marked point processes. The replenished data will provide more robust estimates of the hazard function.
CP02	**Wednesday, 27 June 2018, 8:30 AM to 10:10 AM, ⑨ S16-06118**	
Chair	Takayuki Kawashima, Graduate University for Advanced Studies	

CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:30 AM to 8:50 AM	Takayuki Kawashima Graduate University for Advanced Studies	"ROBUST AND SPARSE REGRESSION IN GENERALIZED LINEAR MODEL BY STOCHASTIC OPTIMIZATION" The generalized linear model (GLM) plays a key role in regression analyses. In high-dimensional data, the sparse GLM has been used but it is not robust against outliers. Recently, the robust methods have been proposed for the specific example of the sparse GLM. Among them, we focus on the robust and sparse linear regression based on the γ-divergence. The estimator of the γ-divergence has strong robustness under heavy contamination. In this talk, we extend the robust and sparse linear regression based on the γ-divergence to the robust and sparse GLM based on the γ-divergence with a stochastic optimization approach in order to obtain the estimate. We adopt the randomized stochastic projected gradient descent as a stochastic optimization approach and extend the established convergence property to the classical first-order necessary condition. By virtue of the stochastic optimization approach, we can efficiently estimate parameters for very large problems. Particularly, we show the linear regression, logistic regression and Poisson regression with L_1 regularization in detail as specific examples of robust and sparse GLM. In numerical experiments and real data analysis, the proposed method outperformed comparative methods. This is a joint work with Prof. Hironori Fujisawa (ISM).
8:50 AM to 9:10 AM	Mineaki Ohishi Hiroshima University	"A FAST ALGORITHM FOR SOLVING MODEL SELECTION CRITERION MINIMIZATION PROBLEM IN GENERALIZED RIDGE" In this presentation, we deal with the optimization method of ridge parameters in a generalized ridge regression by minimizing a model selection criterion (MSC). The optimization methods based on minimizations of generalized C_p criterion and GCV criterion are fast because minimizers of the two criteria can be derived as closed forms. Although the methods are fast, they do not work well when the number of explanatory variables is larger than the sample size. Methods based on minimizations of other MSCs that will work well even when the number of explanatory variables is larger than the sample size are not fast because minimizers of the MSCs cannot be derived as closed forms. Even though a minimization problems of MSC cannot be solved analytically, we propose a fast optimization algorithm to minimize MSC by specifying small number of minimizer candidates.
9:10 AM to 9:30 AM	Kengo Fujisawa Tokyo University of Science	"ASYMMETRY MODEL BASED ON F-DIVERGENCE AND ORTHOGONAL DECOMPOSITION OF SYMMETRY FOR SQUARE TABLES" In a square contingency table with ordinal row and column classifications, we are interested in determining whether the structure in a contingency table is symmetric or asymmetric since the independence between the row and column classifications is unlikely to hold. Many statisticians have proposed a variety of symmetry and asymmetry models. We propose the asymmetry model based on the f-divergence which generalizes various symmetry and asymmetry models, and we give the theorem that the proposed model is the closest model to the symmetry model in terms of f-divergence under certain conditions. Caussinus (1965) proved the decomposition that the symmetry model holds if and only if both the quasi-symmetry model and the marginal homogeneity model hold. The decomposition may be useful to determine the reason for the poor fit when the symmetry model fits poorly. Hence, we are also interested in considering the decomposition of symmetry using the proposed model, and we partition test statistics for the symmetry model.
9:30 AM to 9:50 AM	Md. Abud Darda National University, Bangladesh	"A GENERALIZATION OF DIFFUSION MODELS IN TERMS OF LOGISTIC DISTRIBUTION FOR CUMULATIVE GROWTH" In the existing social dynamics, individuals share their evaluations of an innovation that results the diffusion speedy or delayed. The fundamental work done by Bass is an initial approach to describe the evolutionary pattern of diffusion of innovations within the existing social structure. Further modifications and considerations have been adopted later, to capture the local interventions and heterogeneity of related agents. Recent research on diffusion innovations is much concentrated to model individual level adoption rate and obtain the aggregate diffusion pattern, which can be discussed as agent based modelling or bottom to the top modelling approach. In this paper we have examined and show that the Bass model is a modified logistic one under a monomolecular seeding action governed by (1-e^(-θt)). A number of homogeneous and heterogeneous diffusion models can be obtained by further re-parameterization of Bass model considering local interventions and agents' heterogeneity. Therefore, it can be concluded that the diffusion model works as the accumulation with basic foundation in the logistic model. Individual evaluation performance and corresponding sharing tendency in the existing social network reforms the diffusion dynamics.
9:50 AM to 10:10 AM	Jieming Wang Beijing Institute of Technology	"HEAT KERNEL ESTIMATE FOR FINITE RANGE PROCESSES" In this paper, we consider a class of non-symmetric finite range non-local operator, which can be viewed as the stable like operator in the paper by Chen-Zhang with large jumps more than 1 removed. The aim of this paper is to establish a strong Markov process with transition density function associated with the operator and to give the two-sided estimates of the transition density function and its gradient estimate in $\R^d$ for $t\in (0,1)$.
CP03	**Friday, 29 June 2018, 1:30 PM to 3:10 PM, ⑨ S16-06118**	
Chair	Lu Hung Chen, National Chung-Hsing University	
1:30 PM to 1:50 PM	Lu Hung Chen National Chung-Hsing University	"STATISTICAL DOWNSCALING OF PM2.5 FORECASTS BY EOF-VARX MODELS" We suggest a strategy for downscaling PM2.5 forecasts that incorporates their spatial-temporal relationships. The strategy is based on empirical orthogonal functions (EOFs) and vector autoregressive model with exogenous variables (VARX). The spatial correlation of PM 2.5 observations and forecasts are captured by PCAs, and their temporal information are utilized by VARX models. We evaluate the strategy by some empirical studies in Taiwan. Results show that our EOF-VARX model provides much more accurate downscaled results in terms of root-mean-square error and correlation scores.

CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
1:50 PM to 2:10 PM	Junhyeon Kwon Seoul National University	"FREQUENCY-DOMAIN PRINCIPAL COMPONENT ANALYSIS WITH MISSING VALUES" Frequency-domain principal component analysis (FDPCA), known as dynamic PCA, was first proposed by Brillinger (1981) to decompose multivariate time series data into PC series. A main advantage of FDPCA is its capability of extracting crucial components from the data reflecting the serial dependence of the data. However, its beneficial property cannot be utilized once there exist missing values in the data, which should not be simply ignored in the process of estimating the spectral matrix. In this study, we suggest a new method of principal component analysis in the frequency domain when missing values are present. For this purpose, FDPCA is coupled with the Monte-Carlo Self-Consistent (MCSC) algorithm of Lee and Zhu (2009), which exploits the concept of self-consistency and uses an iterative procedure for the optimal imputation of missing values to estimate the spectral density. As a result, the PC series obtained by our method still retains good properties of FDPCA. The effectiveness of the method is shown by simulation and real data examples in various missing rate situations.
2:10 PM to 2:30 PM	Masaki Narukawa Okayama University	"EFFICIENT TAPERED SEMIPARAMETRIC ESTIMATION OF MULTIVARIATE FRACTIONAL PROCESSES" This paper considers the semiparametric estimation of multivariate fractional processes based on the tapered periodogram of the differenced series. We construct multivariate local Whittle estimators by incorporating the maximal efficient taper developed by Chen (2010). The proposed estimation method allows a wide range of potentially nonstationary long-range dependent series and is invariant to the presence of deterministic polynomial trends with the same extent of the differencing order. We establish the consistency and asymptotic normality of the estimators and show that the asymptotic variance is the same as that of the nontapered local Whittle estimation by increasing the order of a taper to infinity with a moderately slow rate. The finite sample behavior of the proposed estimators is examined by a Monte Carlo simulation.
2:30 PM to 2:50 PM	Shun Matsuura Keio University	"ESTIMATION OF PRINCIPAL POINTS FOR MULTIVARIATE LOCATION-SCALE-ROTATION FAMILIES AND APPLICATIONS" (This is a joint work with Professor Thaddeus Tarpey, Wright State University, USA.) A set of principal points of a probability distribution provides a set of representative points for the distribution and a generalization of the mean (one-point representation) to multiple points, which is useful especially for multivariate distributions and can be connected and applied to various areas such as principal component analysis, cluster analysis, functional data analysis, and statistical quality and process control fields. In practice, principal points may need to be estimated using a random sample from the distribution. In this talk, we discuss the estimation of principal points of multivariate distributions with location, scale, and rotation parameters including multivariate normal distributions, under the criterion of expected mean squared distance. The optimal estimators of principal points are derived and compared to nonparametric and maximum likelihood estimators. Some applications are also discussed.
2:50 PM to 3:10 PM	Xiaohui Chang Oregon State University	#N/A
CP04	**Wednesday, 27 June 2018, 10:30 AM to 12:10 PM, ⑨ S16-06118**	
Chair	Tomonari Sei, University of Tokyo	
10:30 AM to 10:50 AM	Tomonari Sei University of Tokyo	"AN OBJECTIVE GENERAL INDEX FOR SEPARATION" Consider a multivariate data, where every variable has the meaning that a larger value indicates higher score. Our aim is to separate the observations into two groups: one is at overall higher score and the other is not. The problem is unsupervised. Separation is typically based on a weighted sum of the variables, where weights are assumed to be positive. A naive weighting method is just to take the sum of standardized variables. We call it the simple sum method. However, we found a paradoxical phenomenon: under the simple sum method, a particular variable of the higher-score group may be less than that of the lower-score group. The phenomenon is not desirable since then a part of variables is not much taken into account. We provide an algorithm without such shortcomings as long as decision around borderline is randomized. The problem is reduced to a convex optimization problem with respect to the weight variables, and therefore effectively solved. Numerical examples are presented to illustrate the properties of the proposed method.
10:50 AM to 11:10 AM	Dustin Pluta University of California, Irvine	#N/A
11:10 AM to 11:30 AM	Michiko Okudo University of Tokyo	"PRIORS FOR THE FACTOR ANALYSIS MODEL BASED ON ITS GEOMETRIC STRUCTURE" Factor analysis is an important tool for analysis of multivariate data, especially in psychology. Factor analysis models are latent variable models, and the observation is divided into two parts, "common factor" and "specific factor", using latent variables. The maximum likelihood estimation of these models can be difficult when the dimension of common factor space is high. Hence Bayesian estimation methods have been studied. We propose new priors for factor analysis models. These priors take advantage of the model manifold's geometric structure.
11:30 AM to 11:50 AM	Yuling Jiao Zhongnan University of Economics and Law	#N/A
11:50 AM to 12:10 PM	Yoga Dwi Nugroho Sekolah Tinggi Ilmu Statistik, Jakarta	#N/A
CP05	**Wednesday, 27 June 2018, 3:30 PM to 5:10 PM, ⑨ S16-06118**	
Chair	Yugo Nakayama, University of Tsukuba	

CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
3:30 PM to 3:50 PM	Yugo Nakayama University of Tsukuba	"BIAS CORRECTION OF SUPPORT VECTOR MACHINE FOR HIGH-DIMENSIONAL DATA" In this talk, we consider the classification in High-dimension, low-sample-size (HDLSS) context. HDLSS data situations occur in many areas of modern science such as genetic microarrays, medical imaging, text recognition, finance, chemometrics, and so on. In the HDLSS context, Hall et al. (2008, JRSS), Chan and Hall (2009, Biometrika) and Aoshima and Yata (2014, AISM) considered distance-based classifiers. In the field of machine learning, there are many studies about the classification in the context of supervised learning. A typical method is the support vector machine (SVM). The SVM has versatility and effectiveness both for low-dimensional and high-dimensional data. It is known that the SVM is effective because the data points are linearly separable in high-dimensional space. Asymptotic properties of the linear SVM were investigated by Hall et al. (2005, JRSS), Chan and Hall (2009), Qiao and Zhang (2015, JMLR) and Nakayama et al. (2017, JSPI) in the HDLSS settings. Huang (2017, JMLR) investigated it in high-dimension, large-sample-size settings. In this talk, we investigate asymptotic properties of the SVM for HDLSS data. We show that the SVM holds a consistency property in which misclassification rates tend to zero as the dimension goes to infinity under certain severe conditions. We show that the SVM is very biased in HDLSS settings and its performance is affected by the bias directly when sample sizes and covariance matrices are unbalanced. In the worst case, the SVM has a inconsistency property in the sense that one error rate tends to zero although the other tends to one as the dimension increases. In order to overcome such difficulties, we propose a bias-corrected SVM (BC-SVM). We show that the BC-SVM has consistency properties under milder conditions and improves the SVM even when sample sizes or covariance matrices are unbalanced. Finally, we check the performance of the BC-SVM by numerical simulations and use it in real data analysis.
3:50 PM to 4:10 PM	Huiming Zhang Peking University	#N/A
4:10 PM to 4:30 PM	Febry Utami STIS 56 Computer Division	"RAW DATA CAPTURE AS EASY AS LIKERT-TYPE BUT HAVING QUASI INTERVAL SCALE WITH SMART PHONE (PART ONE)" computer division : febry utami STIS 56 maya narang STIS 44 economics division : nela yesiana STIS 47 Likert - type is often used to capture raw data using numeral or integer. Captured raw data is not interval scale by respondent intention but often mistreated by statistician as interval scale. Mistreatment such as end to end symmetrical distance suggestive midpoint / category suggestive equal category width Certain culture tends to respond with positive evaluation so that end to end distance is not symmetrical. Present a line vertically or horizontally with labels only at both ends. Since there is not any suggestive midpoint, respondent is free to answer evaluation anywhere on the line. This line on hard copy paper is difficult to process as a statistical raw data. This line on respondent desk top monitor is difficult to express although respondent is sure of the Likert - type evaluation Touch screen such as that of smart phone facilitates both respondent graphical Likert - type evaluation and statistical raw data capture. Respondent may choose to be more accurate by activating touch screen option such as show pointer location. To avoid lazy choice by respondent the line has not any category marking. Graphical Likert - type line can be zoomed or transposed. A thicker line preferably from two adjacent colors with fine gradation can be used. Example from red to blue continuum pick A020F0 purple on one end and pick FF00FF magenta on other end. Interpolation can be used to create fine gradation, example F503FD, ECD8FC, E209FA, D90CF9, CF10F7, C813F8, BC18F4, B319F3, A91CF1. Color gradation is directly converted to quasi interval scale raw data. Computer programmer can use software function such as show touch or touchpose or fingertips. Keywords: Likert – type, fine gradation
4:30 PM to 4:50 PM	Ting Wang University of Otago	"BAYESIAN MODELLING OF MARKED POINT PROCESSES WITH INCOMPLETE RECORDS" Modelling point processes with incomplete records is a challenging problem, especially when the degree of record completeness varies over time. For example, there are more missing data from early time periods compared to recent periods for volcanic eruption records, and smaller events are more likely to be missing than substantial ones. In this study, we propose a time-varying intensity function for a marked point process to model the nonstationary variation of the observed point process caused by missing data. We apply this model to global and regional volcanic eruption records and use Bayesian inference to obtain hazard estimates and their uncertainties based on the observed incomplete records.
4:50 PM to 5:10 PM		#N/A
CP06	**Friday, 29 June 2018, 8:30 AM to 10:10 AM, ⑨ S16-06118**	
Chair	Yoon Bae Jun, Seoul National University	
8:30 AM to 8:50 AM	Yoon Bae Jun Seoul National University	"BAYESIAN SPATIAL REGRESSION WITH NONPARAMETRIC MODELING OF SPECTRAL DENSITIES" We propose a nonparametric modeling of spatial autocorrelation function under a Bayesian framework to predict a stationary spatial random field over a regularly spaced grid. Especially, we switch from the space domain to the frequency domain, and consider a Gaussian process prior for the log spectral density. Some interpolation technique to convert an estimated spectral density to a covariance matrix is suggested which does not require matrix inversion for the spatial prediction. Simulation study shows that our approach gives good estimation and prediction results when the true model is relatively smooth. Furthermore, our approach is robust in a sense that our approach does not require parametric form and/or isotropic assumption of a covariance function. In spite of estimating spectral density at Fourier frequencies during the Bayesian procedure, our approach does not lose much on computational efficiency compared to a conventional Bayesian spatial regression under a parametric covariance model. As a real data analysis, we illustrate Korean ozone exposure study from Moderate Resolution Imaging Spectroradiometer (MODIS) atmosphere dataset in 2017.

CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:50 AM to 9:10 AM	Jin Hong Park College of Charleston	"ROBUST ESTIMATION OF CONDITIONAL VARIANCE OF TIME SERIES" Suppose Z_{t} is the square of a time series Y_{t} whose conditional mean is zero. We do not specify a model for Y_{t}, but assume that there exists a $p \times 1$ parameter vector Φ such that the conditional distribution of $Z_t\|{\Z}_{t-1}$ is the same as that of $Z_t \| {\Phi}^{T} {\Z}_{t-1}$, where ${\Z}_{t-1} = (Z_{t-1}, \ldots, Z_{t-p})^{T}$ for some lag $p \ge 1$. Consequently, the conditional variance of Y_{t} is some function of ${\Phi}^T {\Z}_{t-1}$. To estimate ${\Phi}$, we propose a robust estimation methodology based on Density Power Divergences (DPD) indexed by a tuning parameter $\alpha \in [0,1]$, which yields a continuum of estimators, $\{{\hat {\Phi}}_{\alpha}; \alpha \in [0,1]\} $, where α controls the trade-off between robustness and efficiency of the DPD estimators. We develop data-dependent criteria for the selection of optimal α and lag p in practice. In fact, we illustrate the usefulness of our DPD methodology via simulation studies for the variance models, where the errors are drawn from a gross-error contamination model and the conditional variance is a linear and/or nonlinear function of ${\Phi}^T {\Z}_{t-1}$. Furthermore, we analyze the Chicago Board Options Exchange Dow Jones volatility index data and methodically show that our DPD approach yields viable models for the conditional variance, which are as good or superior to ARCH/GARCH models.
9:10 AM to 9:30 AM	Yoonsuh Jung Korea University	"EFFICIENT TUNING PARAMETER SELECTION BY CROSS-VALIDATED SCORE IN HIGH DIMENSIONAL MODELS" As DNA microarray data contain relatively small sample size compared to the number of genes, high dimensional models are often employed. In high dimensional models, the selection of tuning parameter (or, penalty parameter) is often one of the crucial parts of the modeling. Cross-validation is one of the most common methods for the tuning parameter selection, which selects a parameter value with the smallest cross-validated score. However, selecting a single value as a 'optimal' value for the parameter can be very unstable due to the sampling variation since the sample sizes of microarray data are often small. Our approach is to choose multiple candidates of tuning parameter first, then average the candidates with different weights depending on their performance. The additional step of estimating the weights and averaging the candidates rarely increase the computational cost, while it can considerably improve the traditional cross-validation. We show that the selected value from the suggested methods often lead to stable parameter selection as well as improved detection of significant genetic variables compared to the tradition cross-validation via real data and simulated data sets.
9:30 AM to 9:50 AM	Kenichi Satoh Hiroshima University	"TEXT MINING OF DECLARATION OF HIROSHIMA 1947-2016 AND ITS VISUALIZATION" While the atomic bomb survivors are aging, it is becoming more important social problems in Hiroshima to tell the experience to the next generation. On the other hand, some public facilities have stored a huge amount of text information including notes, testimonies and books. The new methodology for plainly summarizing these textual information by statistical methods or text-mining could provide great help for easily handing the information down to posterity. In the study we made an animation for visualizing keywords of Peace declaration presented by the Mayor of Hiroshima every year, from 1947 to 2016, in total 69 times. Keywords arrangement on the map was shown to well reflect changes of the time, that should be useful for us to understand the recovery history of Hiroshima.
9:50 AM to 10:10 AM	Wiji Triwilujeng Price Directorate, BPS	"REGIONAL PRIORITY FOR HOUSING ASSISTANCE OR HOUSING SUBSIDY BY IKK STRATIFICATION" Wiji Wilujeng price directorate, BPS wiji.triwilujeng@gmail.com Dwi Jayanti stis 49 economics dwijayanti.stis49@gmail.com Yurisman Adidharma price directorate, BPS In the past popular housing assistance or housing subsidy was based on poverty line. An alternative is proposed using IKK (Indeks Kemahalan Konstruksi) or physical construction cost regional index (as opposed to temporal index). High IKK value reflects that it is more expensive to build physical infrastructure such as hospital, bridge, airports, harbor, education facilities Firstly 42 regions in Papua and Western Papua are separated due to over the top IKK value. Secondly seven regions are separated (surabaya - based ikk > 129) Thirdly four and half hundred regions are stratified into five strata namely top, high, middle, low, bottom. Next bivariate Human Development Index (HDI) & Gross Domestic Product (GDP) are used as proxy respectively to regional skill capacity & regional monetary capacity. Bivariate Minimum Covariance Determinant [MCD] identifies leverage point and non-Extreme outlier. Regions within top IKK stratum identifed as leverage points having relatively high GDP are considered as not in priority for allowance or subsidy such as Bogor, Kutai Kartanegara, Bengkalis. Regions within high IKK stratum identifed as leverage points having relatively high GDP are considered as not in priority for allowance or subsidy such as Surabaya, Batam, Bekasi Top IKK stratum has a hundred regions with some non-Extreme outliers including Berau, Rokan Hilir High IKK stratum has a hundred regions with some non-Extreme outliers including Gayo Lues, Halmahera Selatan, Alor, Boalemo Regions identifed as non-Extreme outlier are prioritized for housing assistance or housing subsidy. Assistance or subsidy for instance can be referred to WikiHouse.cc or Ubuntu-Blox (www.youtube.com/watch?v=ugqhdms1rg4) or Build Change (www.newscientist.com/article/mg21028138-300). keywords: IKK, GDP, HDI, leverage
CP07	**Tuesday, 26 June 2018, 3:30 PM to 5:10 PM, ⑨ S16-06118**	
Chair	Mitsunori Ogawa, University of Tokyo	
3:30 PM to 3:50 PM	Mitsunori Ogawa University of Tokyo	"THE HOLONOMIC GRADIENT METHOD FOR THE MOMENTS OF TRUNCATED CENTERD MULTIVARIATE NORMAL DISTRIBUTION" We apply the holonomic gradient method to compute the moments of truncated centered multivariate normal distributions. Our method utilizes the A-hypergeometric system associated with the second Veronese configuration. By numerical experiments, it is confirmed that our method accurately computes the moments. We present some applications including the robust graphical modeling based on the alternative t-distribution, which is the original motivation of this study. This talk is based on the joint work with Kazuki Nakamoto and Tomonari Sei.
3:50 PM to 4:10 PM	Yihua Wang Tamkang University, Taiwan	"ON THE MONITORING OF AUTOCORRELATED GENERAL LINEAR PROFILES" If the quality of a process is better represented by a functional relationship between response variables and explanatory variables, a collection of this type of quality data is called a profile. Profile monitoring is used to understand and check the stability of this relationship or curve over time. The independent assumption for the error term is commonly used in the literature. However, in many applications, the profile data show correlations over time. Therefore, in this study, we focus on a general linear regression model with a first-order autocorrelation between profiles. We propose a MEWMA (multivariate exponentially weighted moving average) charting scheme to monitor this type of profile. The simulation study shows that our proposed methods outperform the existing schemes based on the average run length criterion.

CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
4:10 PM to 4:30 PM	Lili Yue Beijing University of Technology	"REGRESSION ADJUSTMENT FOR TREATMENT EFFECT WITH MULTICOLLINEARITY IN HIGH DIMENSIONS" Randomized experiment is an important tool for studying the Average Treatment Effect (ATE). In this paper, we consider the regression adjustment estimation of the Sample Average Treatment Effect (SATE) in high-dimensional case, where the multicollinearity problem is often encountered and needs to be properly handled. Many existing regression adjustment methods fail to achieve satisfactory performances. To solve this issue, we propose an Elastic-net adjusted estimator for SATE under the Rubin causal model of randomized experiments with multicollinearity in high dimensions. The asymptotic properties of the proposed SATE estimator are shown under some regularity conditions, and the asymptotic variance is proved to be no greater than that of the unadjusted estimator. Furthermore, we propose an Neyman-type conservative estimator for the asymptotic variance, which yields tighter confidence intervals than both the unadjusted and the Lasso-based adjusted estimators. Some simulation studies are carried out to show that the Elastic-net adjusted method is better in addressing collinearity problem than existing methods. The advantages of our proposed method are also shown in analyzing the dataset of HER2 breast cancer patients.
4:30 PM to 4:50 PM	Jiangyan Wang Nanjing Audit University	"ORACALLY EFFICIENT ESTIMATION OF THE FACTOR RETURN AND SPECIFIC ERROR DISTRIBUTIONS WITH SCB" The distributions of the factor return and specific error for an individual variable are important in forecasting and applications. However, they are not identified with low dimensional observations. Using the recently developed theory for large dimensional approximate factor model for large panel data, the factor return and specific error can be estimated consistently. Based on the estimated factor returns and residual errors, we construct the empirical processes for estimation of the distribution functions of the factor return and specific error, respectively. We prove that the two empirical processes are oracle efficient when T = o(p) where p and T are the dimension and sample size, respectively. This demonstrates that the factor and residual empirical processes behave as well as the empirical processes pretending that the factor returns and specific errors for an individual variable are directly observable. Based on this oracle property, we construct the simultaneous confidence bands for the distributions of the factor return and specific error. For the first order consistency of the estimated factor and residual distributions, T^(1/2) = o(p) suffices. Extensive simulation studies check that the estimated bands have good coverage probabilities. Our real data analysis shows that the factor return distribution has a structural change during the crisis in 2008, while the idiosyncratic return distribution does not change much.
4:50 PM to 5:10 PM	Gabriela Ciolek Telecom ParisTech	"ROBUST ESTIMATION FOR PDMPS" The purpose of this talk is to present a method to build robust estimators for Piecewise Deterministic Markov Processes (PDMPs) when the underlying embedded chain is regenerative .That is of particular interest when the process is contaminated by outliers. The aforementioned method relies on a renewal theory for Markov\nchains, i.e. we divide the sample path of the chain into data blocks and eliminate blocks having either too big contribution to the statistics of interest or having too big length resulting in an important bias on the statistics. We will illustrate this result on selected statististical learning procedure and a ruin model arising in insurance.
CP08	**Friday, 29 June 2018, 10:30 AM to 12:10 PM, ⑨ S16-06118**	
Chair	Joonpyo Kim, Seoul National University	
10:30 AM to 10:50 AM	Joonpyo Kim Seoul National University	"EXPECTILE FUNCTIONAL DATA CLUSTERING" This talk considers the problem of clustering functional data with various probabilistic structures that includes heavy-tailed or asymmetric distribution. In literature, numerous methods for functional data clustering have been studied. However, most of them have focused on mean information of the data using symmetric measures; hence, they are not efficient in clustering the data that have similar mean responses but different probabilistic structure. The talk presents a new clustering method that can adapt to the probabilistic characteristics of functional data. The method is based on expectile curves that is defined as a minimizer of asymmetric L2 criterion. Hence, the proposed method is capable of capturing inherent characteristics of functional data beyond mean structure. Results from a simulation study and a real example demonstrate the promising empirical properties of the proposed approaches.
10:50 AM to 11:10 AM	Wenlu Tang Chinese University of HongKong	#N/A
11:10 AM to 11:30 AM	Yuan Tsung Chang Mejiro University	"A NEW ASPECT OF ESTIMATION OF POSITIVE NORMAL MEANS, INDIVIDUAL AND SIMULTANEITY" We study estimation of non-negative normal mean, when variance is known and propose a generalized Bayes estimator for non-negative normal means. Based on the proposed generalized Bayes estimator we consider the Stein-type estimator for estimation several unknown non-negative normal means, simultaneously.
11:30 AM to 11:50 AM	Xingjie Shi Duke-NUS Medical School	"A FORWARD AND BACKWARD STAGEWISE ALGORITHM FOR NONCONVEX LOSS FUNCTIONS WITH ADAPTIVE LASSO" Penalization is a popular tool for multi- and high-dimensional data. Most of the existing computational algorithms have been developed for convex loss functions. Nonconvex loss functions can sometimes generate more robust results and have important applications. Motivated by the BLasso algorithm, this study develops the Forward and Backward Stagewise (Fabs) algorithm for nonconvex loss functions with the adaptive Lasso (aLasso) penalty. It is shown that each point along the Fabs paths is a δ-approximate solution to the aLasso problem and the Fabs paths converge to the stationary points of the aLasso problem when δ goes to zero, given that the loss function has second-order derivatives bounded from above. This study exemplifies the Fabs with an application to the penalized smooth partial rank (SPR) estimation, for which there is still a lack of effective algorithm. Extensive numerical studies are conducted to demonstrate the benefit of penalized SPR estimation using Fabs, especially under high-dimensional settings. Application to the smoothed 0-1 loss in binary classification is introduced to demonstrate its capability to work with other differentiable nonconvex loss function.
11:50 AM to 12:10 PM	Claudia Rivera University of Auckland	#N/A
CP09	**Friday, 29 June 2018, 3:30 PM to 5:10 PM, ⑨ S16-06118**	
Chair	Yi Shen Lin, Academia Sinica, Taiwan	

CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
3:30 PM to 3:50 PM	Yi Shen Lin Academia Sinica, Taiwan	"OPTIMAL STOPPING PROBLEMS WITH LOGCONCAVE REWARD FUNCTIONS" In the literature, the problem of maximizing the expected discounted reward over all stopping times has been explicitly solved for a number of reward functions (including $(x^+)^{ u}$, $(e^{x}-K)^+$, $1-e^{-x^{+}}$, $x\in\mathbb{R}$, $ u\in(0,\infty)$ and $K>0$) when the underlying process is either a random walk in discrete time or a Le ̀vy process in continuous time. All of such reward functions are increasing and logconcave while the corresponding optimal stopping problems have one-sided solutions (i.e. the optimal stopping times are of threshold type). In this talk, we show that all optimal stopping problems with increasing and logconcave reward functions admit one-sided solutions for general random walks in discrete time and Le ̀vy processes in continuous time. Furthermore, we explore the principle of smooth fit in the case of Le ̀vy processes. (This talk is based on joint work with Dr. Yi-Ching Yao.).
3:50 PM to 4:10 PM	Seoncheol Park Seoul National University	"MULTIRESOLUTION ANALYSIS FOR SPATIO-TEMPORAL DATA" This talk presents a new method for multiresolution analysis of spatially scatterd spatio-temporal data. The proposed method is based on bottom-up network design and lifting scheme. For multiresolution analysis, we suggest a data-adaptive lifting scheme with having a predetermined network, which sequentially decomposes numerous time series at fine levels into a few time series at coarse levels. Unlike conventional lifting schemes, a data-adaptive lifting scheme is a method based on edge-weighted graphs where weights are decided by partitioned around medoids (PAM) clustering for time series data, which gives robust clustering results for temporal outliers or extremes. Simulation studies and real data analysis applied to environmental time series are provided.
4:10 PM to 4:30 PM	Guebin Choi Seoul National University	"HEAVY-SNOW TRANSFORM: A NEW MULTISCALE METHOD FOR TIME SERIES" This talk proposes a new multiscale method for time series, termed 'heavy-snow transform', which is motivated by observing how much snowfall accumulates on the one-dimensional data. The concept of the proposed method is shared with that of scale-space theory in computer vision. Heavy-snow transform is defined by a collection of observations over various depths of snow. Heavy-snow transform uses some characteristics of the shape of snow accumulation for analyzing time series; hence, it might be useful for evaluating some probabilistic structures of time series whether a particular set of observations is similar to nearby others or not. Furthermore, some statistical applications such as detection and smoothing are studied.
4:30 PM to 4:50 PM	Weichang Yu University of Sydney	"VARIATIONAL NONPARAMETRIC DISCRIMINANT ANALYSIS" We propose a Bayesian nonparametric classification method that is suitable for high dimensional data. The method seamlessly integrates classification and variable selection – the two common objectives of data analysis in genomics studies. By specifying a Pólya tree prior on the unknown variable distributions and using variational Bayes for approximate posterior inference, we arrive at an intuitive variable selection and classification rule (with potentially good asymptotic properties). Through simulated and public-available gene expression datasets, we demonstrated the classifier’s competitive performance with existing nonparametric classifiers and its superiority over Gaussian-based classifiers.
4:50 PM to 5:10 PM	Saba Riaz Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology	#N/A
CP10	**Thursday, 28 June 2018, 8:30 AM to 10:10 AM, ⑨ S16-06118**	
Chair	Ismiana Putri, STIS 53 Computer Division	
8:30 AM to 8:50 AM	Ismiana Putri STIS 53 Computer Division	"REPRESENTATION OF REGIONAL CONSTRUCTION COST" dwi jayanti(1) ismiana putri(2) lita pertiwi(3) febry utami(4) (1) STIS 49 Economics, dwijayanti.stis49@gmail.com (2) STIS 53 Computer, ismianaptri@gmail.com (3) STIS 53 Social, litapertiwi.stis53@gmail.com (4) STIS 56 Computer, febry.utami.id@gmail.com Research Statement: There are occasions when IKK is used as raw materials but not released to general public. This occasion call for quick but rough estimate of IKK by covering only part of the regions as well as part of the goods and services. Approach: Firstly a particular aggregated data for all regions is subject to quality set for example by split into more reliable and less reliable. We use minimum covariance determinant (mcd) to split the regions into one hundred regions for further process and set aside other regions. There are more than four hundred goods and services in raw data input and we decide to have twenty eight most important ones. Similar goods or services can be represented by a single variate which reduces the number of goods or services to seventeen. Secondly we apply principal component analysis (pca) to obtain goods and services (variates). That is from seventeen variates we obtain seventeen principal components. If seventy per cent of explained variance is considered sufficient then only ten principal components are retained. Result: Only less than ten variates with more variability are used for quick but rough estimate of IKK. Conclusion: This proposition is based on particular aggregated data of certain year then it is not valid for other purpose far from reference time Keywords : variability, quick rough estimate, IKK, MCD, PCA, dimensionality reduction

CONTRIBUTED PAPER SESSIONS

Session ID Time	Full Name Affiliation	Title Abstract
8:50 AM to 9:10 AM	Evi Sanjaya STIS 55 Social Division	"ALTERNATIVE METHOD TO IDENTIFY NON-EXTREME OUTLIER (PART ONE)" STIS 55 Social Division : Deja Firda, Isni Sukrianah, Evi Sanjaya STIS 53 Computer Division : Risna Yuliani Most people usually consider outlier as extreme, this point of view can be true for univariate observation. For bivariate observation, another term is introduced that is leverage (Gujarati, 2008). The research goal is to apply different treatment for the identified non-Extreme outlier(s). Mechanism to achieve this expectation consists of four steps. Firstly regions in the universe (not sample) are stratified into top, high, middle, low, bottom GDP [by Dewi Takarini & Laura Fadilah & friends 3rd Asian Population Association July 2015 conference]. Secondly almost all regions within each stratum are contained by ellipse [by Risna Yuliani & friends 14th Pacific Regional Science Conference Organization June 2016]. Thirdly regions outside ellipse are considered either as leverage point or as non-Extreme outlier Fourthly shrink or expand the ellipse to consider different number of regions as leverage point or as non-Extreme outlier. Non-Extreme outliers are expected to reveal strength or weakness. Mahalanobis' distance is used to identify the regions which considered as non-extreme outliers. The research goal is to apply different treatment for the identified non-Extreme outlier(s). Acknowledgement The author and co-authors declare there is not any potential conflict of interest with respect to the research, authorship, and/or publication of this article. The views expressed here are those of the individual author and co-authors and not necessarily those of BPS or STIS or its board, or officers, or staff. References Gujarati, D. N. & Porter, D. C. (2008). Basic Econometrics: Fifth Edition. New York, United States: McGraw-Hill/Irwin. World Bank. per capita GDP (Current US$ 2013). Retrieved from http://data.worldbank.org/indicator/NY.GDP.PCAP.CD Keywords: Mahalanobis distance, non-Extreme Outlier
9:10 AM to 9:30 AM	Ryota Takami Tokyo University of Science	"ON POINT-ASYMMETRY OF MULTINOMIAL PARAMETER AND DECOMPOSITION OF POINT-SYMMETRY" Some trials have more than two possible outcomes. We suppose that each of n independent, identical trials can have outcome in any of c ordinal categories. Then the counts, which is the sum of the multinomial trials, have the multinomial distribution. This paper treats a problem of modeling for parameter of a multinomial distribution. Especially, we focus on the structure of symmetry with respect to the center point (or cell) of categories. Then, we show that the symmetry for the parameter of multinomial distribution can be separated into the asymmetric structure of it and equality of some moments. This result can also be applied for the multi-way contingency tables. By using the proposed model, we consider the decomposition of the point-symmetry model proposed by Wall and Lienert (1976). The decomposition may be useful to see the reason for the poor fit of the point-symmetry model when the point-symmetry model fits poorly for the real dataset.
9:30 AM to 9:50 AM	Keisuke Yano University of Tokyo	"NON-ASYMPTOTIC MINIMAX ADAPTATION AND WEAK ADMISSIBILITY USING RANDOM SIEVE PRIORS" This talk focuses on theoretical aspects of Bayesian nonparametrics. We study random sieve priors in nonparametric regression from the viewpoints of non-asymptotic minimax adaptation and weak admissibility.Non-asymptotic minimax adaptation is an important property that the estimator is expected to have. An estimator is said to be non-asymptotic minimax adaptive if it is minimax up to a universal constant and if it does not depend on any hyper-parameter of the function class of the regression function.We show that the Bayes estimator based on the random sieve prior is non-asymptotic minimax adaptive and it also has weak admissibility.This talk is based on the joint work with Fumiyasu Komaki (the university of Tokyo).
9:50 AM to 10:10 AM	Li Xie Beijing University of Technology	"PARTIALLY LINEAR ADDITIVE SPATIAL AUTOREGRESSIVE MODEL" This paper considers the partially linear additive spatial autoregressive models. The nonparametric functions are approximated by polynomial spline. Based on the profile quasi-likelihood estimation method, we propose the estimation for the parametric and nonparametric components. Under mild conditions, the asymptotic normal of the corresponding estimators of the linear components are established , as well as the convergence rate of the estimators of the nonparametric components. The finite sample performance of the proposed estimation method is assessed by Monte Carlo simulation studies. Finally the proposed model and estimation method are applied to a real data analysis. This is a joint work with Cao Ruiyuan and Wang Xiaorui.

Last updated
7/2/2018