Joint Meeting of

The 2011 Taipei International Statistical Symposium and

7th Conference of the Asian Regional Section of the IASC

7th IASC-ARS

Σ joint 2011

Taipei Symposium

December 16 - 19, 2011

Academia Sinica, Taipei, Taiwan

http://joint2011.stat.sinica.edu.tw/

中央研究院 ACADEMIA SINICA 1928

Abstract

December 17 (Saturday)

Today's Highlights:

08:30 – 09:30	***Keynote Speech (II) by Wing Hung Wong***
09:30 – 10:10	***The Statistica Sinica Special Invited Session (I) by Jun Liu***
10:30 – 12:00	***Parallel Sessions 17a1 – 17a7***
12:00 – 13:00	***Poster Session 1 – 11***
13:00 – 14:30	***Parallel Sessions 17b1 – 17b7***
14:40 – 16:10	***Parallel Sessions 17c1 – 17c7***
16:30 – 17:30	***Parallel Sessions 17d1 – 17d7***
18:30 –	***Welcome Party & Dinner Speech by George C. Tiao***

Abstract Pages:

Special Sessions:	Keynote Speech (II) Page 90	SS Invited Session (I) Page 90	Poster Session Page 103

Session No.	17a1	17a2	17a3	17a4	17a5	17a6	17a7
Starting Page	91	93	94	96	98	99	102

Session No.	17b1	17b2	17b3	17b4	17b5	17b6	17b7
Starting Page	110	111	113	114	116	118	119

Session No.	17c1	17c2	17c3	17c4	17c5	17c6	17c7
Starting Page	121	123	124	126	128	129	131

Session No.	17d1	17d2	17d3	17d4	17d5	17d6	17d7
Starting Page	133	135	136	138	140	142	145

Keynote Speech (II)
December 17 (Saturday), 8:30 - 9:30, HSS Center International Conference Hall
Speaker: Wing Hung Wong
Chair: Chuhsing Kate Hsiao

Statistical Issues in Personalized Genomics & Medicine

Wing Hung Wong
Department of Statistics, Stanford University, U.S.A.

Dramatic progress in sequencing technology has recently made it feasible to consider ”mega-cohorts” where both longitudinal medical records and personal genome data are obtained for a substantial fraction of a well defined population. If utilized effectively this data will enable new and powerful approaches to medical discovery and health service delivery. In this talk I will discuss several computational and statistical issues that are relevant to the planning and analysis of such studies.

[Wing Hung Wong, Department of Statistics, Stanford University, U.S.A.; whwong@stanford.edu]

The Statistica Sinica Special Invited Session (I)
December 17 (Saturday), 9:30 - 10:10, HSS Center International Conference Hall
Speaker: Jun Liu
Chair: Naisyin Wang

Dictionary Models for Item Association with Applications to Chinese Text Analysis

Jun Liu
Department of Statistics, Harvard University, U.S.A.

Pattern discovery is a ubiquitous problem in many disciplines. It is especially prominent in recent years due to our greatly improved data-generation capa- bilities in science and technologies. The model and methods I present here is motivated by the ”motif-finding” and ”module-finding” problems in biology, the ”market-basket problem” in data mining, and text analysis in studying chinese history books. In these problems, a common challenge is to discover which ”items” (or, key words in text mining, and regulatory elements in biology) tend to co-occur with which others, i.e., to find association rules among the items. In market-basket problems, the observations are customers’ transactions (i.e., ”bastkets”), each contains multiple items. We can imagine that each basket is composed by a few ”themes” selected by the custermer and each theme is a set of items that are bought together (an analogy is stamp-collecting: a person’s collection of stamps can be organized as ”sets”). Our goal is to discover these themes from only the transactions. Inspired by a dictionary

model proposed by Bussemaker, Li and Siggia (2000), we propose a ”theme dictionary model”, which prescribes a probabilistic rule for generating each transaction. We then used both the EM and Monte Carlo strategies to aid our inference of the themes. In text analysis and biological sequence analysis, an added difficulty is that the ”items” are some phrases and sequence patterns, which are not all known in advance. In this case, we can combine a motif finding strategy with the theme dictionary model to complete the analysis. Existing motif-finding methods are mostly ”bottom-up” approaches, i.e., to build up the dictionary starting with single-letter words and then concatenate some existing words that appear to occur next to each other in sentences more frequently than chance. Our new approach is a top-down strategy, which uses a tree structure to represent the relationship among all possible existing words and uses the EM algorithm to estimate the U.S.A.ge frequency of each word. It automatically trims down most of the incorrect ”words” by letting their U.S.A.ge frequencies converge to zero. I will demonstrate its applications in a few examples including an analysis of a Chinese novel, some Chinese history books, and publications in PNAS in the past 50 years. This is based on a joint work with Ke Deng, Zhi Geng, Chunlin Ji, and Peter Bol.

[Jun Liu, Department of Statistics, Harvard University, U.S.A.; jliu@stat.harvard.edu]

17a1-Statistical Genetics
December 17 (Saturday), 10:30 - 12:00, HSS 1st Conference Room
Organizer: Chao A. Hsiung
Chair: Chao A. Hsiung

17a1-1 On Selection of Endophenotypes in Genetic Mapping

Yen-Feng Chiu
Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan, ROC

The complex etiology of common diseases often challenges investigators in finding the relationships between phenotypes and genotypes. One option to dissolve the complexity is to select an ”endophenotype” to signify quantitative measures that are more proximate to the biological etiology of a clinical disorder than its signs and symptoms and thus less genetically complex than the disorder’s underlying mechanism. In practice, it is common for investigators to measure numerous phenotypes on individuals. To achieve a substantial possibility of success in gene mapping, a systematic statistical method on election or validation of endophenotypes would be very helpful. In this study, we evaluated a statistical approach on endophenotype selections in association mapping. This approach allows investigators to select or validate an endophenotype in a pathway and can be applied to all kinds of study designs in association mapping as well as in linkage mapping.

[Yen-Feng Chiu, Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taiwan, ROC; yfchiu@nhri.org.tw]

17a1-2 Mapping Complex Traits in Genetically Admixed Populations

Hua Tang
Department of Genetics, Stanford University, U.S.A.

Mapping Complex Traits in Genetically Admixed Populations To date, the majority of genome-wide association studies (GWAS) have focus on populations of European descent. GWAS in ancestrally admixed populations offer exciting opportunities for identifying genetic variants that underlie phenotypic diversity between populations, as well as between individuals. At the same time, the het- erogeneous genetic background and population structure pose special challenge for traits mapping analyses. I will describe a study of skin and eye color in an African-European admixed population, in which genotype-based and ancestry- based association approaches are applied and compared. Analysis of these data also provided insights into the genetic architecture of complex traits.

[Hua Tang, Department of Genetics, Stanford University, U.S.A.; huatang@stanford.edu]

17a1-3 Distribution of the Number of False Discoveries in Large-Scale Family Based Association Testing with Application to the Association between PTPN1 and Hypertension and Obesity

Wen-Chang Wang
Chao A. Hsiung
Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Taiwan

Lan-Chao Wang
National Institute of Cancer Research, National Health Research In- stitutes, Taiwan

Lee-Ming Chuang
Department of Internal Medicine, National Taiwan University, Taipei, Taiwan

Thomas Quertermous
Division of Cardiovascular Medicine, Falk Cardiovascular Research Center, Stanford University, CA, U.S.A.

I-Shou Chang
Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences and National Institute of Cancer Research, National Health Research Institutes, Taiwan

We present a model-free approach to the number of false discoveries for a largescale simultaneous family based association tests (FBATs), in which the set of discoveries are decided by applying a threshold to the test statistics. When the association between a set of markers in a candidate gene and a group of phenotypes is studied by a class of FBATs, we indicate that a joint null hypothesis distribution for these statistics can be obtained by the fundamental statistical method of conditioning on sufficient statistics for the null hypothesis. Based on the joint null distribution of these statistics, we can obtain the distribution of the number of false discoveries for the set of discoveries defined by a threshold, referred to as tail counts. Simulation studies are presented to indicate that conditional distribution of the tail counts, not the unconditional one, is appropriate for the study of false discoveries. The usefulness of this approach is illustrated by re-examining the association between PTPN1 and a group of blood pressure related phenotypes reported by Olivier et al. (2004); our results refine and reinforce this association.

[I-Shou Chang, National Institute of Cancer Research, National Health Research Institutes, 35, Keyan Road, Zhunan Town, Miaoli County 350, Taiwan; ischang@nhri.org.tw]

17a2-Statistical Methods for Large Scale Data: from Gene to Sun
December 17 (Saturday), 10:30 - 12:00, HSS 2nd Conference Room
Organizer: I-Ping Tu
Chair: Tzee-Ming Huang

17a2-1 **Statistical Computing in Protein Folding**

Samuel Kou
Department of Statistics, Harvard University

Predicting the native structure of a protein from its amino acid sequence is a long standing problem. A significant bottleneck of computational prediction is the lack of efficient sampling algorithms to explore of the configuration space of a protein. In this talk we will introduce a sequential Monte Carlo method to address this challenge: fragment regrowth via energy-guided sequential sampling (FRESS). The FRESS algorithm combines statistical learning (namely, learning from the protein data bank) with sequential sampling to guide the computation, resulting in a fast and effective exploration of the configurations. We will illustrate the FRESS algorithm with both lattice protein model and real proteins.

[Samuel Kou, Department of Statistics, Harvard University; kou@stat.harvard.edu]

17a2-2 **Dynamic Tomographic Imaging of the Solar Corona**

Yuguo Chen
University of Illinois at Urbana-Champaign, U.S.A.

We address the image formation of a dynamic object from projections by formulating it as a state estimation problem. To overcome the curse of dimensionality in large scale state space models, we discuss the localized ensemble Kalman filter, a Monte Carlo state estimation procedure that, unlike the standard particle filter and many other state estimation techniques, remains computationally tractable when the state dimension is large. We give the asymptotic behavior of the localized ensemble Kalman filter, and apply it to time-dependent tomographic imaging of dynamic objects such as the solar corona.

[Yuguo Chen, Department of Statistics, University of Illinois at Urbana-Champaign, 725 S. Wright Street, Champaign, IL 61820 U.S.A.; yuguo@illinois.edu]

17a2-3 **Normalization and Testing for RT-PCR Experiments**

Marc A. Coram
Stanford University, U.S.A.

Quantitative real-time polymerase chain reaction technology (qRT-PCR) allows the scientist to measure RNA expression over a wide dynamic range with impressive precision and sensitivity. It can be applied to measure the expression quantify the trace expression from tuberculosis bacteria living in host tissue or of a single human cell. It can be multiplexed to measure thoU.S.A.nds of genes per sample. For data of this kind, one might think that a straightforward application of microarray analysis methodology would be suitable, and in many ways it is, but careful consideration reveals somewhat different statistical problems in play. As examples: sample handling, cDNA synthesis, and PCR pre-amplification introduce variability; censoring occurs when the expression of a gene passes below the genes detection limit; normalization of the samples is essential for their comparison but hard to pin down. This will not be a talk about the technology of qRT-PCR, and neither do I have all the answers, but it would be my hope to present a cogent introduction to the domain, an illustration of some of the issues of concern, and to indicate possible approaches to their resolution.

[Marc A. Coram, 259 Campus Dr., Stanford, CA 94305, U.S.A.; mcoram@stanford.edu]

17a3-Medical Statistics
December 17 (Saturday), 10:30 - 12:00, HSS Media Conference Room
Organizer: Yi-Ting Hwang
Chair: Hung Hung

17a3-1 **Modeling Left-truncated and Right Censored Survival Data with Longitudinal Covariates**

Jane-Ling Wang
University of California, Davis, U.S.A.

In this talk, we explore the modeling of survival data in the presence of longitudinal covariates. In particular, we consider survival data that are subject to both left truncation and right censoring. It is well known that traditional approaches, such as the partial likelihood approach for the Cox proportional hazards model encounter difficulties when longitudinal covariates are involved in the modeling of the survival data. A joint likelihood approach has been shown in the literature to provide an effective way to overcome those difficulties for right censored data. However, in the presence of left truncation, there are additional challenges for the joint likelihood approach. We propose an alternative likelihood to overcome these difficulties and establish the asymptotic theory, including semiparametric efficiency of the new approach. The approach will also be illustrated numerically. The talk is based on joint work with Yuru Su, National Cheng-Kun University.

[Jane-Ling Wang, Department of Statistics, University of California, Davis, CA95616, U.S.A.; wang@wald.ucdavis.edu]

17a3-2 Joint Modeling of Multivariate Survival and Longitudinal Data

Ya-Fang Yang
Yi-Kuan Tseng
National Central University, Taiwan

In the literature, joint modeling approaches have been employed to analyze both survival and longitudinal processes and to investigate their association. Early attention has mostly been placed on developing adaptive and flexible longitudinal processes based on a prespecified univariate survival model, most commonly chosen as the Cox proportional model. We propose a marginal likelihood approach to handle multivariate survival time in joint model framework which implements the similar idea of marginal methods used in literature by ignoring the dependency among event times. The marginal likelihood could be easily incorporated various survival model in the likelihood function including two popular survival models, Cox and AFT models, or others such as extended hazard model. The maximization of the marginal likelihood is conducted through Monte Carlo EM and the standard error estimates are obtained via bootstrap method. The performance of the procedure is demonstrates through simulation study.

[Yi-Kuan Tseng, No. 300, Jhongda Rd., Jhongli, Taoyuan, Taiwan; tsengyk@ncu.edu.tw]

17a3-3 Statistical Methods for Evaluating ED Interventions

W.Y. Wendy Lou
University of Toronto, Toronto, Ontario, Canada

Strategies for reductions in wait times within various health care settings, such as the emergency department (ED), are often implemented to improve health outcomes and quality of care. Methods

for evaluating the effectiveness of interventions involving protocol or policy changes within the ED are often based on unrealistic assumptions. Methodology that allows for systematic assessment, under more realistic settings, of the changes attributed to interventions will be presented. The approach utilizes information derived from retrospective and prospective system monitoring, and involves structural multi-dimensional time series and statistical process control methods. The proposed methodology will be illustrated through a motivating example from a study involving multiple hospitals of a health system in Ontario, Canada. Some challenges as well as opportunities for such statistical methodologies will be discussed.

[W.Y. Wendy Lou, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Toronto, ON M5T 3M7, Canada; wendy.lou@utoronto.ca]

17a4-Analysis of Social Networks: Detecting Patterns and Rules of Interaction in Humans
December 17 (Saturday), 10:30 - 12:00, AC 1st Conference Room
Organizer: Wei-Chung Liu
Chair: Hwai-Chung Ho

17a4-1 Global and Local Assessment of Distributional Inequality in Networks

Yen-Sheng Chiang
Department of Sociology & Institute for Mathematical Behavioral Science University of California, Irvine, U.S.A

Measuring distributional inequality has long been a core research topic in the social sciences. Inequality is usually assessed globally across the population, yet with increasing studies in social psychology indicating that people evaluate distributional inequality through comparison with their own reference groups, how inequality is assessed locally and how it differs from global assessment of inequality merits our attention. The underlying structure of ”who compares with whom?” for local inequality assessment can be treated as a network. Understanding how social network is formed thus helps us study the difference between global and local assessment of inequality. In this study, I consider a set of standard inequality measures and manipulate different principles of network formation through computer simulation to investigate how differently these inequality measures perform locally as opposed to globally. I also report a preliminary experiment study with human subjects to show whether inequality can be perceived differently between global and local assessment.

[Yen-Sheng Chiang, University of California, Irvine; yenshenc@uci.edu]

17a4-2 Comparing Local Structure in Social Networks

Katherine Faust
Department of Sociology and Institute for Mathematical Behavioral Sciences

Our understanding of network structure and process is advanced when we can compare, describe, and model networks of different relational contents or arising in different contexts. In this talk I discuss local structural comparison of social networks from a variety of animal species and social relations to identify classes of networks with similar structural signatures. Such classes might indicate traces of different network generating processes and thus be useful for characterizing fundamental types of networks or social relations. Results show that discrete clusters of network structural signatures are not readily apparent in this collection of social relations and raise questions about generalizability across settings, measurement of social relations, selection network configurations for comparison, and appropriate baselines for evaluation of structural tendencies.

[Katherine Faust, Department of Sociology, University of California, Irvine; Irvine, CA U.S.A. 92697-5100; kfaust@uci.edu]

17a4-3 Parametrizing Order-Stable Exponential Family Random Graph Models

Cater Butts
Department of Sociology and the Institute for Mathematical Behavioral Sciences, University of California, Irvine, U.S.A.

Parametrizing Order-Stable Exponential Family Random Graph Models General random graphs (i.e., stochastic models for networks incorporating heterogeneity and/or dependence among edges) are increasingly widely used in the study of social and other networks. Discrete exponential family representations for these models (the "exponential family random graph models," or ERGMs) are attractive due to the their flexibility, generality, and connections with existing statistical theory. Unfortunately, however, many seemingly natural and well-motivated ERGM parametrizations are ill-behaved, with one problem being the tendency of such models to concentrate probability mass on mixtures of nearly complete and nearly empty graphs as graph order (i.e., the number of vertices) increases. In this talk, I employ a recently introduced technique for bounding the behavior of ERGMs to show how one may parametrize ERGM families in ways that are order-stable, in the sense of avoiding large mean degree fluctuations as graph order increases. While not guaranteed to be well-behaved in all respects, these models can be shown to be immune to many pathologies of standard model families.

[Cater Butts, Department of Sociology and the Institute for Mathematical Behavioral Sciences, University of California, Irvine, U.S.A.; buttsc@uci.edu]

17a5-CSA/JSS/KSS International Session: Time series Analysis and Stochastic Processes
December 17 (Saturday), 10:30 - 12:00, AC 2nd Conference Room
Organizer: Ming-Yen Cheng, Sangyeol Lee, and Jinfang Wang
Chair: Byeongchan Seong

17a5-1 Functional Limit Theorems of Ruin Behavior for Levy Insurance Risk Process

Hyun Suk Park
Hallym University, Korea.

The interest of this paper is on when and how ruin occurs for large premium levels. For this, we consider a general Levy process for convolution equivalent Levy measures using the quintuple law, and the time to ruin which have a proper limiting distribution conditional on ruin occurring when non-Cramer condition holds. This yields a weak limit theorems for asymptotic behavior of ruin probability as the initial premium u tends to infinity. This asymptotic results and proofs are shown to simplify by comparison with the results of Griffin and Maller [Ann. Appl. Probab. 2011], which have investigated asymptotic path-decomposition conditionally on the rare event that the process crosses a large level. Simulation of finite sample versions is given for specific illustrations of asymptotic behavior of the time to ruin.

[Hyun Suk Park, Department of Finance & Information Statistics, Hallym University, Chuncheon 120-702, Korea; hspark@hallym.ac.kr]

17a5-2 Principal Volatility Components and Their Applications

Yu-Pin Hu
National Chi Nan University, Taiwan

Ruey S. Tsay
University of Chicago, U.S.A.

This research considers a measure for quantifying multivariate volatility and proposes a method that decomposes a vector time series into principal volatility components. We define a generalized covariance matrix for a vector time series and study properties of the new covariance matrix. We then perform an eigenvalue-eigenvector analysis on the squares of the generalized covariance matrix that enables us to achieve the decomposition. Test statistics are proposed to detect the number of components that have conditional heteroskedasticity. Asymptotic distributions of the test statistics are derived. Simulation study shows that the proposed test statistics work well in finite samples. For applications on financial data, we apply the proposed analysis to dimension reduction in multivariate volatility modeling and to search for common factors in volatility.

[Yu-Pin Hu, Department of International Business studies, National Chi Nan University, No. 1 University Road, Puli, Nantou, Taiwan; huyp@ncnu.edu.tw]

17a5-3 Adaptive Estimation for Diffusion Processes

Yoichi Nishiyama
The Institute of Statistical Mathematics, Japan

We consider a semiparametric estimation problem in an ergodic diffusion process model, where the drift coefficient contains an unknown finite-dimensional parameter of interest and the diffusion coefficient does a nuisance parameter with infinite-dimension. We propose two kinds of approaches, namely, semiparametric Z- and Bayes estimations. Both approaches bring us some adaptive estimators, in the sense that their asymptotic distributions of the estimators for the drift coefficient coincide with the optimal one in the Hajek-Le Cam theory when the diffusion coefficient is known.

[Yoichi Nishiyama, The Institute of Statistical Mathematics, Japan; nisiyama@ism.ac.jp]

17a6-Statistical Modeling and Analysis of Suicide Data
December 17 (Saturday), 10:30 - 12:00, AC 3rd Conference Room
Organizer: Jeng-Tung Chiang
Chair: Anne Chao

17a6-1 An Illness and Death Model for Evaluating Cost Effectiveness of Suicide Prevention Programs

Paul Siu Fai Yip
The University of Hong Kong, Hong Kong, China

An illness and death model is proposed to assess the effectiveness of suicide prevention model. Here we consider the impact on reducing suicide risk for the general population and the mental ill respectively. It is shown that the effect would be more significant in reducing the risk for the general population than that of the mentally ill. An empirical data from Hong Kong is used for illustration and some discussion will be provided. It echoes the famous Rose Theorem ”reducing a small risk to a large population would be more effective than reducing a large risk to a small population.”

[Paul Siu Fai Yip, The University of Hong Kong, Hong Kong, China; sfpyip@hku.hk]

17a6-2 Spatial Autocorrelation Statistics and Spatial Clustering in the Areas in Japan with Low Suicide Rates

Takafumi Kubota
The Institute of Statistical Mathematics, Tokyo, Japan

Makoto Tomita
Tokyo Medical and Dental University, Tokyo, Japan

Fumio Ishioka
Okayama University, Okayama, Japan

Toshiharu Fujita
The Institute of Statistical Mathematics, Tokyo, Japan

We focus on the areas in Japan with low suicide rates. We use ”Statistics of Community for the Death from Suicide” [*T.Fujita (2009): available in http:// ikiru.ncnp.go.jp/ikiru-hp/genjo/toukei/index.html*] to define non-suicide persons as total population minus suicide persons, and non-suicide rate as the number of non-suicide persons per 100,000 persons in each second medical care zone. Our first objective is to measure the degree of dependency among non-suicide rate. We use the spatial autocorrelation statistics especially Moran’s I. Our second objective is to detect cumulated area of non-suicide rate. We use spatial scan statistics to determine hotspot. We use ”coolspot” of suicide rate as the hotspot of non-suicide rate. Our last objective is to compare areas between the coolspot and the hotspot in [M.Tomita et al., Bulletin of the Computational Statistics of Japan, (2010):23(1), 25 - 43] to discuss the tendency as well as countermeasures to prevent suicide cases.

[Takafumi Kubota, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan; tkubota@ism.ac.jp]

17a6-3 Age-Period-Cohort Analysis of Suicide Mortality in Taiwan–Gender-Specific Differences

Jeng-Tung Chiang
National Chengchi University

Age-Period-Cohort model has been widely used to analyze temporal patterns associated with age, period and cohort effects. When data are stratified, researchers may also want to explore whether the above mentioned temporal patterns vary with strata. Unfortunately, almost all the analyses results appeared in the literature thus far relied only on researchers subjective judgments on whether the curves look different. Multivariate Age-Period-Cohort model was introduced only recently, it, however, also suffers the well-known non-identification problem. Thanks to the intrinsic estimator proposed by Fu (2000), the problem can be avoided. In this study, we incorporate the idea of intrinsic estimator into multivariate age-period-cohort model, and provide some useful statistical tests that can be useful for use in detecting strata differences. Taiwan suicide data will be used for an illustration.

[Jeng-Tung Chiang, National Chengchi University; chiangj@nccu.edu.tw]

17a6-4 Analysis of the Repeated Suicide Attempt: A Follow-up Study of the National Suicide Surveillance System in Taiwan

Wei-Hwa Chang
Taiwan Suicide Prevention Center, Taipei, Taiwan, R.O.C.

Chen-Hsin Chen
Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.

Chia-Ming Chang
Department of Psychiatry and Suicide Prevention Center, Chang Gung Memorial Hospital at Lin-Ko and Chang Gung University, Taoyuan, Taiwan, R.O.C.

Ming-Been Lee
Departments of Psychiatry and Social Medicine, National Taiwan University College of Medicine, Taipei, Taiwan, R.O.C.

Suicide attempt is a strong predictor of further repeated attempts and suicide complete. The National Suicide Surveillance System (NSSS) in Taiwan was established in 2006 for collecting and analyzing the information of suicide attempters, and providing aftercares to prevent their repeated attempts. A follow-up study included 51182 subjects registered in the NSSS for an index suicide attempt during 2006-2008. Information about whether these attempters had consented at the index attempts to receive aftercares and actually received aftercares were recorded in the NSSS with their characteristics such as gender, age, history of past mental disorders and suicide attempt method.
The Kaplan-Meier survival curve of repetition flattens to a level plateau in the long run as the probability of non-susceptibility. It implies that index attempters, who have not experienced reattempts, would either have a reattempt later or be non-susceptible to the event afterward. We propose the time-dependent transformation cure model as an extension of the transformation cure model, and use it to handle the time-dependent covariate of receiving aftercare in investigating both the probability of susceptibility to repeated suicide attempt and the time to the first reattempt of a susceptible case. The results show that early aftercares to the suicide attempters can significantly prevent them from repetition of attempts, especially for those subjects who had consented to receive aftercares right after their index attempts.

[Wei-Hwa Chang, Taiwan Suicide Prevention Center, Taipei, Taiwan, R.O.C.;]

17a7-Computer Experiments
December 17 (Saturday), 10:30 - 12:00, AC 4th Conference Room
Organizer: Ray-Bing Chen
Chair: Chien-Yu Peng

17a7-1 Efficient Particle Swarm Methods for Optimal Space-Filling Designs on GPU

Weichung Wang
Department of Mathematics, National Taiwan University, Taiwan, R.O.C.

Space-filling design is indispensable to computer experiments. Many optimal designs of computer experiments involve solving large-scale discrete optimization problems. We concentrate on two types of optimal space-filling designs in this talk: (i) the Latin hypercube designs based on the phip criterion and (ii) the uniform designs based on central composite discrepancy for irregular regions. Main challenges for generating such designs are due to the huge number of feasible points and irregular sensitivity of the objective function values. We first formulate these two space-filling design problems as discrete optimization problems with respect to different criteria. We propose variants of particle swarm optimization algorithms to solve the target discrete optimization problems. In addition, we use massively parallel graphic process units (GPU) to accelerate the computations. Numerical results will be presented to illustrate the characteristics of the proposed schemes with comparisons with other existed alternatives. This is a joint with Ray-Bing Chen (NCKU), Dai-Ni Hsieh (NTU), Yen-Wen Shu (NTU), and Ying Hung (Rutgers University).

[Weichung Wang, Department of Mathematics, National Taiwan University, Taiwan; wwang@math.ntu.edu.tw]

17a7-2 Iterative Kernel Fitting with Massive Data: Simultaneously Achieving Numeric and Nominal Accuracy

Peter Z. G. Qian
Department of Statistics, University of Wisconsin-Madison

In today's data-deluge world, fitting kernel models with massive data is a frequently encountered problem in statistics, applied mathematics and computer science. On one hand, the nominal accuracy of a kernel model increases with the number of data points. On the other hand, kernel fitting with a massive amount of data leads to numerical singularity. To reconcile this contradiction, we present a sequential method to simultaneously achieve numerical stability and theoretical accuracy in kernel fitting. This method forms nested space-filling subsets of the data, builds kernel models for different subsets and combines these submodels together to obtain an accurate final model. We introduce a decomposition of the overall error of a kernel model into nominal and numeric portions. Bounds on

the numeric and nominal error are developed to show that substantial gains in overall accuracy can be attained with this sequential method. Examples from computer experiments are given to bear out the effectiveness of the developed method.

[Peter Z. G. Qian, Department of Statistics University of Wisconsin-Madison; peterq@stat.wisc.edu]

17a7-3 Analysis of Computer Experiments with Functional Response

Ying Hung
Department of Statistics and Biostatistics, Rutgers University, U.S.A.

Most existing methods for analyzing computer experiments with single outputs such as kriging cannot be easily applied to functional outputs due to the computational problems caused by high-dimensionality of the response. In this paper, we develop an efficient implementation of kriging for analyzing functional responses. The main contribution of this paper is to develop a two-stage model building procedure and a general framework, which can be used irrespective of the data structure. When the functional data are observed on a regular grid, we show that the computations can be simplified using an application of Kronecker products. The case when the functional data are observed on a nonregular grid is quite complex, but we show that it can be handled using a Gibbs sampling- based estimation algorithm. The methodology is illustrated using a computer experiment conducted for optimizing residual stresses in machined parts.

[Ying Hung, Department of Statistics and Biostatistics, Rutgers University, U.S.A.; yhung@stat.rutgers.edu]

Poster Session
December 17 (Saturday), 12:00 - 13:10, HSS 4th Floor

1 Optimum Settings of Process Mean, Economic Order Quantity, and Commission Fee

Chung-Ho Chen
Department of Management and Information Technology, Southern Taiwan University

In this study, the author proposes a modified Chen and Liu's model with consignment policy. Taguchi's quadratic quality loss function will be adopted for evaluating the product quality. The optimum producer's process mean and retailer's economic order quantity and commission fee will be jointly determined by maximizing the expected total profit of society.

[Chung-Ho Chen, Department of Management and Information Technology, Southern Taiwan University, 1 Nan-Tai Street, Yung-Kang, Tainan 71005, Tai- wan; chench@mail.stut.edu.tw]

2 Optimizing Latin Hypercube Designs by Particle Swarm with GPU Acceleration

Dai-Ni Hsieh
Institute of Statistical Science, Academia Sinica

Ray-Bing Chen
Department of Statistics, National Cheng Kung University, Taiwan, R.O.C.

Ying Hung
Department of Statistics and Biostatistics, Rutgers University, New Jersey, U.S.A.

Weichung Wang
Department of Mathematics, National Taiwan University, and Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R.O.C.

Due to the expensive cost of many computer and physical experiments, it is important to carefully choose a small number of experimental points uniformly spreading out the experimental domain in order to obtain most information from these few runs. Although space-filling Latin hypercube designs (LHDs) are popular ones that meet the need, LHDs need to be optimized to have the space-filling property. As the number of design points or variables become large, the total number of LHDs grow exponentially. The huge number of LHDs makes this a difficult discrete optimization problem. In order to search optimal LHDs efficiently, we propose a population based algorithm which is adapted from the standard particle swarm optimization (PSO) and customized for LHD. Moreover, we accelerate the adapted PSO for LHD (LaPSO) via a graphic processing unit. The numerical results show that the proposed LaPSO outperforms genetic algorithm in speed, stability, and solution quality.

[Dai-Ni Hsieh, Institute of Statistical Science, Academia Sinica; dnhsieh@webmail.stat.sinica.edu.tw]

3 Some Properties of Näıve Bayes

Hyun Ji Kim
Byongsoo Choi
Sungkyunkwan University and Hansung University, Seoul, KOREA

The Näıve Bayes(NB) is a well known method for its simplicity and efficiency. This method

is an approximation to Bayes-optimal decision rule (BD) and assumes that all x variables are conditionally independent given class variable. Then, the problem is how close NB approximates to BD. We investigate this problem in the sense of entropy. We have obtained the necessary condition for the equality of the two methods. We will also investigate the improvement of NB approximation to BD when discretization is applied to numerical data.

[HyunJi Kim, Sungkyunkwan University and Hansung University, Seoul, KOREA; polaris7867@gmail.com]

4 An Improvement of CFS Feature Subset Selection

Jaeeun Lee
Hosung Lee
Sungkyunkwan University, Seoul, KOREA

CFS is a correlation-based feature subset selection algorithm (Hall, 1999) for classification. Original CFS first transforms continuous features to nominal features using MDLP(minimum description length principle) and applies bestfirst search algorithm for subset selection. In this work, we propose a new search algorithm that reduces the candidate feature space into a smaller subset. This ”search space reducing algorithm(SSRA)” can be applied to any subset selection algorithm. We will investigate the efficiency of SSRA when applied to forward selection, best-first search selection and all possible subset selection. Furthermore, we consider the efficiency of CFS when the original numeric data set is discretized using other discretization methods rather than MDLP, and also consider the efficiency when the original numeric data set is used.

[JaeEun Lee, Sungkyunkwan University, Seoul, KOREA; wosilver77@gmail.com]

5 Application of Coefficient of Intrinsic Dependence in Measuring Multivariate Association

Ya-chun Hsiao
Li-yu D Liu
National Taiwan University, Taipei, Taiwan, R.O.C.

The inquiry of methods to measure dependence among two multi-dimensional random variables has been raised, especially after the rapid development of modern high throughput technologies (e.g. microarrays) for gene expression studies. However, the application of conventional methods such as canonical correlation and projection pursuit regression have their own limitations. In this study, we proposed the coefficient of intrinsic dependence (CID) in measuring multivariable associations. It is easy to compute and applicable to various kinds of occasions. In addition, we derived a stepwise variable selection procedure accordingly. Our methods were exercised on both

simulated and real microarray data. The results showed that CID was capable to identify relevant features and then improved the accuracy of predition.

[Li-yu D Liu, Department of Agronomy, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan; lyliu@ntu.edu.tw]

6 Asymptotic Expansions of the Distributions of the Polyserial Correlation Coefficients

Haruhiko Ogasawara
Otaru University of Commerce, Japan

Asymptotic expansions of the distributions of the sample polyserial correlation coefficients and associated parameter estimators are derived up to order $O(1/N)$ when the estimators are obtained by full maximum likelihood. The asymptotic results are given under the assumption of multivariate normality for several observed continuous variables and a single unobserved variable underlining the corresponding ordered categorical variable. Asymptotic expansions of the distributions of the pivotal statistics studentized by using the estimate of the information matrix are obtained up to the order next beyond the conventional normal approximation. Numerical examples with simulations are shown in order to illustrate the accuracy of the asymptotic results in finite samples.

[Haruhiko Ogasawara, Otaru University of Commerce, 3-5-21 Midori, Otaru 047-8501 Japan; hogasa@res.otaru-uc.ac.jp]

7 On the Calculation of Win Probability for a Baseball Game

Norio Torigoe
Tokai University, Kanagawa, Japan

Win Probability Added (WPA) is one of the important facet of sabermetrics. WPA attempts to measure a player's contribution to a win by figuring the factor by which each specific play made by that player has altered the outcome of a game. It is necessary for calculate WPA to obtain Win Probability (WP) beforehand. In baseball WP basically measures the probability one team will win based on score, inning, outs, and runners on base. In this study, we build the algorithm for calculating the WP at each of the situations including batter's count and calculate the WP using the observed data taken from the result of games played in Nippon Professional Baseball (NPB) during the last five seasons. This study is supported by Data Stadium Inc.

[Norio Torigoe, Regular mailing address; torigoe@tokai-u.jp]

8 Matrix Variate Logistic Regression Model with Application to EEG Data

Hung Hung
Institute of Epidemiology and Preventive Medicine, National Taiwan University

Chen-Chien Wang
Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.

Logistic regression has been widely applied in the field of biomedical research for a long time. It aims to model the conditional probability of an event as the logit function of a linear combination of covariates. In some applications, covariates of interest have a natural structure, such as being a matrix, at the time of being collected. The rows and columns of the covariate matrix then have certain physical meanings, and they must contain useful information regarding the response. If we simply stack the covariate matrix as a vector and fit the conventional logistic regression model, relevant information may be discarded and the problem of inefficiency will arise. Motivated from this reason, we propose in this paper the matrix variate logistic (MV-logistic) regression model. The most important feature of MV-logistic regression model is that it retains the inherent structure of the covariate matrix. Another advantage is the parsimony of parameters needed. These features lead to a good performance of MV-logistic regression in many applications. Besides two simulation studies, the EEG Database Data Set is analyzed to demonstrate the usefulness of the proposed method, where the structure effects of covariate matrix are extracted, and a high classification accuracy is achieved.

[Hung Hung, Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan, R.O.C.; hhung@ntu.edu.tw]

[Chen-Chien Wang, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R.O.C.; ccwang@stat.sinica.edu.tw]

9 On Multilinear Principal Component Analysis of Order-two Tensors

Hung Hung
Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan, R.O.C.

Peishien Wu
Iping Tu
Suyun Huang
Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R.O.C.

Principal component analysis is a commonly used tool for dimension reduction in analyzing high dimensional data. Multilinear principal component analysis aims to serve a similar function for analyzing tensor structure data, and has been shown effective in reducing dimensionality both through real data analyses and through simulations. In this paper, we investigate statistical properties of multilinear principal component analysis and provide explanations for its ad- vantages. Conventional principal component analysis, which vectorizes the tensor data, may lead to inefficient and unstable prediction due to the oftentimes extremely large dimensionality involved. On the other hand, multilinear principal component analysis, in trying to preserve the data structure, searches for low-dimensional multilinear projections and, thereby, decreases dimensionality more efficiently. Asymptotic theory of order-two multilinear principal component analysis, including asymptotic efficiency and asymptotic distributions of principal components, associated projections, and the explained variance, is developed. A test of dimensionality is also proposed. Finally, multilinear principal component analysis is shown to improve conventional principal component analysis in analyzing the Olivetti Faces data set, which is achieved by extracting a more modularly-oriented basis set in reconstructing the test faces.

[Hung Hung, Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan, R.O.C.; hhung@ntu.edu.tw]

[Peishien Wu, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R.O.C.; peishien1987 @gmail.com]

[Iping Tu, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R.O.C.; iping@stat.sinica.edu.tw]

[Suyun Huang, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R.O.C.; syhuang@stat.sinica.edu.tw]

10 An Iterative Algorithm for Robust Kernel Principal Component Analysis

Hsin-Hsiung Huang
Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL, U.S.A.

Yi-Ren Yeh
Research Center for Information Technology Innovation Academia Sinica, Taipei, Taiwan, R.O.C.

We introduce a technique to improve iterative kernel principal component analysis (KPCA) robust to outliers due to undesirable artifacts such as noises, alignment errors, or occlusion. The proposed iterative robust KPCA (rKPCA) links the iterative updating and robust estimation of principal

directions. It inherits good properties from these two ideas for reducing the time complexity, space complexity, and the influence of these outliers on estimating the principal directions. In the asymptotic stability analysis, we also show that our iterative rKPCA converges to the weighted kernel principal kernel components from the batch rKPCA. Experimental results are presented to confirm that our iterative rKPCA achieves the robustness as well as time saving better than batch KPCA.

11 Cancer Mortality Analysis by the Cox Proportional Hazard Model in Nagasaki Atomic Bomb Survivors

Kenichi Yokota
Mariko Mine
Yoshisada Shibata
Atomic Bomb Disease Institute, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

We evaluated terrain shielding effects on cancer mortality in Nagasaki Atomic bomb survivors. We specified four region groups of Nagasaki atomic bomb survivors: group1 was bombed in the unshielded region at 2.0-3.0km from the hypocenter (UR2-3km, 1663 subjects), group2 was bombed in the shielded region at the similar distance from the hypocenter (SR2-3km, 2341), and the group3 and group4 was bombed in the unshielded region at 3.0-4.0km (UR3-4km, 5062) and the unshield4.0-5.0km (UR4-5km, 3313) from the hypocenter, respectively. The hazard ratio (95% C.I.) of cancer mortality in SR2-3km was 0.76 (0.63-0.93)-fold for UR2-3km. The hazard ratios in UR3-4km and UR4-5km for UR2-3km were 0.75 (0.64-0.89) and 0.80 (0.67-0.97), respectively. The results suggest the survivors bombed in the shielded area were less irradiated than those bombed in the unshielded area. The shielding by the mountains will have the same effect as the distance of 3km and over from hypocenter.

[Kenichi Yokota, 1-12-4 Sakamoto Nagasaki, Japan 852-8523; kyokota@nagasakiu.ac.jp]

12 Predictor Selection for First-order Autogressive Processes with Positive Disturbances

Ching-Kang Ing
Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.
Chiao-Yi Yang
National Central University, Taiwan, R.O.C.

We consider the problem of selecting the best predictor (from the point of view of mean squared predictionerror (MSPE)) for a first-order autoregressive process with non-negative errors. A new one-step-aheadpredictor based on the minimum of two adjacent observations is introduced

and compared with the leastsquares predictor. In particular, asymptotic expressions for the MSPEs of these two predictors are obtained.These expressions show that their rankings are crucially determined by the underlying error distribution,which is unknown in most practical situations. To resolve this difficulty, we further investigate theasymptotic performance of the accumulated prediction error (APE) of these two predictors and show thattheir rankings on the APE asymptotically equivalent to their rankings on the MSPE. As a result, the bestpredictor can be identified asymptotically through the APE. A simulation study is conducted to illustrate thepractical implication of our asymptotic result.

[Chiao-Yi Yang, National Central University, Taiwan, R.O.C.; chiaoyiy@stat.sinica.edu.tw]

17b1-Industrial Statistics
December 17 (Saturday), 13:00 - 14:30, HSS 1st Conference Room
Organizer: Hsiuying Wang
Chair: Lyinn Chung

17b1-1 Reliability in Nano-Technology and Logistics Industry

Jye-Chyi (JC) Lu
The School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, U.S.A.

This presents first reviews reliability studies in nano-technology and outlines a few future research areas. The second part of this presentation will focus on formulating reliability concepts and procedures in logistics and supply-chain industries.

[Jye-Chyi (JC) Lu, The School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205 U.S.A.; nino@math.kyushuu.ac.jp]

17b1-2 Misspecification Analyses of Gamma with Inverse Gaussian Degradation Processes

Sheng-Tsaing Tseng
Ye-Chern Yao
Institute of Statistics, National Tsing-Hua University, Taiwan, R.O.C.

Degradation models are widely used these days to assess the lifetime information of highly reliable products. In this study, motivated by a laser data, we investigate the mis-specification effect on the prediction of product's MTTF (mean- time-to-failure) when the degradation model is wrongly fitted. More specifically, we derive an expression for the asymptotic distribution of quasi MLE (QMLE) of the product's MTTF when the true model comes from gamma degradation process, but is wrongly treated as Inverse Gaussian degradation process. The penalty for the model mis-

specification can then be addressed sequentially. The result demonstrates that the effect on the accuracy of the product's MTTF prediction strongly depends on the ratio of critical value to the scale parameter of the gamma process. The effects on the precision of the product's MTTF prediction are observed to be serious when the shape and scale parameters of the gamma degradation process are large. Furthermore, we also carry out a simulation study to evaluate the penalty of the model mis-specification when the sample size and termination time are not large. It demonstrates that the simulation results are quite close to the theoretical ones.

[Sheng-Tsaing Tseng, Institute of Statistics National Tsing-Hua University Hsin-Chu, Taiwan, R.O.C. 30013; sttseng@stat.nthu.edu.tw]

17b1-3 **On the Monitoring of Mixture Linear Profiles**

Yi-Hua T. Wang
Department of Statistics, Tamkang University, Taiwan, R.O.C.

Hsiuying Wang
Institute of Statistics, National Chiao Tung University,Taiwan, R.O.C.

In some applications, the quality of a process or product is better characterized and summarized by a functional relationship between a response variable and one or more explanatory variables. Profile monitoring is used to understand and to check the stability of this relationship or curve over time. The normality assumption for the error term is commonly used in the existing simple linear profile monitoring models. However, in certain applications, the mixture nor- mal assumption for the error term may be more appropriate in real situations. Therefore, a process with mixture simple linear profiles is considered in this article. We propose new control schemes for Phase II monitoring, which are shown to have good performance in the simulation study.

[Hsiuying Wang, Institute of Statistics, National Chiao Tung University, Taiwan, R.O.C.; wang@stat.nctu.edu.tw]

17b2-Statistical Genetics
December 17 (Saturday), 13:00 - 14:30, HSS 2nd Conference Room
Organizer: Chen-Hung Kao
Chair: Chen-Hung Kao

17b2-1 **Study Genetic Basis and Pathways of Complex Traits**

Zhao-Bang Zeng
William Neal Reynolds Distinguished Professor Bioinformatics Research Center Department of Statistics and Department of Genetics North Carolina State University, U.S.A.

The genetic basis of many complex traits can be very complex. There could be multiple genes that work interactively to influence trait variation. Identification of those genes through mapping, cloning and characterization of gene pathways is an important component for many biological, biomedical and agricultural studies. Due to their complex genetic basis (multiple genes) and environmental effects, mapping of quantitative trait loci (QTL) requires complicated statistical models and analytical methods. Over the years, we have worked on developing many innovative statistical methods as well as computer softwareQTL Cartographerfor QTL mapping analysis on complex traits. We have also applied the methodology to a number of genomics studies on the genetic basis of complex traits related to speciation between Drosophila simulans and D. mauritiana, selection response in D. melanogaster, and heterosis in maize and rice. More recently, we have also developed statistical methods for mapping gene expression QTL (eQTL), for inferring gene effect networks from eQTL hotspots, and for inferring gene pathways from gene knock-out experiments. The overarching goal of our research is to systematically develop methods for inferring genetic basis and pathways for a systems oriented study of complex traits.

[Zhao-Bang Zeng, Bioinformatics Research Center, North Carolina State University; zeng@stat.ncsu.edu]

17b2-2 Optimal Significance Analysis of Microarray Data in a Class of Tests whose Null Statistic can be Constructed

Hironori Fujisawa
Institute of Statistical Mathematics, Tokyo, Japan
Takayuki Sakaguchi
Yamagata University, Yamagata, Japan

Microarray data consist of a large number of genes and a small number of replicates. We have examined testing the null hypothesis of equality of mean for detecting differentially expressed genes. The p-value for each gene is often estimated using permutation samples not only for the target gene but also for other genes. This method has been widely used and discussed. However, direct use of the permutation method for the p-value estimation may not work well, because two types of genes are mixed in the sample; some genes are differentially expressed, whereas others are not. To overcome this difficulty, various methods for appropriately generating null permutation samples have been proposed. In this paper, we consider two classes of test statistics that are naturally modified to null statistics. We then obtain the uniformly most powerful (UMP) unbiased tests among these classes. If the underlying distribution is symmetric, the UMP unbiased test statistic is similar to that proposed by Pan (2003). Under another condition, the UMP unbiased test statistic has a different formula with one more degree of freedom and therefore is expected to give a more powerful test and a more accurate p-value estimation from a modified null statistic. In microarray data, because the number of replicates is small, differences in the degree of freedom

produce large effects on the power of test and the variance of the p- value estimation. By simulation study and analysis of HIV data, we investigated the performances of the methods.

[Hironori Fujisawa, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan; fuji- sawa@ism.ac.jp]

17b3-Advances in Functional Data Analysis
December 17 (Saturday), 13:00-14:30, HSS Media Conference Room
Organizer: Jeng-Min Chiou
Chair: John Aston

17b3-1 Nonlinear Representations for Functional Data

Hans-Georg Müller
Department of Statistics, University of California Davis, U.S.A.

For functional data lying on a nonlinear low-dimensional space, concepts of manifold mean, of manifold modes of functional variation and of functional manifold components are introduced. These nonlinear representations of functional data complement classical linear representations such as eigenfunctions and functional principal components and are implemented with manifold learning methods. The performance of these nonlinear methods compares favorably to the established linear representations, if functional data lie on a manifold. For example, manifold representations provide a unified framework for timewarped functional data. The application of functional manifold representations is demonstrated with various data examples.

[Hans-Georg Müller, Department of Statistics, University of California Davis, U.S.A.; mueller @wald.ucdavis.edu]

17b3-2 Forecasting EDF Electricity Demand: A Curve Time Series Approach

Haeran Cho
Qiwei Yao
London School of Economics, London, UK

Motivated by the EDF practice of forecasting the daily demand for electricity consumption, we propose a curve time series framework which consisting of two key steps: (i) modelling seasonal patterns such as the impact of temperature semiparametrically, and (ii) modelling dynamical structure of the curves using some dimension-reduction techniques.

[Qiwei Yao, Department of Statistics, London School of Economics, Houghton
Street, London WC2A 2AE, UK; q.yao@lse.ac.uk]

17b3-3 Parameter Estimation for Ordinary Differential Equations: An Alternative View on Penalty and beyond

Naisyin Wang
Department of Statistics, University of Michigan, U.S.A.

Dynamic modeling through solving ordinary differential equations has ample applications in the fields of physics, engineering, economics and biological sciences. The recently proposed parameter-cascades estimation procedure with a penalized estimation component (Ramsay et al., 2007) combines the strengths of basis-function approximation, profile-based estimation and computation feasibility. Consequently, it has become a very popular estimation procedure. In this manuscript, we take an alternative view through variance evaluation on the penalized estimation component within the parameter-cascades procedure. We found, through some theoretical evaluation and numerical experiments, that the penalty term in the profile component could increase estimation variation. Further, contrary to the traditional belief established from the penalized spline literature, this penalty term in the ordinary differential equations setup also makes the procedure more sensitive to the number of basis functions. By taking the penalty parameter to its limit, we eliminate this problem. Our numerical experiences indicate that through more time and computation-consuming task of penalty parameter selection, the popular penalty-based method performs similarly to the method without the traditional penalty term. For other casually selected penalty parameters and numbers of basis functions, the method without penalty outperforms the penalty-based methods. We observe this phenomenon in a numerical study even when the underlying ordinary differential equations model is mis-specified. This recognition enables us to address the goodness of fit problems by considering a more flexible parameter structures across a long period of study time by using an alternative penalty structure. In this talk, we will illustrate our findings on both theoretical and numerical aspects. This is joint work with Yun Li and Ji Zhu of University of Michigan.

[Naisyin Wang, Department of Statistics, University of Michigan, U.S.A.; nwangaa@umich.edu]

17b4-IASC Committee on CS & DM in KD Special Invited Session: Massive Data Analysis Focused on Symbolic Data Analysis
December 17 (Saturday), 13:00 - 14:30, AC 1st Conference Room
Organizer: Hiroyuki Minami
Chair: Ci-Ren Jiang

17b4-1 A Feature Extraction Method for Mixed Feature-type Symbolic Data

Ikufumi Takagi
Hiroshi Yadohisa
Doshisha University, Kyoto, Japan.

Mixed feature-type symbolic data analysis (MixSDA) is one of the research area in symbolic data analysis. The main aim of the studies in MixSDA is classification or feature extraction from the symbolic data table, which is composed of some types of symbolic data. Many clustering methods and feature extraction methods have been proposed [De Carvalho De Souza 31 (2010):430–443]. In this paper, we propose a new feature extraction method for mixed feature-type symbolic data. We define a new algorithm as the feature extraction method. By using this algorithm we can comprehend not only the relations among observed units, but also the characters of the units.

[Ikufumi Takagi, Department of Culture and Information Science, Doshisha University, Kyoto, 610-0394, Japan.; dik0010@mail4.doshisha.ac.jp]

17b4-2 **Analysis for Distribution Valued Dissimilarity Data**

Masahiro Mizuta
Hokkaido University, Japan

We deal with methods for analyzing complex structured data, especially, distribution valued dissimilarity data. Nowadays, there are many requests to analyze various types of data including spacial data, time series data, functional data and symbolic data. The idea of symbolic data analysis proposed by Professor Diday covers a large range of data structures. Prof. Diday said in his book ”distributions are the number of the future.” We focus on distribution valued dissimilarity data. Typical methods for analysis of dissimilarity data are multidimensional scaling (MDS) and cluster analysis. MDS is a powerful tool for analyzing dissimilarity data. The purpose of MDS is to construct a configuration of the objects from dissimilarities between objects. In conventional MDS, the input dissimilarity data are assumed (non-negative) real values. Dissimilarities between objects are sometime given by probabilistic; dissimilarity data may be represented as distributions. We assume that the distributions between objects i and j are non-central chi-square distributions multiplied by a scalar. We propose a method of MDS under this assumption. The purpose of the proposed method is to construct a configuration with normal distributions. Cluster analysis can be widely applied in statistical data analysis and data mining. The purpose of cluster analysis is to identify groups (clusters) of individuals who have similar abilities or common views. Sometimes, we divide methods of cluster analysis into two groups: hierarchical clustering methods and nonhierarchical clustering methods. We usually use hierarchical clustering methods for dissimilarity data. An important issue for hierarchical clustering methods for distribution valued dissimilarity data is a way to compare two distributions. We propose a method for it.

[Masahiro Mizuta, Information Initiative Center, Hokkaido University, N11, W8, Kita-ku, Sapporo 060-0811; mizuta@iic.hokudai.ac.jp]

17b4-3 **Toward Complementary Application with Symbolic Data Analysis and Rough Set Theory**

Hiroyuki Minami
Hokkaido University, Japan

Symbolic Data Analysis (SDA) [Diday and Noirhomme-Fraiture(eds.) Symbolic Data Analysis and the SODAS Software (2008). Wiley] is one of the powerful idea for data analysis, which enables us to handle a set of data as it is. A typical SDA study is for interval data, however, the idea originally has a potential to handle much complex ones, based on its ”Concept” which consists of its ”intent” and ”extent”. The ”Rough set” theory [Pawlak. Rough sets. Theoretical Aspects of Reasoning about Data. Kluwer Adacemic Publishers (1991).] is known as a way to handle data according to ”possible” and ”positive” regions. This idea and SDA might share some similar feature since some studies in SDA utilize the upper and lower sets in the analysis for interval data. Both might have common features and complementary application might be developed. In the study, we revisit the originalities and reveal the specific and common features for their mixed approaches.

[Hiroyuki Minami, N11W5, Kita-ku, Sapporo 060-0811 Japan; min@iic.hokudai.ac.jp]

17b5-CSA/JSS/KSS International Session: Modeling and Inference for Biomedical/High Dimensional Data
December 17 (Saturday), 13:00 - 14:30, AC 2nd Conference Room
Organizer: Ming-Yen Cheng, Sangyeol Lee, and Jinfang Wang
Chair: Ming-Yen Cheng

17b5-1 A Two-step Procedure for Development of Powerful Robust Genetic Association Tests using Case-parents Triad Data

Jiun-Yi Wang
Asia University, Taiwan, R.O.C.

John Jen Tai
Fu Jen Catholic University, Taiwan, R.O.C.

It has been shown that under the conditional likelihood framework the efficient score tests can be derived for test of genetic association using case-parent families. Although these score tests are more powerful when the underlying model is correctly specified, they suffer from the risk of model misspecification which may lead to a loss of substantial power. To deal with this problem, a couple of approaches, which include the maxmin efficient robust test (MERT) statistic and the maximum (MAX) test statistic, were proposed. These types of statistics do hold the robustness against misspecification problem in general; however, loss of somewhat power in some situations or involvement of complex computation cannot be avoided. In this study, we adopt a weighted combination approach to construct two new robust association tests to try to enhance the power

in analysis. From the simulation results we find that the two new methods do outperform other methods in many situations.

[Jiun-Yi Wang, Department of Healthcare Administration, Asia University, Taichung 41354, Taiwan, R.O.C.; jjwang@asia.edu.tw]

17b5-2 Identification of Dissimilarities in High-dimensional Gene Regulatory Systems by Predicted Discrepancies from State Space Model

Rui Yamaguchi
University of Tokyo, Japan

Detection of differences in gene regulatory systems among individual groups of cells from gene expression data is an important task for elucidating their distinct behaviors, e.g., those after drug stimulations. However, there are fundamental difficulties to identify such dissimilar regulations by simply identifying differentially expressed genes between cells, since identical systems may produce differential expressions of genes. We developed a methodology to distinguish differentially regulated genes between case-control samples from time-course gene expression data by using a state space model (SSM) as a model of gene regulatory system. By using SSM, we can infer gene regulatory relationships and also obtain a predictive model. By employing predictive ability of SSM, we can discriminate the following two situations behind differentially expressed genes in time-course: 1) genes that are differentially expressed from the different regulatory systems for the case and control, and 2) genes that are differentially expressed from the same regulatory system but with different states of regulators. The method was applied to time-course gene expression data of human normal lung cell treated with(case)/without(control) gefitinib, an inhibitor of EGFR and found candidates of genes under differential regulations between the case and control. Furthermore, the identified gene set was applied to build a classifier for prognosis prediction of lung cancer patients and showed good performances for independent data sets. These results suggest that the proposed method is a promising tool for systems biology and translational medicine.

[Rui Yamaguchi, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan; ruiy@ims.u-tokyo.ac.jp]

17b5-3 Nonparametric Analysis of Recurrent Events with Incomplete Observation Gaps

Jinheum Kim
University of Suwon, Korea

Eun hee Choi
Yonsei University College of Medicine, Korea

Yanh-Jin Kim
Sookmyung Women's University, Korea

This talk is concerned with the analysis of recurrent events with possibly multiple observation gaps. These gaps may occur whenever a subject temporarily disappears from a study. If the duration of the observation gaps is completely known, we can simply define the subject's contribution to a risk set as zero during the period. Unfortunately the information about the duration sometimes is collected of incomplete form. We treat these incomplete observation gaps as so-called interval-censored data. We proposed an estimate for a cumulative mean function and derive its asymptotic distribution. Additionally we investigate its finite-sample performance through a simulation. We also propose a weighted log-rank test for comparing the cumulative mean functions of several populations and investigate an empirical size and power of the proposed test. We illustrate our proposed methods to the Young Traffic Offenders Program (YTOP) taken from Sun et al. (2001).

[Jinheum Kim, Department of Applied Statistics, University of Suwon, Hwaseong, 445-743, Korea; jkimdt65@gmail.com]

17b6-Time Series and Spatial Data Analysis
December 17 (Saturday), 13:00 - 14:30, AC 3rd Conference Room
Organizer: Hsin-Cheng Huang
Chair: Uwe Hassler

17b6-1 **Threshold Modeling of Martingale Differences**

Kung-Sik Chan
Department of Statistics and Actuarial Science, University of Iowa, U.S.A.

Martingale differences play a key role in probability and statistics. For them, this paper develops systematically the theory and practice of a simple yet versatile multiple-threshold model, or a TMD model for short. Advantages enjoyed by the TMD model include the almost complete free choice of the model parameters while maintaining stationarity, the availability of fairly comprehensive theoretical underpinnings for statistical practice, which are supported by pertinent computation procedures and algorithms. We report some of our experiences in TMD modelling of real time series from diverse disciplines and highlight some insights that the new approach is capable of shedding. Finally, we indicate potentials of the approach to multivariate time series as well as random fields. (This talk is based on joint work with D. Li, S. Ling and H. Tong)

[Kung-Sik Chan, Department of Statistics and Actuarial Science, University of Iowa, U.S.A.; kchan@stat.uiowa.edu]

17b6-2 On Estimating the Parameters in Conditional Heteroskedasticity Models by Empirical Likelihood Estimation

Tsung-Lin Cheng
Department of Mathematics, National Changhua University of Education, Taiwan, R.O.C.

In most time series models, the data sets that we might be confront with are not statistically independent. While the celebrated empirical likelihood (EL) estimation proposed by owen (1988) has been widely used in a framework of independent data without having to know the distribution of the population, it is also challenging to apply EL estimation to the models with dependent data. In this talk, we will exploit EL method to estimate the parameters emerging in some important econometrical models including ARCH, GARCH, EGARCH and TGARCH. In addition, we conduct some illustrative simulations to compare EL approach with other methods of estimation (e.g. MLE and OLS). Finally, we analyze the data of the West Texas Intermediate (WTI) Crude Oil Prices by fitting it into the GARCH model.

[Tsung-Lin Cheng, Department of Mathematics, National Changhua University of Education, Taiwan, R.O.C. ; tlcheng@cc.ncue.edu.tw]

17b6-3 Quasi-likelihood Scan Statistics for Detection of Spatial Clusters with Covariates

Pei-Sheng Lin
National Health Research Institutes

In this talk, we use a generalized linear mixed model, which simultaneously includes geographic clusters, covariates and spatial correlation in the model, for analysis of spatial clusters. We combine quasi-likelihood estimating equations and scan statistics to develop a method for multiple cluster detection. In the proposed method, parametric bootstrapping and quasi-deviance criteria are used to select hot/cool clusters for rejection. The proposed estimates of cluster coefficients are consistency even under misspecified cluster regions and correlation models. We conduct simulations to evaluate performance of the proposed method. The method is applied to an enterovirus data set in Taiwan for illustration.

[Pei-Sheng Lin, National Health Research Institutes; pslin@nhri.org.tw]

17b7-Statistical Computing in Multisensory Network Systems
December 17 (Saturday), 13:00 - 14:30, AC 4th Conference Room
Organizer: Yunmin Zhu
Chair: Yunmin Zhu

17b7-1 Optimal Dimensionality Reduction of Sensor Data in Multisensor Estimation Fusion

Enbin Song
Mathematical College, Sichuan University Chengdu 610064

When there exists the limitation of communication bandwidth between sensors and a fusion center, one needs to optimally pre-compress sensor outputs-sensor observations or estimates before sensors' transmission to obtain a constrained optimal estimation at the fusion center in terms of the linear minimum error variance criterion. We will give an analytic solution of the optimal linear dimensionality compression matrix for the single sensor case and analyze the existence of the optimal linear dimensionality compression matrix for the multisensor case, as well as how to implement a Gauss-Seidel algorithm to search for an suboptimal solution to linear dimensionality compression matrix.

[Enbin Song, Mathematical College, Sichuan University Chengdu 610064; e.b.song@163.com]

17b7-2 A Finite Iterative Algorithm for Best Linear Unbiased Estimation Fusion in Distributed Systems

Jie Zhou
Sichuan University, Chengdu, China

We propose a new formula for the best linear unbiased estimation (BLUE) of the generalized Gauss-Markov model and develop its iterative algorithm. As applications, we provide a finite iterative algorithm for the BLUE fusion in the distributed multi-sensor systems, in which the square matrix involving the inverse or generalized inverse has the same size as the parameter or state at each iteration. The proposed algorithm not only is a generalization of Bar-Shalom and Campo's fusion formula in more general case but also has low computational complexity for the large sensor networks.

[Jie Zhou, Sichuan University, Chengdu, China; jzhou@scu.edu.cn]

17b7-3 Diffusion based EM Algorithm for Distributed Estimation of Gaussian Mixtures in Wireless Sensor Networks

Yang Weng
Math. Dept., Sichuan University

Distributed estimation of Gaussian mixtures has many applications in wireless sensor network (WSN), and its energy-efficient solution is still challenging. This paper presents a novel diffusion

based EM algorithm for this problem. A diffusion strategy is introduced for acquiring the global statistics in EM algorithm in which each sensor node only needs to communicate its local statistics to its neighboring nodes at each iteration. This improves the existing consensus based distributed EM algorithm which may need much more communication overhead for consensus, especially in large scale networks. The robustness and scalability of the proposed approach can be achieved by distributed processing in the networks. In addition, we show that the proposed approach can be considered as a stochastic approximation method to find the maximum likelihood estimation for Gaussian mixtures. Simulation results show the efficiency of this approach.

[Yang Weng, Math. Dept., Sichuan University; wengyang@scu.edu.cn]

17b7-4 Distributed Kalman Filtering Fusion with Packet Loss or Intermittent Communications

Yingting Luo Yunmin Zhu Xiaojing Shen Enbin Song
College of Mathematics, Sichuan University, Chengdu, Sichuan, P. R. China

This paper considers the distributed Kalman filtering fusion with passive packet loss or initiative intermittent communications from local estimators to fusion center while the process noise does exist. When the local estimates are not lost too much, we propose an optimal distributed fusion algorithm which is equivalent to the corresponding centralized Kalman filtering fusion with complete communications even if the process noise does exist. When this condition is not satisfied, based on the above global optimality result and sensor data compression, we propose a suboptimal distributed fusion algorithm. Numerical examples show that this suboptimal algorithm still works well and significantly better than the standard distributed Kalman filtering fusion subject to packet loss even if the process noise power is quite large.

[Yunmin Zhu, College of Mathematics, Sichuan University, Chengdu, Sichuan, P. R. China; ymzhu@scu.edu.cn]

17c1-ISBIS (The International Society for Business and Industrial Statistics) Special Invited Session: Reliability Evaluation
December 17 (Saturday), 14:40 - 16:10, HSS 1st Conference Room
Organizer: Guoying Li
Chair: May-Ru Chen

17c1-1 System Reliability Evaluation in Absence of Partial Components Test Data

Dan Yu
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China

In this research, a novel system reliability evaluation method is proposed under the situation that life test data is incomplete for part of its components. Specifically, for components with complete life data, we can have their Fiducial distributions through Fiducial inference approach. While for components with incomplete life data, a novel approach based on the reliability prediction is applied. With proper assumptions, a systematic evaluation approach is proposed. Simulation results show the applicability and effectiveness of this approach.

[Dan Yu, 306, Siyuan Building, East Zhongguancun Road, Haidian, Beijing, China; dyu@amss.ac.cn]

17c1-2 An Adaptive Design to Assess the Reliability of the Pyrotechnic Control Subsystem in Opening the Solar Array

Yubin Tian
Dianpeng Wang
School of Sciences, Beijing Institute of Technology, Beijing, 100081, China

It is important to assess the reliability of the pyrotechnic control subsystem in opening the solar array of a satellite. Such an assessment requires determining the level of a control factor that is needed to cause the desired response and energy outputs with high probability. We propose a two-phase adaptive design to estimate the level of interest. We also prove the convergence of the design. A simulation study shows that the estimate is accurate and robust. The design is used to assess the reliability of a real pyrotechnic control subsystem.

[Yubin Tian, Department of Mathematics, Beijing Institute of Technology, P. R. China; tianyb@bit.edu.cn]

17c1-3 Software Reliability Modeling and Analysis with Non-homogeneous Poisson Processes

Qingpei Hu
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China

Software reliability evaluation is critical for the decision making on the development process. Usually non-homogenous Poisson process is adopted for the associated fault detection process modeling. A comprehensive review on the application of NHPP is conducted, with focus on the recent progress on this research topic. The application is illustrated with practical software projects data.

[Qingpei Hu, 307, Siyuan Building, East Zhongguancun Road, Haidian, Beijing, China; qingpeihu@amss.ac.cn]

17c2-Statistical Bioinformatics
December 17 (Saturday), 14:40 - 16:10, HSS 2nd Conference Room
Organizer: Jae Won Lee
Chair:Yih Huei Steven Huang

17c2-1 Clustering, Classification and Ordering via a Graph Theory

Choongrak Kim
PU.S.A.n National University, BU.S.A.n, S.KOREA

Assume that adjacency (closeness) between any pairs of n observations are given. Then we can construct a Laplacian matrix based on the adjacency matrix, and can show that the clustering, classification and ordering information on n observations are contained in the Fiedler vector which is an eigenvector corresponding to the non-zero smallest eigenvalue of the Laplacian matrix. By using the idea of hard-thersholding we show that clustering, classification and ordering have conceptually the same spirit, and we apply this result to clustering and classification problem of microarray data, and genetic map construction.

[Choongrak Kim, Department of Statistics, PU.S.A.n National University, Jangjeon Dong, Gumjeong Gu, BU.S.A.n, KOREA, 609-735; crkim@pU.S.A.n.ac.kr]

17c2-2 Minimax Estimation for Mixtures of Wishart Distributions

Ja-Yong Koo
L.R. Haff
P.T. Kim
Donald Richards
Korea University, Korea

The space of positive definite symmetric matrices has been studied extensively as a means of understanding dependence in multivariate data along with the accompanying problems in statistical inference. Many books and papers have been written on this subject, and more recently there has been considerable interest in high-dimensional random matrices with particular emphasis on the distribution of certain eigenvalues. Our present paper is motivated by modern data acquisition technology, particularly, by the availability of diffusion tensormagnetic resonance data. With the availability of such data acquisition capabilities, smoothing or nonparametric techniques are required that go beyond those applicable only to data arising in Euclidean spaces. Accordingly, we present a Fourier method of minimax Wishart mixture density estimation on the space of positive definite symmetric matrices.

[Ja-Yong Koo, Korea University; jykoo@korea.ac.kr]

17c2-3 Exploration of Gene-Gene Interactions in Case-Control Study Using Information Gain

Jaeyong Yee
Mira Park
Eulji University, Daejeon, Korea

Detection of gene-gene interactions has become a hot topic in complex disease genetics in recent years. Various methods for the detection and characterization of these interactions have been proposed including regression-based analysis, neural networks, and multidimensional reduction (MDR). In this study, we developed an entropy-based method to explore the interaction effects. We define an entropy measure which is based on a two-way contingency table of the trait and the genotype combination, and calculate the relative information gain. The measure is standardized to compare directly the intensities of interactions in different dimensions. The empirical p-values are calculated for selected SNPs by permutation method. We illustrate this method using both artificial and real genotype data. Simulation study is conducted to compare the power of proposed method and two types of MDR. Successful identification of the genetic associations and the gene-gene interactions has been accomplished.

[Mira Park, 143-5 Yongdu-dong, Jung-gu, Daejeon 301-832 KOREA; mira@eulji.ac.kr]

17c3-Machine Learning and Its Applications
December 17 (Saturday), 14:40 - 16:10, HSS Media Conference Room
Organizer: Su-Yun Huang
Chair: Chen-Hsiang Yeang

17c3-1 Projective Power Entropy Based Learning for Unsupervised Data

Shinto Eguchi
Osamu Komori
Institute of Statistical Mathematics, Tokyo, Japan

Akifumi Notsu
Graduate University of Advanced Studies, Tokyo, Japan

The projective power entropy is a one-parameter family of generalized entropy including the Boltzmann-Shannon entropy with a case of power index 0. The minimum principle of projective power entropy for a given data set provides a statistical method, which is parallel to that of Boltzmann-Shannon entropy, or equivalently the maximum likelihood. For example, under a normal distribution model, it gives one parameter family of estimators for the normal mean. If the power index equals 0, the estimator is nothing but the sample mean; if the power index is positive,

the estimator becomes a reweighed iterative mean. Surprisingly, the reweighed iterative mean has k equilibrium solutions if the data comes from *k* separate means. We discuss a close relation of Lebesgue L_p space and the projective power entropy. The minimum projective power entropy principle for unsupervised data is proposed to new approaches for boosting of density estimation and *k*-means clustering.

[Shinto Eguchi, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan; eguchi@ism.ac.jp]

17c3-2 Multiple Kernel Learning for Dimensionality Reduction

Yen-Yu Lin
Research Center for Information Technology Innovation, Academia Sinica

This talk focuses on our research effort in improving the performances of a set of computer vision applications by using multiple, heterogeneous features of data. In solving complex vision tasks, adopting multiple descriptors to more precisely characterize the data has been a feasible way for improving performance. The resulting data representations are typically high dimensional and assume diverse forms. Thus finding a way to transform them into a unified space of lower dimension generally facilitates the underlying tasks, such as supervised recognition or unsupervised clustering. To this end, we propose an approach (called MKL-DR), which generalizes the framework of multiple kernel learning for dimensionality reduction, and distinguishes itself with the following three main contributions. First, it provides the convenience of using diverse descriptors to describe useful characteristics of various aspects about the data. Second, it extends a broad set of existing dimensionality reduction techniques to consider multiple kernel learning, and consequently improves their effectiveness. Third, the formulation of MKL-DR introduces a new class of applications with the multiple kernel learning framework to address not only the supervised learning problems but also the unsupervised and semi-supervised ones.

[Yen-Yu Lin, Research Center for Information Technology Innovation, Academia Sinica; yylin@citi.sinica.edu.tw]

17c3-3 Global Quantification of Mammalian Gene Expression Control

Bjrn Schwanhusser
Dorothea Busse
Na Li
Gunnar Dittmar
Johannes Schuchhardt
Jana Wolf
Wei Chen
Matthias Selbach

Max Delbrueck Center for Molecular Medicine, Robert-Roessle-Str. 10, D-13092 Berlin, Germany

Gene expression is a multistep process that involves the transcription, translation and turnover of messenger RNAs and proteins. Although it is one of the most fundamental processes of life, the entire cascade has never been quantified on a genome-wide scale. Here we simultaneously measured absolute mRNA and protein abundance and turnover by parallel metabolic pulse labelling for more than 5,000 genes in mammalian cells. Whereas mRNA and protein levels correlated better than previously thought, corresponding half-lives showed no correlation. Using a quantitative model we have obtained the first genomescale prediction of synthesis rates of mRNAs and proteins. We find that the cellular abundance of proteins is predominantly controlled at the level of translation. Genes with similar combinations of mRNA and protein stability shared functional properties, indicating that half-lives evolved under energetic and dynamic constraints. Quantitative information about all stages of gene expression provides a rich resource and helps to provide a greater understanding of the underlying design principles

[Wei Chen, Max Delbrueck Center for Molecular Medicine, Robert-Roessle-Str. 10, D-13092 Berlin, Germany; wei.chen@mdc-berlin.de]

17c4-Statistics in Systems Biology
December 17 (Saturday), 14:40 - 16:10, AC 1st Conference Room
Organizer: Seiya Imoto, Rui Yamaguchi
Chair: Seiya Imoto, Rui Yamaguchi

17c4-1 Prediction for Severe Adverse Drug Events by Systems Biology and Statistical Learning

Pei-Ling Liu
Liang-Chin Huang
Henry Horng-Shing Lu
Institute of Statistics, National Chiao Tung University, Taiwan, R.O.C.

We propose an integrated approach based on systems biology and statistical learning for severe adverse drug events (ADEs) prediction in this study. This study utilizes systems biology informatics for Drugomics feature extraction of 1163 drugs based on DrugBank, HPRD PPIDB, KEGG and GODB. The FDA report system provides historical ADEs among actual patients. Among all the adverse patient responses associated with properly prescribed medicine, we choose a set of severe ADEs such as death, life threatening events and toxicity to build the prediction models. We use advanced statistical learning methods to predict the ADE incidences in the general population.

[Henry Horng-Shing Lu, Institute of Statistics, National Chiao Tung University, Hsinchu 30010, Taiwan; hslu@stat.nctu.edu.tw]

17c4-2 **Bayesian Supercomputing Tackles Cancers**

Ryo Yoshida
The Institute of Statistical Mathematics, Research Organization of Information and Systems

The advent of high speed DNA sequencers is bringing into reality a new era of personal genome and personalized medicine. Within the next several years, this technological breakthrough will cause significant changes to every -omics studies in terms of both quantity and quality. To reveal a complex world of cellular systems from such a vast amount of information, we aim to create a new research infrastructure of data assimilation systems involving experimental bioscience, modeling, simulation, and state-of-art data science. The invention of DNA microarray chip has enabled us to measure transcript levels of more than 20,000 genes in human genome, simultaneously. Our collaborators (Professor Miyano and his research team at Institute of Medical Science, the University of Tokyo) successfully monitored time-dependent change of all genes in two phenotypes of human lung cancers under the treatment with an anticancer agent; one demonstrating exquisite sensitivity to the drug treatment and the other being resistant. Recently, several studies have advocated relatively rapid acquisition of resistance within a few years after initiation of the anticancer drug treatment. The experimental data that we obtained will certainly be vital to uncovering molecular basis of maintaining the viability of drug resistant cancer population. For the first time in the world, we succeeded in the development of whole gene transcription simulators that are highly reproducible to the observed gene expression of the drug sensitive and resistant cancer cells. The developed simulation models will be utilized in the discovery of key molecules that are promising for pharmacological treatment to improve drug efficacy, and a principle of action in maintaining the viability of the acquired drug resistance. High-throughput generation of large-scale simulators so as to reproduce observed data on 20,000 endogenous variables would be a quite hard challenge, computationally and statistically, which is to be addressed with state-of-the art supercomputer systems. Some researchers in our team are now developing a new life science data assimilation system as members involved in a national project on the Next Generation of Supercomputer 'K'.

[Ryo Yoshida, Institute of Statistical Mathematics, Research Organization of Information and Systems; yoshidar@ism.ac.jp]

17c4-3 **Gene Network Analysis in Cancer Biology**

Hiromitsu Araki
Bioinformatics of Disease, Department of Molecular Medicine & Pathology, School of Medical Sciences, Faculty of Medical and Health Sciences, University of Auckland

Cancer is composed of diverse dysregulated molecular pathways involving cell cycle, apoptosis,

DNA repair, etc.. In cancer cells, a large number of genes are differentially expressed compared to normal cells, and they contribute such dysregulation of molecular pathways. The identification of master regulators from such huge gene lists might provide clues as to underlying molecular pathogenesis of cancer, which is also informative for cancer drug discovery. This presentation will talk about our genome-wide gene network analysis by using Bayesian networks applying to microarray data of cancer cells to identify prognosis markers. The talk will also provide a dynamic gene network analysis applying to time course microarray data of anti-tumor compounds treated cells to reveal compound mode of actions.

[Hiromitsu Araki, Bioinformatics of Disease, Department of Molecular Medicine & Pathology, School of Medical Sciences, Faculty of Medical and Health Sciences, University of Auckland; hiromitsu araki@cell-innovator.com]

17c5-CSA/JSS/KSS International Session: Multivariate Analysis
December 17 (Saturday), 14:40 - 16:10, AC 2nd Conference Room
Organizer: Ming-Yen Cheng, Sangyeol Lee, and Jinfang Wang
Chair: Jinfang Wang

17c5-1 **Effective Methodologies for High-Dimensional Statistical Inference**

Makoto Aoshima
University of Tsukuba, Japan

A common feature of high-dimensional data such as genetic microarrays is that the data dimension is extremely high, however the sample size is relatively small. This type of data is called HDLSS data. In this talk, we present modern multivariate statistical methodologies that are very effective to draw statistical inference from HDLSS data. We first consider PCA for HDLSS data. We previously gave the cross-data-matrix (CDM) methodology by Yata and Aoshima [J. Multivariate Anal. 101 (2010): 2060-2077]. We introduce the generalized CDM (GCDM) methodology by Aoshima and Yata [Prostate Cancer (2011a): in press]. We apply GCDM to clustering for HDLSS data. Next, we consider classification for HDLSS data. We pay special attention to the quadratic-type classification methodology by Aoshima and Yata [Sequential Anal. 30 (2011b): Editor's Special Invited Paper]. We give a sample size determination for the classification so that the misclassification rates are controlled by a prespecified upper bound.

[Makoto Aoshima, Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan; aoshima@math.tsukuba.ac.jp]

17c5-2 **Flexible Generalized Varying Coefficient Regression Models**

Byeong U. Park
Seoul National University, Korea

This paper studies a very general model that can be used widely to analyze the relation between a response and multiple covariates. The model is nonparametric, yet renders easy interpretation for the effects of the covariates. The model accommodates both continuous and discrete random variables for the response and covariates. It is quite flexible to cover both generalized varying coefficient models and generalized additive models as a special case. We propose a powerful technique of fitting the model and discuss its theoretical properties. In particular, we establish identifiability of the model under weak conditions. The estimators of the nonparametric functions are given as solutions of a system of nonlinear integral equations. We provide a double iteration scheme to solve the system of equations and prove its fast algorithmic convergence. We also derive the limit distributions of the function estimators. Finally, we illustrate the method with a data example.

[Byeong U. Park, Department of Statistics, Seoul National University, Seoul 151-747, Korea; bupark@stats.snu.ac.kr]

17c5-3 A Monotonic Constrained Regression Framework for Image Contrast Enhancement

Lu-Hung Chen
National Chung Hsing University, Taiwan, R.O.C.

This paper introduces a general framework for image contrast enhancement based on histogram equalization (HE) and specification (HS). Traditional HE and HS are simple and effective, but they often amplify the noise level of the image while enhancing it. Furthermore, they may not utilize the entire dynamic range due to the discrete nature of the image. In our framework, image contrast enhancement is posed as a nonparametric monotonic constrained regression problem, in which both the two boundary values and the slopes of the brightness transform function are controlled. We show that such a framework provides an effective way to avoid enlarging the noise level and to utilize the entire dynamic range while performing HS (and also its special case HE). Our method can thus reduce the production of visual artifacts while enhancing the image.

[Lu-Hung Chen, Department of Applied Mathematics, National Chung Hsing University, Taichung, Taiwan; Email: luhung.chen@statistics.twbbs.org]

17c6-The Bernoulli Society Special Invited Session: Mathematical Finance, Applied and Theoretical Developments
December 17 (Saturday), 14:40 - 16:10, AC 3rd Conference Room
Organizer: Arturo Kohatsu-Higa
Chair: Te-Hsin Liang

17c6-1 Dynamic Cointegrated Pairs Trading: Time-Consistent Mean-Variance Strategies

Hoi Ying Wong
Department of Statistics, The Chinese University of Hong Kong

Cointegration is a useful econometric tool for identifying assets which share a common equilibrium. Cointegrated pairs trading is a trading strategy which attempts to take a profit when cointegrated assets depart from their equilibrium. This paper investigates the optimal dynamic trading of cointegrated assets using the classical mean-variance portfolio selection criterion. To ensure rational economic decisions, the optimal strategy is obtained over the set of time-consistent policies from which the optimization problem is enforced to obey the dynamic programming principle. We solve the optimal dynamic trading strategy in a closed-form explicit solution. This analytical tractability enables us to prove rigorously that cointegration ensures the existence of statistical arbitrage using a dynamic time-consistent meanvariance strategy. This provides the theoretical grounds for the market belief in cointegrated pairs trading. Comparison between time-consistent and precommitment trading strategies for cointegrated assets shows the former to be a persistent approach, whereas the latter makes it possible to generate infinite leverage once a cointegrating factor of the assets has a high mean reversion rate. (Joint with Dr. M.C. Chiu)

[Hoi Ying Wong, Department of Statistics, The Chinese University of Hong Kong; hywong@cuhk.edu.hk]

17c6-2 **Option Prices in terms of Probability Functions**

Ju-Yi Yen
Vanderbilt University, Nashville, Tennessee, U.S.A.

Marc Yor
Universite Pierre et Marie Curie, Paris, France

The Black-Scholes model is an important starting point for studying nancial derivatives. In the Black-Scholes formula, the evolution of prices of a risky asset is described by an exponential martingale associated to a Brownian motion and, as a consequence, the Black-Scholes function is increasing and bounded and can be written as a distribution function. We shall explore the connection between Black-Scholes functions and their distribution functions. We study the distribution function in terms of the last passage times, and extend the underlying martingale beyond the Brownian framework. Explicit examples of computations of these laws are given.

[Ju-Yi Yen, Department of Mathematics, Vanderbilt University, Nashville, Tennessee 37240, U.S.A.; ju-yi.yen@vanderbilt.edu]

17c6-3 A First-Passage Model of Default with Markovian Credit Migration

Cheng-Der Fuh
Graduate Institute of Statistics,National Central University,Jhongli, Taiwan, R.O.C.

Rating systems imply that firms with distinct ratings have different chances to default, but it is clear that firms within the same rating differ in default probabilities, generating both intra-rating spread and inter-rating overlap in observed yields. We propose a credit risk model that incorporates both exogenous default jumps and endogenous defaults via the firm value process. Closed-form approximations for expected default time and tail probabilities are derived with empirically testable yield consequences. Our model demonstrates strong empirical fit as the S-shaped yield curve and aforementioned spread and overlap are captured. We conclude with an application that may be used to evaluate rating systems (Joint work with Charles Chang and Chu-Lan Kao).

[Cheng-Der Fuh, Graduate Institute of Statistics,National Central University,Jhongli, Taiwan, R.O.C.; cdfuh@cc.ncu.edu.tw]

17c7-Statistical Graphics and Visualization
December 17 (Saturday), 14:40 - 16:10, AC 4th Conference Room
Organizer: Dae-Heung Jang
Chair: Han-Ming Wu

17c7-1 Visualizing Trade-offs Between Multiple Objectives: Tools to Help Decision-Makers

Christine M. Anderson-Cook
Los Alamos National Laboratory, Los Alamos, New Mexico, U.S.A.

When decision-makers make important decisions, they are often forced to balance competing objectives that require weighing the importance of different alternatives and assessing the merits of different choices. We present a suite of graphical tools, based on the Pareto front multiple criteria optimization method, which allow the trade-offs between choices to be compared and assessed. The tools are presented in the context of two examples: a designed experiment where good estimation and protection against model misspecification are considered; and a resource allocation problem about what future data to collect when evaluating reliability for a population of systems based on several different data types.

[Christine M. Anderson-Cook, P.O. Box 1663, Los Alamos, NM 87545, U.S.A.; candcook@gmail.com]

17c7-2 Exploring Symbolic Data Structure Using Matrix Visualization

Yin-Jing Tien
Chun-houh Chen
Academia Sinica, Taipei, Taiwan, R.O.C.

Chiun-How Kao
Chuan-kai Yang
National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C.

Junji Nakano
The Institute of Statistical Mathematics, Tokyo, Japan

Sheau-Hue Shieh
National Taipei University, New Taipei City, Taiwan, R.O.C.

Matrix visualization (MV) techniques such as GAP (generalized association plots) are useful exploratory data analysis tools for visualizing and clustering high-dimensional data structure. However all available MV methods treat samples as independent ones and as base units. In the current study we are enhancing GAP environment for handling data with hierarchical (multi-level) structure and for huge data sets. Both tasks, modules for hierarchical and huge data, are closely related to that of symbolic data analysis (SDA). In this presentation we will summarize our development on matrix visualization for symbolic data analysis with some examples.
[Yin-Jing Tien, 128 Academia Road Sec.2, Nankang Taipei 115 Taiwan, R.O.C.; gary@stat.sinica.edu.tw]

17c7-3 Correlation-Based r-plot for Evaluating Supersaturated Designs

Dae-Heung Jang
Dept. of Statistics, Pukyong National University, KOREA

Christine M. Anderson-Cook
Los Alamos National Laboratory, U.S.A.

Youngil Kim
Division of Business Administration, Chungang University, Seoul, Korea

Orthogonality or near-orthogonality is an important property in the designed experiments. Supersaturated designs are natural when we wish to to investigate main effects for a large number of factors, but are restricted to a small number of runs. These supersaturated designs can not

satisfy orthogonality. Hence, we need a means to evaluate the degree of near-orthogonality of given supersaturated designs. It is usual to use numerical measures to assess the degree of the orthogonality of given supersaturated designs, but we propose using graphical methods to evaluate the degree of near-orthogonality to compare and select between supersaturated designs.

[Youngil Kim, Division of Business Administration, Chungang University, Seoul, Korea; yik01@cau.ac.kr]

17d1-Algorithmic Trading: Methods and Models
December 17 (Saturday), 16:30 - 17:30, HSS 1st Conference Room
Organizer: Philip Leung-ho Yu
Chair: Philip Leung-ho Yu

17d1-1 Complex Trading Strategy Based on Particle Swarm Optimization

Fei Wang
Philip L.H. Yu
David W.L. Cheung
The University of Hong Kong, Hong Kong

Trading rules have been utilized in the stock trading market to make profit for more than a century. However, only using a single trading rule is not sufficient to predict the stock price trend. Although some complex trading strategies combining different class of trading rules are proposed, they only include one rule for a class, which may lose information from other combinations of different class of rules. In this paper, we propose a complex trading strategy combining the two of the most popular trading rules - Moving Average and Trading Range Break-out. For each of the class, we include different combinations of the rule parameters to get a universe of 140 trading rules in all. Each rule is assigned a start weight and a self-adaptive reward system according to profit is proposed to modify these rules' weights during trading. To get the appropriate combination of the start weights and reward system, a time variant Particle Swarm Optimization (PSO) algorithm is used to optimize the proposed complex trading strategy in terms of the annual net profit. The experiments show that the proposed complex trading strategy outperforms the best Moving Average and Trading Range Break-out rules after optimization using PSO.

[Philip L.H. Yu, Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong; plhyu@hku.hk]

17d1-2 Improving the Power of Stepwise Procedures in Multiple Inequalities Testing– Evidence from Global CTA Funds

Stephen G. Donald
Yu-Chin Hsu
Department of Economics University of Texas at Austin

Chung-Ming Kuan
Department of Finance National Taiwan University

Stéphane Meng-Feng Yen
Department of Accountancy and Institute of Finance National Cheng Kung University

In the multiple hypotheses testing problem, the ability of a procedure to identify false hypotheses is as important as that to control the number of false rejections. When the number of the hypotheses is large, control of the familywise error rate (FWE), the probability of rejecting at least one true hypothesis, becomes so stringent that the ability of the procedure to identify false hypotheses is limited. To increase the power of a procedure, one possibility is to allow for more than one false rejection or equivalently to control the probability of k or more false rejections, k-FWE. In this article, we propose the so-called Generalized-k Step-SPA test, aiming to improve on the testing power of the Generalized-k Step-RC, Romano, Shaikh and Wolf (2008, *Econometric Theory*). The Generalized-k Step-SPA test is a generalized version of Step-SPA test of Hsu, Hsu and Kuan (2010, *Journal of Empirical Finance*) since it can control the k-FWE rate at any given level asymptotically instead of the FWE. It is also identical to the Step-SPA test when k equals 1. To demonstrate how our proposed test tends to reject more false hypotheses than the Generalized-k Step-RC, we apply both approaches to 315 Commodity Trading Advisor (CTA) funds with AUM of at least $20 million during the July-1994 to June-2010 period. We observe that the proposed Generalized-k step-SPA test tends to select more outperforming CTAs than the Generalized-k step-RC in numerous different cases. The increase in the number of identified outperforming CTAs tends to be more pronounced when allowing for a larger k-FWE. Although the increase in the test power is not economically large, a marginal increase in test power suggests a significant increase in the selection rate for elite funds given the small mean sample size (i.e. number of funds) across all in-sample windows. This result might carry a non-trivial contribution from the perspective of asset management: certainly, every individual money or portfolio manager does not want to miss any potential fund that really performs well during any period of time. We also test the performance persistence of outperforming CTAs identified by the proposed Generalized-k step-SPA test. Our empirical results suggest performance persistence of selected good CTAs. Moreover, the equal-weighted portfolio of outperforming CTAs selected from the discretionary CTA family tends to perform better than those selected from the systematic CTA family regardless of the factor model used to measure their performance.

[Stéphane Meng-Feng Yen, Department of Accountancy, National Cheng Kung University, Taiwan, R.O.C.; yenmf@mail.ncku.edu.tw]

17d2-Finance
December 17 (Saturday), 16:30 - 17:30, HSS 2nd Conference Room
Chair: Yen-Hung Chen

17d2-1 On the Pricing of Investment Corporation Bonds

Masakazu Ando
Caiba Institute of Technology, Chiba, Japan

Hiroshi Tsuda
Doshisha University, Kyoto, Japan

Yoko Tanokura
The Institute of Statistical Mathematics, Tokyo, Japan

Seisho Sato
The Institute of Statistical Mathematics, Tokyo, Japan

Genshiro Kitagawa
Research Organization of Information and Systems, Tokyo, Japan

In the present study, it is a purpose ot evaluate the pricing model of the investment corporation bond that the investment corporation issued for the funding by using corporate bond pricing model SCBCSM (Straight Coupon Bond Cross-Sectional Market) proposed by Tsuda (2006). This model is estimated the implied default probability and the recovery rate from the market price simultaneously. The cap rate and the amount of market value of the property that the investment corporation bad were presumed, and the recovery rate model presumption was tried there now. A new finding like the term structure of the default probability at the time of each bond ranking and the effectiveness of the pricing moel was obtained from the market price data of the investment corporation bond.

[Masakazu Ando, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan; andomasa@sun.it- chiba.ac.jp]

17d2-2 Verification of the Effectiveness of Dollar-cost Averaging Investment Method

Kaoru Fueda
Masayuki Touji
Graduate School of Environmental Science, Okayama University, Okayama, Japan

As a method of long term investment for private investor, the dollar cost averaging investment method is well known and seems to reduce the purchase cost because we purchase risk assets with same amount of monye every month, then we purchase many assets when the price of assets is low and few assets when the price is high. On the other hand, if the expectation of the return of the risk assets is positive, we have the maximum expectation of return when we purchase the risk assets with all of money to invest. To reduce the risk of investment, diversified investments are effective. However question whehter we use the dollar cost averaging investment method or investe money all at once to well-diversified risk assets remains. In this study, we validate the effecte of the ddollar cost averaging investment method by Monte Carlo simulation.

[Kaoru Fueda, Naka 3-1-1, Kita-ku, Okayama, 700-8530, Japan.; fueda@ems.okayama-u.ac.jp]

17d2-3 Relationship Between Capital Structure and Financial Performance: An Application by Data Mining

Nermin Ozgulbas
Baskent University Faculty of Health Sciences Department of Healthcare Management, Turkey

Ali S. Koyuncugil
Capital Markets Board of Turkey, Ankara, Turkey

The aim of this study is to examine the relationship between capital structure and financial performance of small and medium-sized enterprises (SMEs) by using data mining. The capital structure refers to the percentage of capital that used for financing. Capital structure decisions are critical and important decisions for any organization because of influencing the firm's financial risk. Due to the financial risk and financing issues, examinations of capital structure determinants and relationship between capital structure and financial performance have a vital importance for SMEs. This study covered 145,308 Small and Medium Enterprises (SMEs) in Turkey during 1992-2007. Data of firms was obtained from Turkish Central Bank (TCB) after permission. CHAID (Chi-Square Automatic Interaction Detector) decision trees data mining algorithm has been used for data analysis which is one of the best ways to identify financial profiles of firms and relationship between capital structure and financial performance.

[Nermin Ozgulbas, Baskent University Faculty of Health Sciences, Department of Healthcare Management, Eskisehir Yolu 20.km. 06810/Ankara/Turkey; ozgulbas@baskent.edu.tr]

17d3-Mathematical Statistics (III)
December 17 (Saturday), 16:30 - 17:30, HSS Media Conference Room
Chair: Wei-Ying Wu

17d3-1 A Broad Class Estimators for Two Ordered Normal Means with Ordered Variances Under Pitman's Comparison

Chang Yuan-Tsung
Mejiro University, Japan

Shinozaki Nobuo
Keio University, Japan

We propose a broad class of improved estimator for estimating two ordered normal means, individually and/or simultaneously, when variances are ordered under Pitman closeness criteria. We show that in estimating the mean with larger variance, the proposed broad improved estimators which take into account to the order restriction on variances are closer to parameter than usual one which has not taken into account the order restriction on variances under modified Pitman's criterion. However, in estimating the mean with smaller variance similarly result can't be obtained. We also consider the estimation of two ordered means, simultaneously, and show that proposed broad improved estimators improve the usual estimators under Pitman's comparison.

[Chang Yuan-Tsung, 4-31-1Nakaochiai, Shinjuku-ku, Tokyo; cho@mejiro.ac.jp]

17d3-2 Empirical Likelihood Inference about the Mean with Ranked Set Samples

Ayman Baklizi
Department of Mathematics and Physics, Qatar University

We consider empirical likelihood intervals for the population mean based on a ranked set sample. We presented the large sample distribution of the log empirical likelihood ratio statistic for the mean. The corresponding interval estimates of the population mean are obtained. We present Simulations carried out to compare the performance of the empirical likelihood intervals with other known nonparametric intervals.

[Ayman Baklizi, Department of Mathematics and Physics, Qatar University, Doha, Qatar; a.baklizi@qu.edu.qa]

17d3-3 Multidimensional Scaling as Regression Analysis

Takashi Shindo
Okayama University Graduate School, Okayama, Japan

We propose a Multidimensional scaling (MDS) as a regression analysis. Here, it is focused to explain differences between objects, not a trend over objects. In the method, a MDS as a Feature Matching Model by Shindo and Tamagake [in preparation] is applied. Regresants are differences between objects on a scale, and regressors are dummy variables for distinctive features or adjacent levels in an ordinal feature. In addition, Partial Order Scalogram Analysis by Lingoes [Mathesis Press. (1973)] is modified and used to construct an ordinal feature from several features. Objects are arranged as grid points in a hyper-rectangle whose side corresponds to a feature. A dissimilarity is expressed by a Minkowski distance in the hyper-rectangle. If we employ a Minkowski distance except the city-block distance, the method will be a non-linear regression analysis. The regression analyses are efficient when more contributive feature is more dominant.

[Takashi Shindo, 700-8530 1-1-1 Tsushima-naka, Kita-ku, Okayama, Japan; fuyousianko@gmail.com]

17d3-4 **Distributions of Resampling Moments**

Yoko Ono
Yokohama City University, Yokohama, Japan

Naoto Niki
Tokyo University of Science, Tokyo, Japan

Our attention is focused upon a class of resampling methods, including Efron's bootstrapping, Rubin's Bayesian bootstrapping and the ***m***-out-of-*n* bootstrapping, where we randomly choose a huge number of distributions with the same support composed of the given sample values from an unknown population in order to numerically estimate the sampling distribution of a statistic. We first give a unifying formulation to the class then we simplify the situation into the case of known population, namely, the standard normality. The sampling moments of resampling moments are given in an exact way of symbolic representation. Examples for some simple statistics are also shown to illustrate the meanings of our results in resampling estimation of the distributions of statistics.

[Yoko Ono, Seto, Kanazawa-ku, Yokohama, 236-0027 Japan; onoyk@yokohamacu.ac.jp]

17d4-Clinical Trials (II)
December 17 (Saturday), 16:30 - 17:30, AC 1st Conference Room
Chair: Chia-Hui Huang

17d4-1 **Statistical Investigation of Bioequivalence based on Comprehensive Nested Hypotheses**

Yasunobu Furukawa
Kyowa Hakko Kirin Co., Ltd., Tokyo, Japan

Masashi Goto
Biostatistical Research Association, NPO, Osaka, Japan

Generic drugs and drugs adopted for the use of alternative routes of administration can be approved by an only bioequivalence (BE) evaluation for human subject. In Japan, following Health, Labour and Welfare Ministry (1997), We evaluate Bioequivalence based on confidence interval for the difference of the population means between test drug and reference drug. On the other hand, the above evaluation process includes various statistical problems. So it is an urgent need to investigate characteristic of BE systematically in practical clinical evaluation. We focused on BE-evaluation on the change-over design and presented a process of statistical inference assuming the power-normal distribution [Res. Rep. NO.93, Res. Instit. Fund. Infor. Sc., Kyusyu University (1979):6?0; Rep. Stat. Appl. Res., JUSE (1983): 30: 8-28] as the underlying distribution of observations, where the power-normal distribution is defined as the distribution specified before the power-normal transformation.

[Yasunobu Furukawa, 1-6-1, Ohtemachi, Chiyoda-ku, Tokyo, 100-8185, Japan; yasunobu.furukawa@kyowa-kirin.co.jp]

[Masashi Goto, 2-22-10-A411, Kamishinden, Toyonaka-shi, Osaka, 560-0085, Japan; gotoo@bra.or.jp]

17d4-2 A Discussion on Environmental Risks for Breast and Liver Cancer by Analyzing (Age, Period)-Tabulated Data

Nobutane Hanayama
Shobi University, Japan

Several experiment and epidemiological studies have suggested that estrogen, which is known to raise breast cancer risk by encouraging the growth of breast tissue, might play an inhibitory in the development of hepatocellular carcinoma. Actually a negative correlation is seen between breast and liver cancer death rates for Japanese women, and it can be considered due to an increase in estrogen and other hormonally agents in the environment. In this study, for examining the effect of environmental risks on breast and liver cancer, (age, period)-tabulated data on breast and liver cancer deaths are analyzed by fitting an extended age-period-cohort model introduced by Hanayama [Statistics in Medicine 26 (2007): 3459-3475.]

[Nobutane Hanayama, Shobi University, Shimomatsubara 655, Kawagoe 350-1153, Japan; nob-hanayama@jcom.home.ne.jp]

17d4-3 A Spatial Surveillance Approach for Breast Cancer Mortality by Races and Ages in the U.S.

Lung-Chang Chien
Washington University School of Medicine, Department of Internal Medicine, Division of Health Behavior Research, U.S.A.

Hwa-Lung Yu
National Taiwan University, Department of Bioenvironmental Systems Engineering, Taiwan, R.O.C.

Mario Schootman
Washington University School of Medicine, Department of Internal Medicine, Division of Health Behavior Research, U.S.A.

Examining racial and geographic disparities in breast cancer is a key of national public health goal with the concern of data limitations. We proposed a new method for identifying racial and geographic disparities in breast cancer mortality across the U.S using spatial statistics as well as circumventing data reliability concerns. An advanced spatiotemporal approach with Markov random fields and kriging identified locations with elevated mortality rates by race and age at the U.S. county level. For White women, 30.11% counties with significantly elevated rates located in the West North Central U.S for all age groups (adjusted RR = 1.19 1.64). For Black women, 49.05% counties with elevated rates significantly located in the West South Central, Pacific, and East North Central U.S. in different age groups (adjusted RR = 2.71 4.32). This study may provide insight to aid policy determination for preventing the increase of breast cancer mortality in higher-risk counties in priority.

[Lung-Chang Chien, 4444 Forest Park Ave. Suite 6700. St. Louis, MO 63108; lchien@dom.wustl.edu]

17d5-Applications (I)
December 17 (Saturday), 16:30 - 17:30, AC 2nd Conference Room
Chair: Yung-Fu Cheng

17d5-1 Optimal Allocation for the Second Elementary Symmetric Function with Different Coefficients

Chien-Yu Peng
Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.

In this paper we consider the problem of determining the optimal size allocation and optimal

number of experimental conditions for the second elementary symmetric function with different coefficients. We derive analytical solutions for different cases of practical applications and use the general formulation to elucidate the foundation between different parametric models found in recent studies. A geometrical interpretation for the structure of some theoretical results will then be given. This leads our results to enable some complex problems to be more tractable than numerical search algorithms currently allow.

[Chien-Yu Peng, Institute of Statistical Science, Academia Sinica, Taipei, 11529, Taiwan, R.O.C.; chienyu@stat.sinica.edu.tw]

17d5-2 A Study on Determinants of the Demand for Insurance Products

Soo-Kyeong Lee
Kwang-Sik Choi
Korea University, Chungnam, Korea

According to increasing consumers' desire to safety and stability of insurance product, insurance industry has been developed naturally. To improve competitiveness of the insurance industry and build database which is information source for proper decision-making, we find cause and effect with regression analysis using various economic indicators such as unemployment rate and mortality. The regression analysis has endogenous variable that is entire sum of real total premium.

[Soo-Kyeong Lee, Department of Economics and Statistics, Korea University, Jochiwon-eup, Yeongi-gun, Chungnam, Korea, 339-700; maybe8318@korea.ac.kr]

17d5-3 A Simultaneous Test of Differential Item Functioning and Validating Item Construct of Measurements

Chih-Chien Yang
National Taichung University, Taichung, Taiwan, R.O.C.

Ching-Lin Shih
National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C.

Yih-Shan Shih
National Taichung University, Taichung, Taiwan, R.O.C.

The study proposes a new statistical test to simultaneously evaluate Differential Item Functioning and Validate Item Construct (DIFVIC), the two critical measurement quality ensuring steps often seen in educational and psychological studies. The DIFVIC test can effectively improve accuracy

rates of Wang, Shih, and Yang (2009) by properly controlling Type I errors. The DIFVIC test is newly utilized and derived from the original Sobel's (1982) approach and theoretical foundations of Baron and Kenny (1986). To demonstrate effectiveness and feasibilities of DIFVIC test, numerical simulation studies are conducted under various designed experimental conditions, including varied magnitudes of DIF (Shih and Wang, 2009) and VIC (MacKinnon, Warsi and Dwyer, 1995). Tables and figures will be shown specifically. The stable and accurate numerical performances of DIFVIC test prove its empirical and practical importance. In particular, the DIFVIC test gains main advantages in comparing with the popular multiple indicators, multiple causes (MIMIC) method of Wang, Shih, and Yang (2009). The DIFVIC test unites typical functions of DIF and VIC and provides a complete linkage of item response theory and confirmatory factor analysis.

[Chih-Chien Yang, Cognitive NeuroMetrics Laboratory, Graduate Institute of Educational Measurement & Statistics, National Taichung University of Education, 140, Min-Shen Road, TaiChung 40306, Taiwan, R.O.C.; noayang@ntcu.edu.tw]

17d5-4 Computing Asymptotic Relative Efficiency Using Logistic Regression

Tai-Ming Lee
Department of Statistics and Information Science, Fu-Jen University

Asymptotic Relative Efficiency, constructed completely by Pitman(1948), is a useful concept in comparing different testing procedure with the same statistics goal. On the other hand, it is difficult to compute the ARE in general, since the power function should be approximated and inverted to obtain sample size requirement. A famers special case computed by Hodgens and Lehmann(1956) showed that the ARE is at least 0.864 in comparing Wilcoxon sum of rank and two sample t test. On this paper, we propose the MLE Monte Carlo method to estimate the power function by Logistic Regression, with which it is easy to invert the function. This general methodology can compute ARE with any specific population to compare different procedure. Several cases will be shown graphically and numerically.

[Lee, Tai-Ming, Department of Statistics and Information Science, Fu-Jen University; stat1006@mail.fju.edu.tw]

17d6-Applications (II)
December 17 (Saturday), 16:30 - 17:30, AC 3rd Conference Room
Chair: Hsiu-Wen Chen

17d6-1 SPARSIMAX: Principal Component Analysis with Direct Sparsity Constraint on Loadings

Kohei Adachi
Osaka University, Japan

Nickolay T. Trendafilov
The Open University, United Kingdom

Sparse principal component analysis (PCA) procedures have been proposed in which penalty functions are used for obtaining sparse loading matrices with a certain number of elements being zeros [Jolliffe, Trendafilov, and Uddin (2003) J. Comp. Graph. Stat, 12, 531-547; Zou, Hastie, and Tibshirani (2006) J. Comp. Graph. Stat, 15, 265-286]. In this paper, we propose a new sparse PCA, Sparsimax, which does not involve penalty functions. Instead, the sparsity of loadings is directly constrained in Sparsimax, that is, the PCA loss function is minimized subject to the constraint that a specified number of loadings are zero, though which loadings are zero is not constrained. We develop an alternating least squares algorithm for Sparsimax. Further, we present a procedure for finding an appropriate number of zero loadings with information criteria. Sparsimax is assessed in a simulation study and illustrated with a real data example.

[Kohei Adachi, Osaka University, Japan; k-adachi@lt.ritsumei.ac.jp]

17d6-2 **A Visualization of Aggregated Symbolic data**

Yoshikazu Yamamoto
Tokushima Bunri University, Sanuki, Japan

Junji Nakano
The Institute of Statistical Mathematics, Tachikawa, Japan

Takeshi Fujiwara
Tokyo University of Information Sciences, Chiba, Japan

A large amount of multivariate data is often divided into some groups according to values of categorical variables. Then, we are mainly interested in the information of the groups, not in the original individual data. Such groups are thought to be aggregated symbolic data, in which variables can take complicated values such as intervals or histograms. In the traditional symbolic data analysis, a variable expresses information of the marginal distribution of individuals in the group. We propose to use information of the joint distribution of variables in the group. In this paper, we consider to use correlation coefficients among variables together with averages and standard deviations of variables, and focus on the visualization of them. For this purpose, we propose to use an extended parallel coordinates plot. We also show the usefulness of interactive operations in the graphics to grasp the characteristics of aggregated symbolic data.

[Yoshikazu YAMAMOTO, 1314-1 Shido, Sanuki, Kagawa 769-2193, Japan; yamamoto@fe.bunriu.ac.jp]

[Junji NAKANO, 10-3 Midoricho, Tachikawa, Tokyo 190-8562, Japan; nakanoj@ism.ac.jp]

[Takeshi Fujiwara, 4-1 Onaridai,Wakaba-ku, Chiba 265-8501, Japan; fujiwara@rsch.tuis.ac.jp]

17d6-3 Identification of Extreme Climate Cycles and Trends from Weather Time Series for Renewable Energy Applications

Shin-Guang Chen
Tungnan University, New Taipei City, Taiwan, R.O.C.

Traditionally, the time series profile of a typical day or year is used to size the renewable energy application systems. However, facing the global warming crisis as well as the fact that no two years for a single site would have the same weather condition, this often makes the traditional way of sizing failed in the extreme weather conditions. This article presents a method to statistically model the trend of climate cycles year by year not only for the solar radiation time series but also for the wind speed time series. At first, we will change the solar radiation data to the time domain equivalence (the Peak Solar Hour) and the wind speed data to the time domain equivalence (the Peak Wind Hour) for our approach. Then, the climate cycles and trends for both time series will be obtained via the Generalized Extreme Value Distribution. Some numerical examples are illustrated to show the effectiveness of our approach.

[Shin-Guang Chen, 152, Sec.3, Bei-Shen Rd., Shen-Ken Dist. New Taipei City, Taiwan, R.O.C.; bobchen@mail.tnu.edu.tw]

17d6-4 Data Driven Modeling of Vertical Atmospheric Radiation

Jaromir Antoch
Charles University in Prague

Vertical profiles of radioactivity in the atmosphere are measured on several places of the world since long time, for details see Bazilevskaya and and Svirzhevskaya (1998), CHMI-RADAC, Hatakka et al. (2000), Li et al. (2007), and references given there. Among the reasons behind these measurements belong study of effects of radioactivity on human beings and on the environment, especially during the nuclear incidents. Aside that, obtained data form an important component of numerical climate and weather prediction models.
Apart from improving general knowledge about the distribution of the radioactivity in the atmosphere, there is another serious reason behind these measurement, namely, to collect enough

of information for creation of the warning system combining both atmospheric and ground measurements. Urgent need of such a system, unfortunately never fully achieved, dates back to the Chernobyl disaster and was thoroughly discussed after every larger nuclear incident. It should be recalled that soon after the Fukushima accident CHMI launched control sondes to monitor the situation and thanks to the fact that obtained measurements were not substantially different from the historical records, tried to calm public opinion confused by contradictory information from the media. Of special interest are the observations of vertical beta and gamma radiation levels, so called beta (gamma) counts. The matter at hand is the average number of beta and gamma counts per second in consecutive ten-second-long intervals. Unfortunately, a physical model for the vertical radioactivity profiles has never been published. Therefore, our goal was to suggest a stochastic data-driven model(s) describing dependence of the radiation on the altitude which are based on the analysis of measured intensities of the radiation and the altitude.
After careful study of the data we decided to use modified growth curves and Poisson process approach. The modification consisted of combining the basic growth curve with either a linear function or some other growth curve with the aim to improve the fit in problematic parts of the analyzed profile, being typical especially for the surface radiation. Growth curves were used in two different ways: (1) As a basic nonlinear regression model describing dependence of the mean amount of the beta or gamma counts on the altitude. (2) As a basic model describing dependence of the intensity of underlying nonstationary Poisson process on the altitude.
In our contribution a stochastic data-driven model based on nonlinear regression and on nonhomogeneous Poisson process will be discussed. The primary goal is to improve understanding of the distribution of environmental radiation as obtained from the measurements of the vertical radioactivity profiles by the radioactivity sonde systems. First, growth curves will be used to establish an appropriate nonlinear regression model. Second, for comparison a nonhomogeneous Poisson process with the intensity based on growth curves will be considered. Computational aspects will be briefly discussed. Suggested methods were applied to the real data and compared. Obtained results will be concisely commented. Keywords: Richards'growth curve, nonlinear regression, nonhomogeneous Poisson process, vertical atmospheric radiation profiles, RADAC data.

[Jaromir Antoch, Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Czech Republic; antoch@karlin.mff.cuni.cz]

17d7-Clinical Trials (III)
December 17 (Saturday), 16:30 - 17:30, AC 4th Conference Room
Chair: May Wang

17d7-1 Sample Size Determination for Clinical Trials with Two Correlated Time-to-Event Co-primary Endpoints

Toshimitsu Hamasaki
Osaka University Graduate School of Medicine, Japan

Scott Evans
Harvard University School of Public Health, U.S.A.

Tomoyuki Sugimoto
Graduate School of Mathematical Sciences, Hirosaki University, Japan

Takashi Sozu
Kyoto University School of Public Health, Japan

The effects of interventions are multi-dimensional (e.g., benefits and harms). Co-primary end-points offer an attractive design feature in clinical trials as they capture a more complete characterization of the effect of an intervention. For example in cancer trials, overall survival is often of primary interest, but relapse-free or progression-free survival is also important. Trials of co-morbidities may also utilize co-primary endpoints, e.g., a trial evaluating therapies to treat Kaposis sarcoma (KS) in HIV-infected individuals may have: (1) the time to KS progression and (2) the time to HIV virologic failure, as co-primary endpoints. We present methods for sample size determination for clinical trials comparing two interventions with respect to two time-to-event co-primary endpoints. The methods are practical and simple, and assume that the time-to-event outcomes are exponentially distributed. Using data generated from bivariate exponential distributions, we describe the empirical power profile for the log-rank test under a computed sample size given by the method.

[Toshimitsu Hamasaki, 2-2 Yamadaoka, Suita, Osaka; hamasakt@medstat.med.osaka- u.ac.jp]

17d7-2 Properties and Effects of Inferences on Hierarchical Generalized Linear Models in Clinical Trial Designs

Kentaro Kuroishi
Astellas pharm Inc, Tokyo, Japan

Yoshimichi Ochi
Oita University, Oita, Japan

In the field of sociology, psychology, and clinical trials, data are collected with hierarchy structures. For examples, individual measurements in a clinical trial are often made at many sites involved in the trial. In addition, those sites in the global clinical study belong to the target regions or countries. This research deals with the application of the hierarchical model to binary responses obtained from such a clinical trial in order to take account of the hierarchical data structure. Inference of a hierarchical model on the basis of generalized linear models can be carried out via the maximum-likelihood method (including EM algorithm), REML and Bayesian approach. Variances of the parameters can also be estimated using information matrix, sandwich estimator and variance

evaluation of posterior distribution. We explore properties and effects of those methods with respect to the design of clinical trials.

[Kentaro Kuroishi, 17-1, Hasune 3-chome, Itabashi-ku, Tokyo 174-8612, Japan; kentarou.kuroishi@jp.astellas.com]

17d7-3 Design and Data Monitoring of Clinical Trials with Co-primary Benefit:Risk Endpoints Using Prediction

Scott R. Evans
Harvard University, Boston, MA U.S.A.

Toshimitsu Hamasaki
Kenichi Hayashi
Osaka University Graduate School of Medicine, Osaka, Japan

The evaluation of both the benefits and the risks of interventions is a fundamental goal in the monitoring and analyses of clinical trials. For example, Data Monitoring Committees that monitor trials evaluate benefits and risks to make recommendations regarding future trial conduct. Although many methods for interim monitoring exist, few of these methods evaluate the joint magnitude of the benefits and risks. Furthermore none use prediction to convey information regarding potential effect size estimates and associated precision, with trial continuation. Building upon the work of Evans and colleagues [Evans SR, Li L, Wei LJ, Drug Inf. J., (2007): 41:733-742; Li L, Evans SR, Uno H, Wei LJ, Stat. Biopharm. Res., (2009):1:4:348-355], we propose use of prediction and ”predicted rings” as a flexible and practical strategy for monitoring trials with co-primary benefit:risk endpoints. These methods will provide a valuable tool for Data Monitoring Committees and other decision-makers when evaluating interim data.

[Scott Evans, Harvard School of Public Health, FXB-513 651 Huntington Ave., Boston, MA 02115 U.S.A.; evans@sdac.harvard.edu]

17d7-4 Graphical Approaches for Seamless Phase II/III Designs with Multiple Doses Involving Multiple Families of Endpoints

Toshifumi Sugitani
Chikuma Hamada
Tokyo University of Science, Tokyo, Japan

Adaptive combination tests [Biometrical Journal, 48 (2006): 623-634] are commonly used in adaptive seamless phase II/III designs. They combine the technique of closure principle [Biometrika,

63 (1976): 655-660] with combination tests [Biometrics, 51 (1994): 1315-1324]. However, adaptive combination tests have a crucial disadvantage, that is, they are not able to incorporate the hierarchy of multiple families of hypotheses into their decision process. To overcome such a problem, we introduce a hybrid method between graphical approaches, which are proposed by Bretz et al. [Statistics in Medicine, 28 (2009): 586-604], and combination tests. The proposed method can take into account the order of families of hypotheses while overcoming the problem of non-consonance related to combination tests [JASA, 105 (2010): 660-669]. The validity of the method is guaranteed in terms of partitioning principle [The Annals of Statistics, 30 (2002): 1194-1213].

[Toshifumi Sugitani, Tokyo University of Science, Tokyo, Japan; sugitani@ms.kagu.tus.ac.jp]

7th IASC-ARS Σ joint 2011 Taipei Symposium

Institute of Statistical Science, Academia Sinica

Asian Regional Section of the IASC

DGBAS, Executive Yuan, R.O.C.

National Science Council, R.O.C.

The Chinese Institute of Probability and Statistics

中國統計學社 Chinese Statistical Association