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

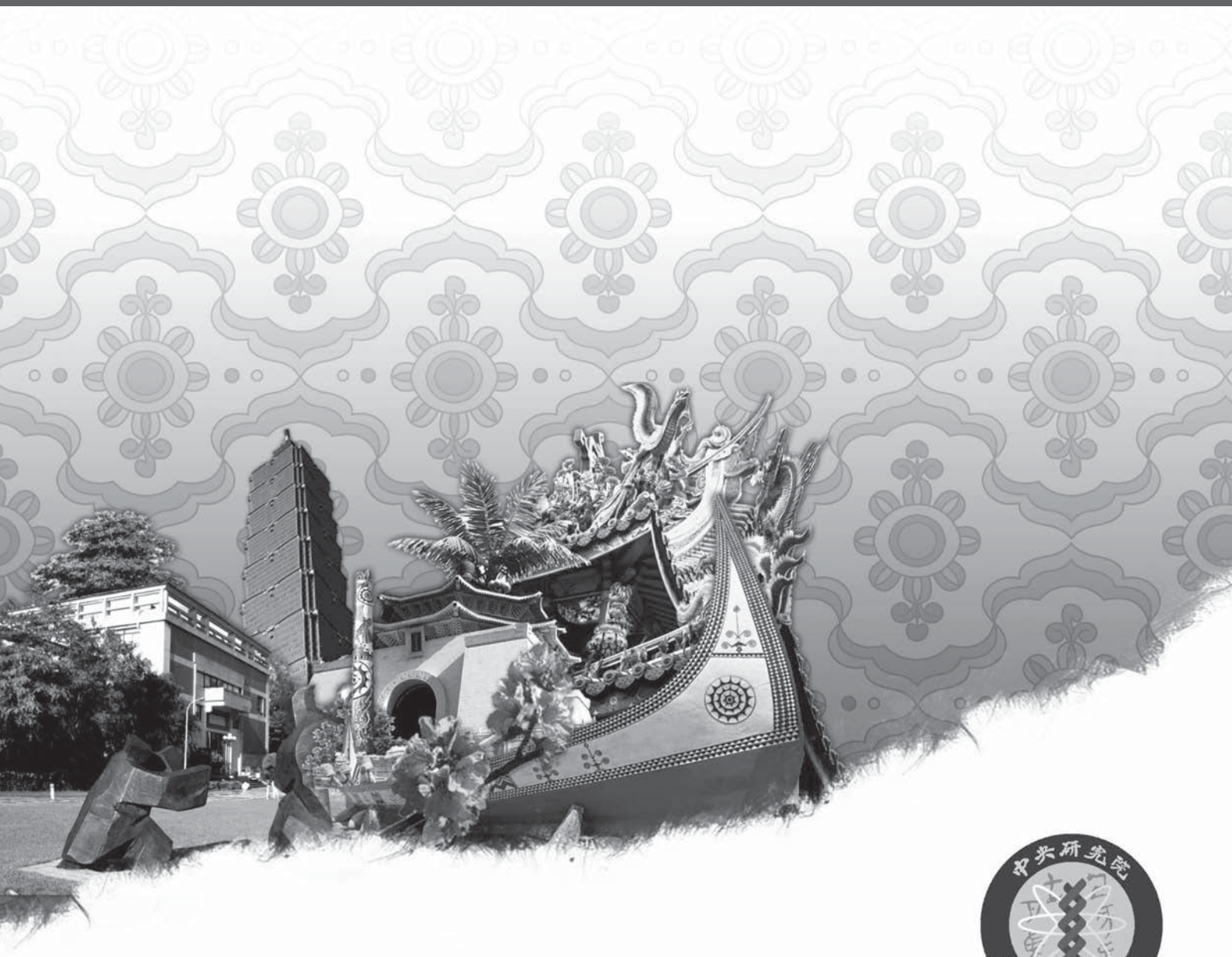
December 16 - 19, 2011

Academia Sinica, Taipei, Taiwan

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

中央研究院 ACADEMIA SINICA 1928

Abstract



December 18 (Sunday)

Today's Highlights:

08:30 – 10:10	***The Future Research and Education in Statistics***
10:30 – 12:00	***Parallel Sessions 18a1 – 18a7***
13:00 – 14:30	***Parallel Sessions 18b1 – 18b7***
14:40 – 16:10	***Parallel Sessions 18c1 – 18c7***
16:30 – 18:00	***Parallel Sessions 18d1 – 18d6***
18:20 – 19:20	***Concert***
19:30 –	***Banquet & Dinner Speech by Jae Chang Lee***

Abstract Pages:

The Future Research and Education in Statistics
December 18 (Sunday), 8:30 - 10:10, HSS Center International Conference Hall
Organizer: Jing-Shiang Hwang
Chair: George Tiao

Statisticians at Work: Inspiration, Aspiration, Ambition

C F Jeff Wu
Georgia Institute of Technology, U.S.A.

A key measure of the maturity and quality of a scientific community is how it judges and values accomplishments and (or versus) scholarship. To address this question, I will describe the motivation or drive for accomplishments and/or scholarship at three levels: inspiration, aspiration, ambition. I will use some prominent examples in statistics history to explain or illustrate the acts of in- spiration, aspiration, and ambition. Then I will share some thoughts on what is good or bad mathematical statistics work. Finally, I will give some advice to young researchers. Throughout this talk, I will use the ”lens” of inspiration, aspiration, and ambition in making my examinations, remarks and suggestions.

[C F Jeff Wu, Georgia Institute of Technology, USA; jeff.wu@isye.gatech.edu]

Statistical Education and Educating Statisticians: Producing Wine Connoisseurs and Master Winemakers

Xiao-Li Meng
Department of Statistics, Harvard University, U.S.A.

In recent years we have taken a broader view of statistical education for Harvard’s undergraduates, by shifting the focus from preparing a few to pursue Ph.D. level quantitative studies to helping many gain a basic appreciation of statistical argument and insight. Intriguingly, the journey, guided by the philosophy that one can become a wine connoisseur without knowing how to make wine, apparently has led us to produce many more future winemakers than when we focused only on producing a vintage. At the Ph.D. level, our focus has always been to produce the best winemakers, but true expert winemakers need to master far more than merely the chemical process of fermenting juice into alcohol. We therefore introduced a Professional Development Curriculum (PDC) parallel to the usual course curriculum, starting from ”Stat 303: The Art and Practice of Teaching Statistics.” This talk shares a number of stories from our intoxicating journey. [Related articles are in Publications on Pedagogy, Professional Development, and Profession Building and Outreach (articles 0.5-0.10) at http://www.stat.harvard.edu /Faculty Content/Meng-cv.pdf or by emailing meng@stat.harvard.edu]

[Xiao-Li Meng, Department of Statistics, Harvard University, USA; meng@stat.harvard.edu]

Graduation Education in Biostatistics: My Personal Viewpoint

Kung-Yee Liang
Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taiwan, R.O.C.

In this talk, I will discuss what I consider as the essences of graduate education in biostatistics. Topics considered include major components in graduate program such as core courses, integration with biomedical sciences, qualification for being admitted to and successful for completing the program, and future challenges, etc. This discussion is primarily drawn from my past experience as the Graduate Program Director (1996-2003) at the Department of Biostatistics, Johns Hopkins University.

[Kung-Yee Liang, Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taiwan, R.O.C.; president@ym.edu.tw]

18a1-Bayesian Modeling and Computation
December 18 (Sunday), 10:30 - 12:00, HSS 1st Conference Room
Organizer: Dedi Rosadi
Chair: Dedi Rosadi and Tsair-Chuan Lin

18a1-1 On The Modeling of District Policy Effects to Household Expenditure: A Hierarchical Bayesian Approach

Nur Iriawan
Pudji Ismartini
Statistics Department, 'Institut Teknologi Sepuluh Nopember', Indonesia

Household data such as household expenditure data are widely used as a basis for policy decision making. Since Household data is often view as hierarchical structured data, with household nested in its district of residence area, both the district condition and its households are influence and are influenced by each other. In this case, it is important to have contextual analysis to model the significance of district policy impacts to household expenditure. This paper proposes to develop a model for analyzing district policy effects to household expenditure by using a hierarchical Bayesian approach. The model is developed by taking into account household characteristics and district characteristics as a reflection of government policy at district level. The result shows that district policy in the economic, health, education and housing facilities does effect to household expenditure through its household characteristics.

[Nur Iriawan, Statistics Department, 'Institut Teknologi Sepuluh Nopember', Indonesia.; nur i@statistika.its.ac.id]

18a1-2 Calibration of Artifical Ensemble Forcast Using Bayesian Model Averaging

Heri Kuswanto
Department of Statistics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia

This paper proposes an alternative of generating probabilistic forecasts by using combination of time series models (e.g. ARIMA models), instead of using numerical weather prediction (NWP) ensemble. The generated forecasts hereafter denoted as artificial ensemble forecasts. The idea combining the models is to capture the uncertainty induced by the model choice. Two weather variables are examined namely temperature and wind speed. Simply combining the forecasts leads to unreliable forecasts and hence it should be calibrated. Applying the standard calibration method i.e. Bayesian Model Averaging (BMA) leads to more reliable forecasts than the forecasts produced by deterministic forecasts.

[Heri Kuswanto, Department of Statistics, Kampus ITS Sukolilo Surabaya; heri k@statistika.its.ac.id]

18a2-Financial Time Series Analysis
December 18 (Sunday), 10:30 - 12:00 , HSS 2nd Conference Room
Organizer: Cheng-Der Fuh
Chair: Cheng-Der Fuh

18a2-1 Clustering of Multiple Time Series with Applications

Ruey S. Tsay
Booth School of Business, University of Chicago

This talk considers clustering of many time series. We propose a new robust model-based method to classify time series into multiple clusters and compare performance of the proposed method with that of other methods available in the literature, including traditional distance-based procedures such as *k*-means. Advantages and applicability of the proposed method are shown. For applications, we consider the prediction of the monthly unemployment rates of the 50 states in the U.S. and the GDP growth rates of many countries. (Joint work with Hannes Ledolter, Keshab Shrestha, and Yongning Wang.)

[Ruey S. Tsay, Booth School of Business, University of Chicago; ruey.tsay@chicagobooth.edu]

18a2-2 Minimum Variance Unbiased Estimator of Integrated Volatility in Stochastic Volatility Models

Meihui Guo
National Sun Yat-sen University, Kaohsiung, Taiwan, R.O.C.

Estimation of the integrated volatility is an important problem in high frequency financial data analysis. In this study, we propose a minimum variance unbiased estimator of integrated volatility for stochastic volatility models with microstructure noise. The proposed estimator is based on a signal-to-noise ratio statistics and its asymptotic normality is established with convergence rate $Op(n^{-1/4})$, which attains the same efficiency as the maximum likelihood estimator in the constant volatility case. An unbiased estimator of the microstructure noise variance is also proposed to facilitate the estimation procedure. Simulation studies are performed to confirm the theoretical results and to compare the proposed estimator with the other estimators proposed in the literatures.

[Liang-ching Lin, National Sun Yat-sen University, Kaohsiung, Taiwan, R.O.C.; linlc@mail.math.nsysu.edu.tw]

18a2-3 Bayesian Inference of Asymmetric Smooth Time-varying Structure in Financial Time Series

Cathy W. S. Chen
Jane-Lin Wu
Department of Statistics and Actuarial Science Feng Chia University, Taiwan, R.O.C.

Motivated by time-varying (TV) GARCH, we consider the asymmetric logistic functions mechanism and incorporate it into the conditional volatility equation for the financial time series. This class of models allows smooth transitions to different break points. We focus on the Bayesian inference of asymmetric TV-GARCH via the Markov chain Monte Carlo (MCMC) methods to obtain parameter estimates. The proposed methodology is illustrated using both simulated and international stock market return series. Marginal likelihood is employed to choose the break number in the TV-GARCH model. We find significant improvements of fitness in asymmetric smooth transition TV-GARCH models.

[Cathy W. S. Chen, Department of Statistics and Actuarial Science Feng Chia University, Taiwan, R.O.C.; Chenws@mail.fcu.edu.tw]

18a3-Robustness and Diagnostics Toward Data Analysis
December 18 (Sunday), 10:30 - 12:00, HSS Media Conference Room
Organizer: Tsung-Chi Cheng
Chair: Hsun-chih Kuo

18a3-1 Very Robust Regression

Anthony C. Atkinson
Department of Statistics, London School of Economics, United Kingdom

Procedures for Very Robust Regression are intended to be unaffected when almost half of the observations are outliers, whatever the values of the p explanatory variables. All algorithms divide the observations into a "good" group which are used in fitting and a "bad" or outlying group which are downweighted. In Least Median of Squares and Least Trimmed Squares the parameters are estimated from many subsamples of p observations. In the "reweighted" versions of these algorithms, a second stage involves fitting to all those observations not rejected as outlying after the first stage. Consequently, subsets of at most two sizes are used. On the contrary, the Forward Search uses a series of subsets of increasing size for fitting the model and for determining which observations should be excluded from the fit. The talk will show the superior properties of this more general approach; the difference is particularly acute when the data are split between two models which are close together. Such data occur in investigations of fraud in international trade. The Forward Search can also be used in building regression models. If time permits, we will describe distributional results used to calibrate a robust version of C_p for model choice and relate it to Akaike's AIC. The procedures to be described are fully automatic. However, insight into the methods is gained through the use of plots. The talk will be illustrated with diagnostic graphics from a series of data analyses.

[Anthony C. Atkinson, Department of Statistics, London School of Economics, United Kingdom; a.c.atkinson@lse.ac.uk]

18a3-2 Diagnostics for the Proportional Odds Regression Model Using Stochastic Processes

Ivy Liu
School of Mathematics, Statistics and Operations Research Victoria University of Wellington, Wellington, New Zealand

The cumulative logit or the proportional odds regression model is commonly used to study covariate effects on ordinal responses. This talk provides some graphical and numerical methods for checking the adequacy of the proportional odds regression model using the cumulative residuals. The methods focus on evaluating functional misspecification for specific covariate effects. An example of a data set from the Normative Aging Study, which studies the effect of two markers of oxidative stress in men in the age group of 48-V94 years: white blood cell count and C-reactive protein on FBG measurement (measured in three clinically defined ordinal categories), is given. This talk also provides a goodness-of-fit method to check the overall adequacy of the proportional odds regression model using stochastic processes.

[Ivy Liu, School of Mathematics, Statistics and Operations Research Victoria University of Wellington Wellington, New Zealand; i-ming.liu@vuw.ac.nz]

18a3-3 Deterministic Algorithms for MCD and LTS

Tim Verdonck
Mia Hubert
Peter Rousseeuw
Department of Mathematics, K.U.Leuven, Celestijnenlaan 200b, Leuven, Belgium

The minimum covariance determinant (MCD) method is a robust estimator of multivariate location and scatter, whereas the least trimmed squares (LTS) is a highly robust regression estimator (Rousseeuw, 1984). Computing the exact MCD and LTS is very hard, so in practice one resorts to approximate algorithms. Most often the FASTMCD and FASTLTS algorithms of Rousseeuw and Van Driessen (1999) are used. The FASTMCD algorithm is affine equivariant, whereas the LTS estimates are regression, scale, and affine equivariant. These algorithms start by drawing many random subsets, followed by so-called concentration steps. Consequently both algorithms are not deterministic, hence not permutation invariant. We present deterministic algorithms for both MCD as LTS, which don't use random subsets and peform much faster. Both algo- rithms will be illustrated on real and simulated data sets.

[Tim Verdonck, Department of Mathematics, K.U.Leuven, Celestijnenlaan 200b, Leuven, Belgium; tim.verdonck@wis.kuleuven.be]

18a4-Statistics in Systems Biology
December 18 (Sunday), 10:30 - 12:00, AC 1st Conference Room
Organizer: Satoshi Kuriki
Chair: Vivian Yi-Ju Chen

18a4-1 Likelihood Ratio Test for Exploratory Factor Analysis Model

Yoshiyuki Ninomiya
Kyushu University, Japan

To select the number of fators in the fator analysis, we consider the following testing proedure: (i) Testing m against $m + 1$ for the number of fators by thelikelihood ratio test statistic is conducted from $m = 1$ as long as the last testing is rejected (ii) When the testing is not rejected, the number of fators in the null hypothesis is selected as the true number. But this proedure needed a Monte Carlo simulation to evaluate its p-value, because the likelihood ratio test statistic does not follow the chi-square distribution asymptotially. To avoid the simulation, we generalize the locally conic parameterization introduced by Dacunha-Castelle and Gassiat [*Ann. Statist.* 27(1999):1178-1209]

to evaluate the asymptotic distribution of the likelihood ratio test statistic.

[Yoshiyuki Ninomiya, Kyushu University, Japan; nino@math.kyushu-u.ac.jp]

18a4-2 Likelihood Ratio Tests for Positivity in Polynomial Regressions

Naohiro Kato
The Graduate University for Advanced Studies, Tokyo, Japan

Satoshi Kuriki
The Institute of Statistical Mathematics, Tokyo, Japan

A polynomial that is nonnegative over a given interval is called a positive polynomial. The set of such positive polynomials forms a closed convex cone K. In this paper, we consider the likelihood ratio test for the hypothesis of positivity that the estimand polynomial regression is a positive polynomial. By considering hierarchical hypotheses including the hypothesis of positivity, we define nested likelihood ratio tests, and derive their null distributions as mixtures of chi-square distributions by using the volume-of-tube method. The mixing probabilities are obtained by utilizing the parameterizations for the cone K and its dual provided in the framework of the Tchebycheff systems when the degree of polynomials is up to 4. Moreover, we propose the associated simultaneous confidence bound for polynomial regression curves. Regarding computation, we demonstrate that symmetric cone programming is useful to obtain the test statistics.

[Naohiro Kato, Department of Statistical Science, The Graduate University for Advanced Studies, 10-3 Midoricho, Tachikawa, Tokyo 190-8562, Japan.; nkato@ism.ac.jp]

18a4-3 Ordinal Data Analysis Based on Kullback-Leibler Distance

Wei Gao
Ning-Zhong Shi
School of Math and Statistics, Northeast Normal University, People's Republic of China

Ordinal data often appear in clinic studies, econometrics and social science data analysis. It can improve statistical inference if "ordinal" information properly incorporate into data analysis. Base on the Kullback-Liebler information, we will propose statistics for ordinal data and present unified generalized iterative scaling (UGIS) to estimate the related parameters, which is a powerful method to deal with ordinal data analysis.

[Wei Gao, School of Math and Statistics, Northeast Normal University, Changchun, Jilin 130024, China; tgaow@nenu.edu.cn]

18a5-Distribution Theory
December 18 (Sunday), 10:30 - 12:00, AC 2nd Conference Room
Organizer: Gwo Dong Lin
Chair: J. S. Huang

18a5-1 Cantor Order Statistics: Without Applications

Barry C. Arnold
Statistics Department, University of California, Riverside, CA 92521, U.S.A.

Building on ideas and concepts introduced by Lad, Taylor and Hosking, a generalized Cantor distribution and a corresponding skew generalized Cantor distribution are described. Associated inverse distributions are also introduced. Some results concerning the corresponding order statistics will be discussed.

[Barry C. Arnold, Statistics Department, University of California, Riverside, CA 92521, U.S.A.; barnold@ucr.edu]

18a5-2 Issues Concerning Constructions of Discrete Life Distributions

Chin-Diew Lai
Massey University, Palmerston North, New Zealand

Although failure data is usually treated as being continuous, it may have been collected in a discrete manner, or in fact be discrete in nature. The question arises as how we define a discrete hazard rate. The classical definition of a discrete hazard rate is found to be wanting as the simple relationship between the cumulative hazard function and the survival function does not hold. An alternative definition of discrete hazard rate has been proposed in the literature and we wish to study further relationships between the two definitions.
It is well known the hazard rate characterizes a lifetime distribution. With this in mind, we wish to explore the issue on how to construct a discrete life distribution from its continuous counter part.
In this talk, we present three possible ways of constructing a discrete life dis- tributions. We explore the differences and similarities between the three constructed discrete distributions.

[C D Lai, Statistics & Bioinformatics, IFS, Massey University, Palmerston North, New Zealand; c.lai@massey.ac.nz]

18a5-3 Multivariate Inverse Gaussian Distributions and Skew - Normal Laws

Harry Joe
Vanamamalai Seshadri

Barry Arnold
University of British Columbia, McGill University and University of California Riverside

Over the years , multivariate extensions of commonly used univariate parametric families have been constructed by researchers either by the use of copulas or other means. But there is no multivariate inverse Gaussian to be found in the literature which has univariate margins just as in the Normal case or multinomial case. In this paper we obtain a very natural extension of the inverse Gaussian law without resorting to the use of copulas. This is made possible by the use of a skew-Normal distribution. In this instance we first study a tranformation of the inverse Gaussian law that results in a skew-Normal distribution . A careful analysis of this skew normal law yields the tools for extension to a multivariate skew Normal distribution. This multivariate skew- Normal distribution is different from that of Azzalini and Dalla Valle (1986). Our multivariate skew-Normal distribution is closed under margins and has range of skewness that is algebraically independent of the dependence parameters, a property that does not hold for the previosly defined competitors.

[Vanamamalai Seshadri, University of British Columbia, McGill University and University of California Riverside; van-sesh@hotmail.com]

18a5-4 Some Properties of a Bivariate Distribution Based on Order Statistics

J. S. Huang
University of Guelph, Canada

G. D. Lin
Academia Sinica, Taiwan

Let $(X, Y) \sim H$, $X \sim F$, $Y \sim G$, and let $X_{k,n}$ and $Y_{k,n}$ be the order statistics from distributions F and G, respectively. To construct a bivarite distribution with given marginals F and G, Bairamov and Bayramoglu (2011, JMVA) proposed $K_{+}^{(n)}(x,y) = \frac{1}{n}\sum_{k=1}^{n} \Pr(X_{k,n} \le x, Y_{k,n} \le y)$, and claimed that if H is positive quadrant dependent (PQD), $K_{+}^{(n)}$ attains a correlation higher than that of a distribution previously proposed by Baker (2008, JMVA), $^{(n)}_{+}(\quad) = \frac{}{n}\sum^{n}_{=} \quad (\quad_{n} \le \quad)\ (\quad_{n} \le \quad)$. Little more is known regarding this new distribution and the underlying joint distribution of order statistics $(X_{r,n}, Y_{s,n})$ when the population distribution $H(x,y) = \Pr(X \le x, Y \le y)$ is not necessarily independent. We prove that (i) for $n = 3$, $\Pr(X_{r,n} \le x, Y_{s,n} \le y)$ is PQD if H is, (ii) for $n = 3$, $\Pr(X_{r,n} \le x, Y_{s,n} \le y)$ is negative quadrant dependent (NQD) if H is, and (iii) for $n = 2, 3, \ldots, 12$, $K_{+}^{(n)}$ is PQD regardless of H. We also present some comparison of $K_{+}^{(n)}$ and $^{(n)}_{+}$.

[J. S. Huang, Department of Mathematics and Statistics, University of Guelph, Ontario, Canada N1G 2W1; jshuang@uoguelph.ca]

18a6-Cure Models in Event History Analysis
December 18 (Sunday), 10:30 - 12:00, AC 3rd Conference Room
Organizer: Yi-Hau Chen
Chair: Wei-Yann Tsai

18a6-1 Estimation and Inference for Accelerated Failure Time Model with a Cure Fraction

Wenbin Lu
Department of Statistics, North Carolina State University, U.S.A.

We study the accelerated failure time model with a cure fraction via kernel-based nonparametric maximum likelihood estimation. An EM algorithm is developed to calculate the estimates for both the regression parameters and the unknown error density, in which a kernel-smoothed conditional profile likelihood is maximized in the M-step. We show that with a proper choice of the kernel bandwidth parameter, the resulting estimates are consistent and asymptotically normal. The asymptotic covariance matrix can be consistently estimated by inverting the empirical Fisher information matrix obtained from the profile likelihood using the EM algorithm. Numerical examples are used to illustrate the finite-sample performance of the proposed estimates.

[Wenbin Lu, Department of Statistics, North Carolina State University, U.S.A.; lu@stat.ncsu.edu]

18a6-2 Nonparametric Method of Mixture Model with Prevalent Sampling

Yu-Jen Cheng
Institute of Statistics, National Tsing Hua University

The objective of this article is to make inference on the survival function and the cure probability subject to left truncation. The problem is especially complex because death and cure are contrast events and both events could be truncated before data recruitment. Mixture model is considered in this article. We addressed the connection between mixture model and competing risk model under prevalent sampling scheme and developed a nonparametric approach to estimate the survival function based on a weaker assumption, conditional independence. Nonparametric MLE of the survival function and the probability of cure are derived in this article. We also show that the model under conditional independence assumption is non-identiable subject to right censoring. Our methodology was motivated by and applied to the intensive care unit (ICU) study in Israel.

[Yu-Jen Cheng, Institute of Statistics, National Tsing Hua University; ycheng@stat.nthu.edu.tw]

18a6-3 The Association Estimation for Multivariate Survival Data with a Cure Fraction

Chyong-Mei Chen
Tai-Fang C. Lu
Department of Statistics and Informatics Science, Providence University, Taichung City, Taiwan, R.O.C.

Substantial research has been devoted to developing methodology for inferring the association of clustered failure time data. However, in the study of familial disease, there may be a proportion of patients cured or nonsusceptible to the disease. Thus, it is necessary to simultaneously consider two types of association, i.e., the association of the susceptibility of the individuals, and that of the ages at onset between the susceptible individuals. In this talk, we consider the pairwise association in both types of association to reduce the mathematical intractability and the difficulty in specifying the full correlation structure. The former association is measured by the pairwise odds ratio of the binary cure statuses, and the latter by the bivariate Clayton copula with a semiparametric marginal regression model for any pair of correlated failure times. A two-stage estimation procedure is adopted for the association estimation. We establish the consistency and asymptotic normality of the estimators for these two types of association. Simulation studies are conducted to assess finite sample properties, and the proposed method is illustrated by a subset of the data in the Australian Twins Study.

[Chyong-Mei Chen, Department of Statistics and Informatics Science, Providence University, 200 Chung-Chi Rd., Salu Dist., Taichung City 43301, Taiwan, R.O.C.; cmchen2@pu.edu.tw]

18a7-Model/Variable Selection
December 18 (Sunday), 10:30 - 12:00, AC 4th Conference Room
Chair: Pai-ling Li

18a7-1 Selecting Variance Structure in Mixed Effect Models by Information Criteria Based on Monte Carlo Approximations

Wataru Sakamoto
Osaka University, Japan

Some information criteria for selecting random effect terms and covariance structure in linear mixed effect models have been proposed and compared by Vaida and Blanchard [*Biometrika* 92 (2005): 351–370], Liang *et al*. [*Biometrika* 95 (2008): 773–778], Greven and Kneib [Johns Hopkins University, Dept. of Biostatistics Working Paper, No. 202 (2009)], Dimova *et al*. [*Comput. Statist. Data Anal.* 55 (2011): 2677–2697] and so on. However, we should notice that some of the regularity conditions to guarantee asymptotic properties of information criteria may not be satisfied. Especially, if a supposed model is over-parametrized, existing information criteria would overestimate the bias in estimating an expected log-likelihood by the maximum log-likelihood, as suggested by Sakamoto [*COMPSTAT 2010: Book of Abstracts* (2010): 141]. A quadratic

approximation of the bias is obtained when the parameter space is constrained. It is represented based on a truncated multivariate normal distribution. Also, some information criteria are proposed by using Monte Carlo approximations of the bias. The performance of these criteria are compared with existing criteria by using example data and in a simulation study.

[Wataru Sakamoto, 1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan; sakamoto @sigmath.es.osaka-u.ac.jp]

18a7-2 Variable Selection in Quantile Regression with Autoregressive Errors

Yaeji Lim
Seoul National University, Seoul, Korea

This paper considers a problem of variable selection in quantile regression with autoregressive errors. Recently, Wu and Liu [*Statistica Sinica* 19 (2009):801–817] investigated the oracle properties of the SCAD and adaptive-LASSO penalized quantile regressions under non identical but independent error assumption. We further relax the error assumption so that the regression model can hold autoregressive errors, and then investigate theoretical properties for our proposed penalized quantile estimators under the relaxed assumption. Optimizing the objective function is often challenging because both quantile loss and penalty functions may be non-differentiable and/ or non-concave. We adopt the concept of pseudo data by Oh et al. [*Biometrika* 94 (2007):893–904] to implement a practical algorithm for quantile estimate. In addition, we discuss the convergence property of the proposed algorithm. The performance of the proposed method is compared with MM algorithm (Hunter and Li, [*The Annals of Statistics* 33 (2005):1617–1642]) and DCA (Wu and Liu, Wu and Liu [*Statistica Sinica* 19 (2009):801–817]) through numerical and real examples.

[Hee-Seok Oh, Seoul National University, Seoul 151-747, Korea; heeseok.oh@gmail.com]

18a7-3 Robust Lasso-type Estimation and Efficient Bootstrap Information Criteria in Regression Modeling

Heewon Park
Fumitake Sakaori
and Sadanori Konishi
Department of Mathematics, Chuo University, 1-13-27 Kasuga, Bunkyo- ku, Tokyo 112-8551, Japan

There is currently much discussion about Lasso-type estimation approaches which is a technique for simultaneous estimation and variable selection (in regression modeling) [*J. Roy. Statist. Soc. Ser.* 58 (1996):267–288]. However, in the situation of data including outlier, Lasso-type estimation approaches produce poor estimation result because they are based on the least square method. To

overcome non robustness of Lasso-type estimation approach, several studies are being produced [*J. Business & Economic. Statistics*. 25 (2007):347–355, *ISCAS09* (2009):1193–1196]. In this study, we introduce robust Lasso-type estimation procedures via the elastic net.[*J. Roy. Statist. Soc.Ser.* 67(2005):301–320]. We also propose to use the efficient bootstrap information criteria [*Biometrika*. 83(4) (1996):875–890] for choosing optimal tuning parameters

[Heewon Park, Department of Mathematics, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan; heewonn@gug.math.chuo-u.ac.jp]

18a7-4 Statistical Diagnosis of Longitudinal Models for Binary Data

Takao Takase
Japan Biostatistics, Clinical Science, SOCS unit, Eisai Co., Ltd., Tokyo, Japan

Wataru Sakamoto
Graduate School of Engineering Science, Osaka University, Osaka, Japan

Masashi Goto
Biostatistical Research Association, NPO, Osaka, Japan

In clinical trials of drug development, we often come across longitudinal binary data. It is common to use the logistic regression model (LLRM) for binary outcomes. Recently, marginal models through Generalized Estimating Equations (GEE) and generalized linear mixed effect models (GLIMMIX) which have parameter of correlation between measurements are becoming popular method for analysis of longitudinal binary data because LLRM cannot be apply. On the other hand, it is important to diagnose the validity of logistic transformation in these models because it is hardly checked. Therefore, we use the asymmetric power transformation (APT) that includes logistic transformation as its core to diagnose the validity of logistic transformation and explore a suitable model [*Biometrika*, 68,2, (1981): 357-36; *Behaviormetrika*, 19 (1986): 17-39]. Thus we can obtain the analysis of diagnosis about the logit from case studies on longitudinal binary data. Also we can compare a suitable model using APT with LLRM and GLIMMIX models.

[Takao Takase, 4-6-10, Koishikawa, Bunkyo-ku, Tokyo, 112-8088, Japan; t2-takase@hhc.eisai.co.jp]

[Wataru Sakamoto, 1-3, Machiakaneyamamachi, Toyonaka-shi, Osaka, 560-8531, Japan; sakamoto@sigmath.es.osaka-u.ac.jp]

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

18a7-5 Model Selection in Multilevel Latent Class Models

Hsiu-Ting Yu
McGill University, Montreal, Canada

The Multilevel Latent Class Models (MLCM) proposed by Vermunt (2003) have demonstrated to be a good framework for analyzing nested data with assumed discrete latent constructs. Examples of application range from clinical research to marketing studies. The two levels of discrete latent components were used to interpret the patterns of cross-level dependency in data. With two levels of latent components, the model selection for decision on the numbers of latent components at each level is challenging and has not been systematically examined. Simulation studies were conducted to examine the performance of various information criteria (AIC, BIC, and cAIC) on two factors: measurement factor: sample size at each level, relative sample size, and class-specific response probability; and structural factors: number of latent components at each level, types of dependent structure, and degree of dependency between cross-level components. Results are summarized and presented to provide guidelines and suggestions for empirical researcher.

[Hsiu-Ting Yu, Department of Psychology, 1205 Dr. Penfield Ave, Montreal, QC H3A 1B1, Canada; hsiu-ting.yu@mcgill.ca]

18b1-IASC-ERS Special Invited Session: Feature Selection and Extraction in Predictive Modeling
December 18 (Sunday), 13:00 - 14:30, HSS 1st Conference Room
Organizer: Vincenzo Esposito Vinzi
Chair: Hsing-Kuo Pao

18b1-1 A Comparative Study of Feature Selection Methods

Agostino Di Ciaccio
University of Rome, La Sapienza, Italy

In a regression model, some explanatory variables may be useless, if not disturbing, to our analysis. In this case, it is useful to implement a variable selection procedure that allows us to identify the ”important” variables, eliminating all the others from the model. There are many proposals in the literature in this area and it is difficult to understand what criterion to choose in real life situations. In the literature, it has been mainly studied the consistency of some criteria, considering different definitions and performing simulations on small samples. In general, the authors do not take into account some characteristics of the data that have relevance in the performance of a criterion: sample size, signal/noise ratio, number of variables analyzed, level of correlation between the variables, adequacy of the chosen class of models. In this communication we will compare the properties of the main methods proposed.

[Agostino Di Ciaccio, University of Rome, La Sapienza. Italy; agostino.diciaccio@uniroma1.it]

18b1-2 Grouped Variable Selection via Nested Spike and Slab Priors

Tso-Jung Yen
Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.

Yu-Min Yen
Department of Finance, London School of Economics and Political Science, UK

In this paper we study grouped variable selection problems by proposing a specified prior, called the nested spike and slab prior, to model collective behavior of regression coefficients. At the group level, the nested spike and slab prior puts positive mass on the event that the l_2-norm of the grouped coefficients is equal to zero. At the individual level, each coefficient is assumed to follow a spike and slab prior. We carry out maximum a posteriori estimation for the model by applying blockwise coordinate descent algorithms to solve an optimization problem involving an approximate objective modified by majorization-minimization techniques. Simulation studies show that the proposed estimator performs relatively well in the situations in which the true and redundant covariates are both covered by the same group. Asymptotic analysis under a frequentist's framework further shows that the l_2 estimation error of the proposed estimator can have a better upper bound if the group that covers the true covariates does not cover too many redundant covariates. In addition, given some regular conditions hold, the proposed estimator is asymptotically invariant to group structures, and its model selection consistency can be established without imposing irrepresentable-type conditions.

[Tso-Jung Yen, Institute of Statistical Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan, R.O.C.; tjyen@stat.sinica.edu.tw]

18b1-3 Variable Selection for Generalized Canonical Correlation Analysis

Arthur Tenenhaus
Supelec, Department of Signal Processing and Electronics Systems, France

Regularized Generalized Canonical Correlation Analysis (RGCCA) is a framework to study association between several blocks of variables observed on the same set of observations. Within this framework, all blocks are not necessary fully connected and this allows RGCCA to include a large number of well-known methods as particular cases. Variable selection within RGCCA is of particular interest because it allows recovering the subset of variables of each block that are active in the relationship between connected blocks. Such a variable selection is achieved by applying an L_1 penalization on the RGCCA weight vectors giving rise to the sparse generalized canonical correlation analysis (SGCCA). Searching for a fixed point of the stationnary equation related to SGCCA, a new monotone convergent algorithm is obtained. One of the distinct advantage of SGCCA is that with a single algorithm all particular cases of RGCCA are automatically sparsified.

[Arthur Tenenhaus, Supelec, Department of Signal Processing and Electronics Systems, France; arthur.tenenhaus@supelec.fr]

18b2-Financial Statistics
December 18 (Sunday), 13:00 - 14:30, HSS 2nd Conference Room
Organizer: Henghsiu Tsai
Chair: Mike Ka Pui So

18b2-1 Efficient Simulation of Markov Models with Applications to Risk Management

Inchi Hu
Department of ISOM, Hong Kong University of Science and Technology, Hong Kong

In Fuh and Hu (2004, 2007), Fuh, Hu, Hsu and Wang (2010) an efficient simulation method for small tail probabilities has been developed. The method can be used to compute confidence regions and VaR under various statistical models. In this talk, we will present the method in a more general framework, which include models for dependent data such as autoregressive models and Markov switching models etc. Presenting the method in more general framework helps us to see the essence of the method without relying on special properties of a specific model. Applications to risk management will also be discussed.

[Inchi Hu, Department of ISOM, Hong Kong University of Science and Technology, Hong Kong; imichu@ust.hk]

18b2-2 Large Volatility Matrix Estimation Based on High-Frequency Financial Data

Yazhen Wang
University of Wisconsin-Madison

Large matrix estimation gains an increasing attention in recent years. This talk will present recent work onhigh dimensional statistical problem where a p-dimensional diffusion process is observed with measurementerrors at n distinct time points, and our goal is to estimate the volatility matrix of the diffusion process. Weestablish the minimax theory for estimating large sparse volatility matrices under matrix spectral norm asboth n and p go to infinity. The theory shows that the optimal convergence rate depends on n and p through $n^{-}\sqrt{\ }$ and avolatility matrix estimator is explicitly constructed to achieve the optimal convergencerate.

[YazhenWang, Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue,Madison, WI 53706, USA; yzwang@stat.wisc.edu]

18b2-3 Price Densities in Incomplete Markets: An Application to Weather Derivatives

Wolfgang Häerdle
Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Germany

Brenda Lopez Cabrera
School of Business and Economics, Humboldt-Universität zu Berlin, Germany

Huei-Wen Teng
Graduate Institute of Statistics, National Central University, Taiwan, R.O.C.

A State Price Density (SPD) is the density function of a risk neutral equivalent martingale measure for option pricing, and is indispensible for (exotic) option pricing and portfolio risk management. Many approaches have been proposed to calibrate SPD using financial options from the Bond and Equity Markets in the last two decades. Non and semi parametric methods are preferred because they can avoid model mis-specification of the underlying and thus give insight into complex portfolio propelling. Despite the recent innovation in financial and insurance markets, markets remain incomplete. We infer the SPD from incomplete markets, where usually there is little liquidity of the assets. We calibrate the SPD for weather derivatives, a classical example of incomplete markets and, whose financial contracts payoffs are linked to nontradable assets, for instance weather indices. We apply the Bayesian Quadrature method as a calibration method. The proposed Quadrature method performs better than other nonparametric methods, since it considers the locations and weights of the support points in the finite representation of the SPD as random variables. It allows for statistical inference, it is computationally efficient in the sense that a Gibbs sampler with slice sampling is provided for exploring the posterior distribution, it avoids the call-put parity by computing the SPD of the options type simoultaneously and it is robust even if the market faces the illiquidity issues. Our empirical results help us to understand the dynamics of SPDs in the weather market and their dependency with other markets well studied in the literature.

[Huei-Wen Teng, Graduate Institute of Statistics, National Central University, Taiwan, R.O.C.; wenteng@ncu.edu.tw]

18b3-Knowledge Creation from a Statistical Viewpoint
December 18 (Sunday), 13:00 - 14:30, HSS Media Conference Room
Organizer: Hideyuki Imai
Chair: Lo-Bin Chang

18b3-1 Sitting Posturography for Fatigue Measurements

Mineichi Kudo
Hisataka Nakane

Jun Toyama
Hokkaido University, Japan

Posturography is a general word to mention a clinical assessment to measure the capability of keeping the posture upright. Typically, if the patient gets tired, he/she cannot keep the standing posture upright. So far, such a measurement is carried out in the standing posture, but it implies that the patient has to cooperate to stand on a special equipment. In this paper, we focus on the sitting posture and propose a system using a pressure sensor mat put on a chair to measure the fatigues of the patient. We report the differences between posturography in standing posture and that in sitting posture. If this system becomes widely available, people working at desks would be warned their ”overworking” by the chair that they sit.

[Mineichi Kudo, Division of Computer Science, Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan; mine@main.ist.hokudai.ac.jp]

18b3-2 A Kind of Overall Weights Representation for Absolute Measurement AHP

Shin-ichi Ohnishi
Tatsuhiko Saito
Takahiro Yamanoi
Hokkai-Gakuen University, Sapporo, Japan

Hideyuki Imai
Hokkaido University, Sapporo, Japan

AHP (Analytic Hierarchy Process) has been widely used in a field of decision making. One extension, an absolute measurement method AHP is effective for bad consistency in case containing too many alternatives and can avoid the rank reversal problem. However using the absolute measurement method, the results often lose reliability because the comparison matrix (of criteria or assessment level) does not always have sufficient consistency. In these cases, fuzzy representation for weighting assessment level, criteria and also alternatives using results from a sensitivity analysis is useful. In this paper, we present fuzzy weights representation for absolute measurement AHP by employing some assumptions. The results show how absolute measurement AHP has fuzziness when the comparison matrix is not enough consistent.

[Shin-ichi Ohnishi, Hokkai-Gakuen University, Sapporo, Japan; ohnishi@hgu.jp]

18b3-3 Agents' Kansei Design Based on Variable Neighborhood Models

Seiki Ubukata
Hokkaido University, Sapporo, Japan

In this paper, we discuss how KANSEI of autonomous agents are designed. In order to design agents' KANSEI, we introduce a variable neighborhood model in which agents select multiple neighborhoods under the given context. We introduce the concept of personal space to the model as kinds of neighborhoods. We observed how agents' KANSEI were designed by alterations in size of multiple neighborhoods through multi-agent simulations.

[Seiki Ubukata, Kita 14, Nishi 9, Kita-Ku, Sapporo, Hokkaido 060-0814, Japan; ubukata@main.ist.hokudai.ac.jp]

18b4-IASC Committee on CS & DM in KD Special Invited Session: Model Based Approach to Knowledge Discovery
December 18 (Sunday), 13:00 - 14:30, AC 1st Conference Room
Organizer: Natale Carlo Lauro
Chair: Shun-Yi Chen

18b4-1 **Topological Learning and Learnability**

Djamel Abdelkader Zighed
University of Lyon 2, France

We propose a statistical approach for characterizing the class separability degree in R_p. This approach is based on a nonparametric statistic called the Cut Edge Weight. We show the principle and the experimental applications of this statistic. First, we build a geometrical connected graph like the Relative Neighborhood Graph on a learning dataset. Second, we cut all edges between two examples belonging to dierent classes. Third, we calculate the relative weight of these cut edges. If the relative weight of the cut edges is in the expected interval of a random distribution of the labels on all the neighborhood graphs vertices, then no neighborhood-based method will give a reliable prediction model. We will say then that the classes to predict are non-separable.

[Djamel A. Zighed, Human Science Institute, 14 av. Berthelot, 69007 Lyon, France; Abdelkader.Zighed@ish-lyon.cnrs.fr]

18b4-2 **Ordinal Rating Models for Financial Evaluation**

Paolo Giudici
Paola Cerchiello
Department of Economics, University of Pavia ITALY

In this contribution we propose to estimate the probability of financial default of companies and the correlated rating classes, using efficiently the information contained in different databases. In this respect, we propose a novel approach, based on the recursive usage of Bayes theorem, that

can be very helpful in integrating default estimates obtained from different sets of covariates. Our approach is ordinal: on one hand, the default response variable is categorized in nine rating classes that are ordered according to the average default probability; on the other hand, covariates induce partitioning of companies into the rating classes, according to their measurement levels, when categorical, or to their quantiles, when they are continuous. The application of our proposal to credit risk databases shows that it performs quite efficiently, allowing to obtain, in a natural way, additional measures such as the default value at risk and the lift index.

[Paolo Giudici, Department of Economics, University of Pavia, ITALY; giu- dici@unipv.it]

18b5-Design of Experiments
December 18 (Sunday), 13:00 - 14:30, AC 2nd Conference Room
Organizer: Frederick Kin Hing Phoa
Chair: Frederick Kin Hing Phoa

18b5-1 On Minimal-Point Designs

Dennis K.J. Lin
Department of Statistics, The Pennsylvania State University

A minimal-point design has its number of experimental runs equals to the number of parameters. This is the minimal effort possible to obtain an unbiased estimate for all parameters. Some recent advances for minimal-point design under various models will be discussed. Specifically, a new class of minimal-point design robust to interactions for first-order model is proposed; a new class of minimal-point design, making use of Conference Matrices, for definitive screening will be explored, and if time permits, new minimal-point designs for full second-order response surface models will be discussed. A related issue on the construction of conference matrix and its applications in design of experiment will be introduced.

[Dennis K.J. Lin, Department of Statistics, Pennsylvania State University, U.S.A.; dkl5@psu.edu]

18b5-2 Design for Tuning Parameter Selection in the Lasso

Chunfang Devon Lin
Department of Mathematics and Statistics, Queen's University, Canada

Multi-fold cross-validation is widely used to select tuning parameters for the Lasso and other variable selection methods. Despite its popularity, this procedure is known to have high variability. To overcome this drawback, we propose an experimental design approach to efficiently selecting tuning parameters. This approach exploits slicing in orthogonal Latin hypercube designs. Cross validation under such a design ensures the orthogonality and space-filling properties of the input

values in each fold. The proposed approach can find optimal tuning parameters more efficiently and lead to more accurate variable selection. The effectiveness of the proposed method is illustrated by several examples. This is joint work with Xinwei Deng and Peter Z. G. Qian at the University of Wisconsin-Madison.

[Chunfang Devon Lin, Department of Mathematics and Statistics, Queen's University, Canada; cdlin@mast.queensu.ca]

18b5-3 The Stepwise Response Refinement Screener (SRRS)

Frederick Kin Hing Phoa
Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.

Supersaturated designs are useful in investigating in a large number of factors in a few experimental runs, particularly in the screening experiments. In this talk, a new analysis procedure called the Stepwise Response Refinement Screener (SRRS) method is proposed to screen important effects. It includes two main procedures: Factor Screening and Model Reconstruction. Analysis of several real-life experiments suggests that the SRRS method can retrieve simialr results as the existing methods do. The SRRS method performs well in the simulation studies when compared to some existing methods in the literature including the Dantzig Selector.

[Frederick Kin Hing Phoa, Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.; fredphoa@stat.sinica.edu.tw]

18b6-Applications (II)
December 18 (Sunday), 13:00 - 14:30, AC 3rd Conference Room
Chair: Osamu Komori

18b6-1 Canonical Correlation Analysis as Maximizing Generalized Coefficients of Determination with Oblique Rotation of Canonical Variables

Hironori Satomura
Graduate School of Human Sciences, Osaka University, Osaka, Japan

A new formulation of canonical correlation analysis is proposed, which gives equivalent weights matrices to the existing one. The formulation is defined as maximizing Yanai [*The Japanese J. of behaviormetrics*.1 (1974):46-54]'s generalized coefficients of determination (GCD), and has freedom for the oblique rotation of two canonical variable matrices with non-identical non-singular matrices, unlike the rotation under the existing formulation in which two matrices can be transformed only with identical orthonormal matrix. We proved the equivalence of the formulation by giving new inequality about GCD, and derived new measures for interpreting

obliquely rotated canonical variables, by which we can control the effects of correlation within each canonical variable matrix. The method to rotate canonical variable matrices is proposed and the usefulness of the technique is demonstrated by an empirical application.

[Hironori Satomura, Graduate School of Human Sciences, Osaka University, 1-2, Yamadaoka, Suita, Osaka 565-0871, Japan; satomura@hus.osaka-u.ac.jp]

18b6-2 **Ordered Statistics for Robust Genomic Meta-analysis**

Chi Song
George Tseng
Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, U.S.A.

Microarray experiments for global gene expression monitoring are widely used in genomic research. As tremendous amount of array data are accumulated and available in public domain, meta-analysis tools for information integration of multiple studies are of great demand. In this paper, we first distinguished different hypothesis settings to detect genes differentially expressed in all studies, in majority or in one or more studies. For detecting genes differentially expressed in majority of studies, we proposed an r^{th} ordered p-value (rOP) method that takes the rth smallest p-value across studies combined as the test statistics. We developed algorithms to estimate the best r for a given application. Theoretically, rOP was found connected to vote counting and can be viewed as a generalized form of vote counting with better statistical property. Simulations were performed to compare with existing methods by statistical power. The method was then applied to three real data sets including major depressive disorder (MDD), brain cancer and diabetes. The results showed that rOP performed well to detect genes differentially expressed in majority of studies in both simulation and real examples. Compared with popular approaches, such as Fishers method, maximum p-value and random effects model, rOP provides a more generalized, sensitive and robust framework to detect disease related markers.

[George C. Tseng, 303 Parran Hall, 130 Desoto Street, Pittsburgh, PA 15261 U.S.A.; ctseng@pitt.edu]

18b6-3 **Cluster Analysis of Mouse Ultrasonic Vocalization Data**

Xiaoling Dou
The Institute of Statistical Mathematics, Tokyo, Japan

Shingo Shirahata
Osaka University, Osaka, Japan

Hiroki Sugimoto
Jichi Medical University, Tochigi, Japan

Tsuyoshi Koide
National Institute of Genetics, Shizuoka, Japan

It has been found that male mice emit ultrasonic vocalizations (USVs) towards females during male-female interaction. The purpose of this paper is to classify the waveforms of the mouse USV data. The data are transformed by FFT. Because the USV data are very noisy, it is impossible to analyze them by existing software. We first smooth the USV waveforms from the noisy data by a moving average method, and then fit them with a polynomial regression. After that, we classify the obtained USV curves by a functional clustering method. This analysis also can help us to find a rule (or grammar) of the USVs in communication between mice.

[Xiaoling Dou, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan; xiaoling@ism.ac.jp]

18b6-4 **Predictive Model Development Based on Customer Segmentation**

N. Y. Yi
S. H. Hwang
J. H. Choi
Department of Information & Statistics, Korea University, Republic of Korea

The purpose of this study is to develop the predictive model using the infomation of customers who got lately credit loans and to establish marketing strategies. Most of continuous variable in data mart tend to skew to the right. We use customer segmentation before missing value imputation so as to develop predictive model for each segment. Logistic regression, decision tree, neural network and partial least squares are applied. Real world data of Korean bank company is used for case study. This case study looks at statistical aspects of customer relationship management.

[J. H. Choi, Department of Information & Statistics, Korea University, Jochiwon- eup, Yeongi-gun, Chungnam 339-700, Republic of Korea; jhchoi@korea.ac.kr]

18b6-5 **The Early Detection of Alzheimer's Diseases Based on MRI Data via Sparse Functional Logistic Classification**

Yuko Araki
Atsushi Kawaguchi
Kurume University, Kukurme, Japan

Fumio Yamashita
Research Association for Biotechnology, Tokyo, Japan

Recent years have seen the evidence indicating that MRI brain scans are applicable to both diagnosis

of Alzheimer's disease itself at an early stage and the identification of people at risk of developing Alzheimer's disease. In this talk, we propose a statistical method for classifying patients with Alzheimer's disease based on signal intensities from volume units. Previous studies have considered only a limited brain area because of the considerably increasing dimensionality and computational burden involved in using a large number of voxels. Our method overcomes this problem by using basis expansions in a logistic classification. Additionally, by imposing L_1 -type regularization on the estimation, the proposed method can easily select the relevant features that affects the classification results. We apply the method to real data from the Open Access Series of Imaging Studies and show that the method offers classification functions with excellent prediction performance in terms of sensitivity and specificity.

[Yuko Araki, Biostatistics Center, Kurume University, 67 Asahi-machi, Kurume 830-0011, Japan; araki yuuko@med.kurume-u.ac.jp]

18b7-Financial Statistics & Other
December 18 (Sunday), 13:00 - 14:30, AC 4th Conference Room
Organizer: Zhuyu Li
Chair: Ben-Chang Shia

18b7-1 Efficient Simulation of Markov Models with Applications to Risk Management

Zhuyu Li
Taiji Wang
Mathematical College, Sichuan University, Chengdu, China

Abstract: The study is based on [the lowest, the highest] of the stock market interval time series. Concepts of fuzzy number will be used to characterize the interval data and explore the structures of the linear and nonlinear dynamic modeling. By changing the view of observation, we can get more information. The talk will be based on the compare analysis to the fuzzy double linear model(*FDLR*) and a modified nonlinear central square mixed regression model (*CSMR*). Partial parameter regression and statistic simulation will be used to study the robustness of two models and compare the fitting results.

[Zhuyu LI, Mathematical College, Sichuan University, Chengdu, China, 610064; zyli@scu.edu.cn]

18b7-2 Linear B-spline Copulas with Applications to Nonparametric Estimation of Copulas

Xiaojing Shen
Yunmin Zhu

Lixin Song
Department of Mathematics, Sichuan university, Sichuan Province, China

In this paper, we propose a method of constructing a new class of copulas. They are called linear B-spline copulas which are a good approximation of a given complicated copula by using finite numbers of values of this copula without loss of some essential properties. Moreover, rigorous analysis shows that the empirical linear B-spline copulas are more effective than empirical copulas to estimate perfectly dependent copulas. For the cases of non-perfectly dependent copulas, simulations show that the empirical linear B-spline copulas also improve the empirical copulas to estimate the underlying copula structure. Furthermore, we compare the performance of parametric estimation of copulas based on the empirical copulas with that based on the empirical linear B-spline copulas by simulations. In most of cases, the latter are better than the former.

[Xiaojing Shen, Department of Mathematics, Sichuan university, Sichuan Province, China; xiao23332@163.com]

18b7-3 **Some Construction Methods for Uniform Designs with Large Size**

Yongdao Zhou
College of Mathematics, Sichuan University, Chengdu 610064, China

The main idea of uniform design is to scatter the design points to be uniformly on the experimental domain. As a measure of uniformity, the centered L_2-discrepancy (CD) is often used. Promoting by practical applications, uniform design tables with large size (large number of runs or/and large number of factors) are urgent. In this talk, some construction methods for uniform design with large size are considered. One method is called mixture method, i.e., the uniform design is constructed by mixturing two nearly uniform designs. Moreover, bases on the relationship between CD and integer programming problem, some stochastic algorithm for solving the quadratic integer programming optimization problem is also considered to construct uniform design with large size. Both of the two algorithms have the better than given property, i.e., the resulted design is better than existed designs in the sense of the pre-decided criterion. Empirical study shows these two algorithms can obtain large uniform designs with lower computational complexity and lower CD comparing with many traditional methods.

[Yongdao Zhou, College of Mathematics, Sichuan University, Chengdu 610064, China; ydzhou@scu.edu.cn]

18c1-Bioinformatics and Translational Medicine
December 18 (Sunday), 14:40 - 16:10, HSS 1st Conference Room
Organizer: Hsuan-Yu Chen
Chair: Hsuan-Yu Chen

18c1-1 Molecular Gene-signatures and Cancer Clinical Trials

Mei-Ling Ting Lee
University of Maryland

Over the last ten years the process to develop molecular biomarkers and genomic tests for assessing the risk of cancer and cancer recurrence has been evolving. High-throughput technologies have increased the rate of discovery of potential new markers and facilitated the development of composite gene signatures that provide prognostic or predictive information about tumors. The traditional method to assess the risk of cancer recurrence is based on clinical/pathological criteria. The conventional design has been challenged, especially when the diseases may be heterogeneous due to underlying genomic characteristics. Recently there has been an increase in cancer clinical trials using gene signatures to assess cancer aggressiveness. For example, in some breast cancer studies, it was hypothesized that by using newly developed gene-signature tools one can identify subgroup of patients who will respond significantly to post-surgery (adjuvant) chemotherapy. Future treatments can then be designed to the individual person receiving it and therefore spare the side effects of treatment to a large subgroup of potentially non-responsive patients.
On the other hand, a parallel goal is to identify what is the best treatment for patients: chemotherapy or hormonal therapy. It is important to note that, if one of the major goals of using genomic biomarkers is to move closer to individualized treatment, the biomarkers or gene signatures need to be both prognostic and predictive. Many studies with genomic biomarker and clinical investigations have been conducted in the past few years. In this talk, we review these investigations and results.

[Mei-Ling Ting Lee, Department of Epidemiology and Biostatistics, University of Maryland; mltlee@umd.edu]

18c1-2 Discovering Influential Variables and Applications: A Review

Shaw-Hwa Lo
Columbia University

Advances in A trend in all scientific disciplines, based on technology, is the increasing availability of high dimensional data in which are buried important information. We review a general computer intensive approach, based on an earlier method (Lo and Zheng 02,04) for detecting which, of many potential explanatory variables, have an influence on a dependent variable *Y*. This approach is suited to detect influential variables, where causal effects depend on the confluence of values of several variables. It has the advantage of avoiding a difficult direct analysis,involving possibly thousands of variables, by dealing with many randomly selected small subsets from which smaller subsets are selected, guided by a measure of influence I.The main objective is to discover the influential variables, rather than to measure their effects. Once they are detected, the problem of dealing with a much smaller group of influential variables should be vulnerable to appropriate analysis. If time

permits, applications and findings on IBD and breast cancer dataset will be presented.

[Shaw-Hwa Lo, Department of Statistics, Columbia University; slo@stat.columbia.edu]

18c2-Simulation Techniques in Quantitative Finance
December 18 (Sunday), 14:40 - 16:10, HSS 2nd Conference Room
Organizer: Wolfgang Härdle
Chair: Ray-Bing Chen

18c2-1 Copula Dynamics in CDOs

Barbara Choroś-Tomczyk
Wolfgang Karl Härdle
Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Germany

Ludger Overbeck
Department of Mathematics, University of Giessen, Giessen, Germany

The recent financial crisis has revealed inefficiencies of methods used in credit risk management. The rapid development of the credit market in the past decade was due to new possibilities that were offered by innovative credit derivatives like CDS, default baskets, collateralized debt obligations (CDOs). We analyse the European market of standardized CDOs using tranches of iTraxx index in the periods before and during the global financial crisis. We compare pricing methods based on different multivariate copula functions. The values of CDOs strongly depend on current economic conditions; therefore, correlations implied from tranches can be seen as measures of the general health of the market. We exploit the time-series dynamics of the calibrated copula parameters and model their joint distribution. We asses the changing risk of iTraxx tranches by calculating VaR over time. For better understanding of the nature of the dependence, we investigate the relations between the implied parameters and observable indicators of economic and financial conditions.

[Barbara Choroś-Tomczyk, Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany; barbara.choros@wiwi.hu-berlin.de]

18c2-2 Measuring the Collective Correlation of Stock Market

Wei-Fang Niu
Henry Horng-Shing Lu
Institute of Statistics, National Chiao Tung University

Correlations are closely related to the crash of the stock market. This paper proposes a measure for the collective correlation of the market from the viewpoint of network. By digitizing pairwise correlations the market network can be easily constructed. Then the transitivity of the network simultaneously addresses the level of correlations and clustering effect and can be used to as a collective correlation index. The statistical properties of the index are also investigated by simulations. With a sample from S&P 500 in the period from January 1996 to August 2009, we found that the index is co-integrated with the CBOE VIX. This result complies with recent theoretical developments for the mechanism of market crash.

[Wei-Fang Niu, Risk Management Institute, National University of Singapore; siegfriedniu@gmail.com]

18c2-3 **Regularized Pairwise Estimator of Realized Covariance**

Ying Chen
Department of Statistics & Applied Probability, Faculty of Science, National University of Singapore

Vladimir Spokoiny
Weierstrass Institute for Applied Analysis and Stochastics

This paper proposes a new estimator of realized covariance based on ultra-high frequency (UHF) data. The new estimator handles 1) asynchrony of raw data, 2) microstructure noise of UHF records and guarantees semi-positiveness of the covariance estimator. In particular, a three-step algorithm is proposed. Firstly, a synchronizing technique for irregularly spaced and asynchronous UHF data is conducted that learns from the dependence structure of raw data and iteratively recovers the ”missing” values of a synchronous series at ultra-high sampling frequency. Then two scaled estimators are computed for each pair elements, which help to reduce the effect of microstructure noise. The combination of all the pairwise estimators leads to a weak covariance estimator that is not necessarily to be semi-positive definite. The regularized covariance estimator is obtained via an optimization of the weak estimator. We compare the feature of the regularized estimator with several alternative realized covariance estimators. The proposed method provides a good performance in terms of accuracy and feature.

[Ying Chen, National University of Singapore: Department of Statistics & Applied Probability, Faculty of Science, Block S16, Level 7, 6 Science Drive 2, Singapore 117546; stacheny@nus.edu.sg]

18c3-Statistics Education
December 18 (Sunday), 14:40 - 16:10 , HSS Media Conference Room
Organizer: Tae Rim Lee
Chair: Sheau-Hue Shieh

18c3-1 Elementary Statistics Education in Smart Phone

Jae Woong Kim
Moon Yul Huh
Dept. of Statistics, Sunkyunkwan University, KOREA

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

In this work, we develop an app of android system for elementary statistics education. The app has three components: firstly, efficient elementary statistics education needs to consider both of the two following elements - basic statistics concept and data analysis; secondly, considering user-friendly interface; thirdly, utilizing abundant information of website.
The app developed in this paper consists of Learning concept, Data analysis, Glossary and Note. Also the system invokes the contents of Wikipedia whenever we need it. These components are implemented on a smart phone with 4 or 4.3 inch screen using android system. The main components are Learning concept, Data analysis and Wikipedia mobile website. Additionally, Glossary and Note components are available through menu button. The system can easily be extended to include other components if we need it.

[Dae-Heung Jang, 599-1 Daeyeon-dong, Nam-gu, Busan, KOREA; dhjang@pknu.ac.kr]

18c3-2 Certification of Statistical Literacy by Japan Statistical Society

Yasuto Yoshizoe
Aoyama Gakuin University, Tokyo, Japan

Japan Statistical Society plans to start a set of examinations for students and professional statisticians or interviewers working in various fields of statistics. The exam called *JSSC*, which stands for JSS Certificate, consists of four levels of statistics (methods and interpretation) for students, and two levels for professional survey workers and researchers. In addition, JSS will introduce a higher level exam jointly with Royal Statistical Society that is conducted both in English and in Japanese, starting next May. By setting up JSSC, JSS wishes to improve educational environment in statistics, enhancing the ability of students in statistical way of thinking, namely, statistical literacy, and help government and other organizations collect better statistical data.

http://www.yoshizoe-stat.jp/index en.html

[Yasuto Yoshizoe, Shibuya, Tokyo 150-8366, Japan; yasuto yoshizoe@post.harvard.edu]

18c3-3 Transforming Statistics Education through ICT Application

Tae Rim Lee
Dept. of Information Statistics, KNOU

As the role of ICT has increasingly taken root in educational institutions, there is a growing need for research to find a way to have statistics education merged in society with ever-changing technology. Mobile learning (m-Learning) is novel in that it facilitates delivery of learning to the right person, at the right time, in the right place using portable electronic devices. In the near future, m-Learning will be a normal part of lifelong education and self-directed learning. In Korea, m-Learning is expanded in almost every sector of educational fields. From last year KNOU kick off the mobile learning system with KT. In this paper the mobile learning for statistics education will be introduced. The objectives of this paper were to evaluate the effectiveness of e-learning courses for teaching Statistics and to propose suggestions for future improvement of e-Learning courses and suggest the future view of more advanced education system of mobile learning and ubiquitous learning system. In the future society with rapid change of educational circumstance and globalization, Statistics education using ICT technology will satisfy educational desires in various classes of learners. KNOU has been provided student with distance education contents through broadcasting and IT. Mobile technologies, like mobile devices and wireless internet services, have the potential to introduce new innovations in the area of education m-learning, a new form of education using mobile internet systems and handheld devices can offer students and teachers the opportunity to interact with and gain access to educational materials, independent of time and spaces. This study suggested some considerable suggestions for preparing the future of statistics education based on Mobile and one more step advanced ubiquitous learning.

[Tae Rim Lee, Dept. of Information Statistics KNOU; trlee@knou.ac.kr]

18c4-Statistics and Visualization
December 18 (Sunday), 14:40-16:10, AC 1st Conference Room
Organizer: Kwan-Liu Ma
Chair: Chuan-kai Yang

18c4-1 Efficient Exploration of 3D Scalar Field based on Statistical Space Analysis

Wei Chen
State Key Lab of CADCG, Zhejiang University, Hangzhou, 310058, P.R.China

Exploring 3D scalar fields focuses on revealing hidden structures in volumetric datasets. Effective exploration is a challenging problem because there is no prior information available with regard to the data distribution. To better disambiguate the various materials and structures, a number of

multi-dimensional transfer functions have been proposed. However, the selection of features within the multi-dimensional transfer function is a time-consuming and trial-and-error process, and it is very likely to yield unsatisfactory results. The gap between the exibility of multi-dimensional transfer function design and the delity requirement of the volume exploration makes the transfer function design challenging. This talk proposes a new scheme that performs the classification in the statistical space by regarding the volumetric exploration as a multi-labeling problem in the 2D statistical space. Three different analysis techniques are employed, namely, the KDE(kernel density estimation), GMM (Gaussian Mixture Model) and semi-supervised learning (SSL). Experimental results are given.

[Wei Chen, State Key Lab of CADCG, Zijingang Campus, Zhejiang University, Hangzhou, 310058, P.R.China; chenwei@cad.zju.edu.cn]

18c4-2 **Visual Analysis of Vehicle Trajectory Data**

Huamin Qu
Department of Computer Science and Engineering, Hong Kong University of Science and Technology

In this talk I will show how visualization techniques can be used together with statistics and data mining methods to analyze vehicle trajectory data. First, for vehicles equipped with GPS devices, their recorded trajectories are often erroneous or at least inaccurate due to various reasons. The errors cannot be easily fixed, especially for trajectories with low sampling rates. I will present a visual analytics system, which can help domain experts detect trajectory error patterns and correct wrong records. Next, I will show how we can gain insight into taxi drivers' mobility intelligence by analyzing a large amount of taxi trajectory data collected over a long time period. Some novel visualization techniques will be introduced to identify the key factors that differentiate top drivers from ordinary drivers. Finally, I will introduce the route diversity problem and present a visualization system, which can show all the major routes for a given source/destination pair and reveal key statistics for each route. Our system can be used for route suggestion, traffic monitoring, and transportation management.

[Huamin Qu, Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong; huamin@cse.ust.hk]

18c4-3 **The Role of Statistics in Medical Visualization**

Ivan Viola
University of Bergen, and Scientific Adviser, Christian Michelsen Research (CMR), Bergen, Norway

Visualization of medical imaging data is nowadays essential for establishing the diagnosis and the treatment plan. For complex datasets, such as longitudinal population studies or perfusion measurements, interactive data visualization alone is becoming inefficient to enable discovery and communication of findings. The reason is the increased data size and lack of data analysis that can guide the clinician to important pieces of information. Coupling visualization with statistical approaches, the clinician can observe the significant population trends, study the outliers, reduce the number of dimensions by statistical methods and use the visualization for displaying the derived statistical data as well as the imaging data. Integrating statistical overview with 3D structural visualization enables comparisons among population, understand functional behavior of pathological and healthy tissue or relate structural tissue changes over aging.
[Ivan Viola, University of Bergen, Norway; ivan.viola@uib.no]

18c5-Recent Advances in Transformation Methods
December 18 (Sunday), 14:40-16:10, AC 2nd Conference Room
Organizer: Byeong Uk Park
Chair: Byeong Uk Park

18c5-1 Thick-pen Transform and Its Extension

Hee-Seok Oh
Seoul National University, Korea

Piotr Fryzlewicz
London School of Economics, UK

Traditional visualization of time series data often consists of plotting the time series values against time and ”connecting the dots”. We propose an alternative, multiscale visualization technique, motivated by the scale-space approach in computer vision. In brief, our method also connects the dots, but uses a range of pens of varying thicknesses for this purpose. The resulting multiscale map, termed the Thick-Pen Transform (TPT) corresponds to viewing the time series from a range of distances. We formally prove that the TPT is a discriminatory statistic for two Gaussian time series with distinct correlation structures. Further, we show interesting possible applications of the TPT to measuring cross-dependence in multivariate time series, classifying time series, and testing for stationarity. In particular, we derive the asymptotic distribution of our test statistic, and argue that the test is applicable to both linear and nonlinear processes under low moment assumptions. Various possible extensions of the methodology are also discussed.

[Hee-Seok Oh, Department of Statistics, Seoul National University, Seoul 151-747, Korea; heeseok@stats.snu.ac.kr]

18c5-2 **Smoothing Empirical Mode Decomposition**

Donghoh Kim
Sejong University, Seoul, Korea

Empirical Mode Decomposition (EMD) has been used in many disciplines of science and engineering since its development by Huang et al. [Proc. Roy. Soc. London A 454 (1998):903–995]. A key feature of EMD is the decomposition of a signal into so-called Intrinsic Mode Functions (IMFs). This work considers extending the scope of EMD method. The extension is aimed at noisy signal, which is necessary for widespread applicability of EMD. The proposed method uses a smoothing technique instead of an interpolation when constructing upper and lower envelopes for extracting IMF. We discuss how to identify noise from a signal, and how to decompose noisy signal into IMFs without distortions.

[Donghoh Kim, 98 Gunja-dong, Sejong University, Seoul 143-747, Korea; donghoh.kim@gmail.com]

18c5-3 **Robust Estimation of Principal Functions Using M-Estimation**

Seokho Lee
Hyejin Shin
Nedret Billor
Seoul National University, Korea

We propose a robust functional principal component analysis method when curve data contain outlying observations. Robust principal functions are estimated by penalized M-estimation with the roughness penalty. In the proposed method, since abnormally observed measurements in a single curve can be separately downweighted by low weights rather removing or downweighting a whole curve containing such abnormalities, we can make a use of maximal information on whole observed data. Natural cubic spline is used to represent the estimated principal function so that computation becomes fast even when data consist of lengthy curves as in typical functional data. We test the performance of the proposed method using lip movement data and simulated datasets in the comparison with the existing functional principal component analysis.

[Hyejin Shin, Seoul National University, Korea; hjshin112@gmail.com]

18c6-Time Series
December 18 (Sunday), 14:40-16:10, AC 3rd Conference Room
Chair: Tsung-I Lin

18c6-1 Test of Mean Difference for Paired Longitudinal Data Using Stationary Bootstrap

Hirohito Sakurai
Masaaki Taguri
National Center for University Entrance Examinations, Tokyo, Japan

We propose a testing method for detecting the difference of two means in paired longitudinal data using the stationary bootstrap of Politis and Romano [*J.Amer. Statist. Assoc*. 89 (1994):1303V1313]. We use four kinds of test statistics: sum of absolute values of difference, sum of squares of difference, area-difference between two mean curves, and the test statistic used in Hall and Hart [*J.Amer. Statist. Assoc*. 85 (1990):1039V1049]. In order to approximate the null distribution of each test statistic, the stationary bootstrap is used. We carry out numerical examinations on empirical level and power for each statistic and compare the behavior among the tests with four statistics. Further, we also compare our level and power results with those obtained by the test of Bowman and Young [*Appl. Statist*. 45 (1996):83V98].

[Hirohito Sakurai, 2V19V23 Komaba, Meguro-ku, Tokyo 153V8501, Japan; sakurai@rd.dnc.ac.jp]

18c6-2 Vine-copula GARCH Model with Dynamic Conditional Dependence

Yuk Ting Yeung
Mike Ka Pui So
Hong Kong University of Science and Technology

Constructing a multivariate distribution for non-Gaussian return series has been a major research agenda in recent papers. Copula-GARCH models, which combine the use of traditional GARCH models and a copula function, allow flexibility on the choice of marginal distributions and dependence structures. However, copula-GARCH models have mainly been explored in low-dimensional cases because of the model complexity for high dimensional problems. In this paper, we propose a vine-copula GARCH model with dynamic conditional dependence. The main idea is to make use of the vine decomposition [Bedford and Cooke (2002), Annals of Statistics] to reduce the complexity of our models. In addition, motivated by the idea in DCC-GARCH models, dynamic dependence is incorporated in each bivariate copula pair. In an empirical study of Europe stock data with 30 minutes interval from 2007 to 2009, we find that estimates from bivariate copula pairs in each layer are similar to each other. We also discuss methods to simplify our models when dealing with a large number of stocks simultaneously.

[Yuk Ting Yeung, Hong Kong University of Science and Technology; ytyeungaa@ust.hk]

18c6-3 Modeling Dynamic Seasonality in Time Series

Mike K.P. So
Ray S.W. Chung
The Hong Kong University of Science and Technology, Hong Kong

In this research, a new class of time series model capturing dynamic seasonality is introduced. Unlike the traditional dynamic seasonality models which focus mainly on dynamic seasonality in the mean process, our approach can accommodate dynamic seasonality in the variance process. This feature allows us to perform statistical inference of the dynamic seasonality in heteroskedastic time series models. Quasi-maximum likelihood estimation and forward selection are adopted for estimation and model selection. Simulation study is carried out to evaluate the efficiency of the estimation method. In empirical examples, our model outperforms deterministic seasonality models in forecasting the monthly domestic electricity consumption in Hong Kong and the intraday stock return variation.

[Ray S.W. Chung, Department of Information System, Business Statistics and Operations Management, The Hong Kong University of Science and Technology, Clear Water Bay Road, Kowloon, Hong Kong; imray@ust.hk]

18c6-4 Asymptotic Behavior of Temporal Aggregates in the Frequency Domain

Uwe Hassler
Goethe University, Frankfurt, Germany

Henghsiu Tsai
Academia Sinica, Taipei, Taiwan, R.O.C.

If a stochastic process in discrete time is integrated of order *r, r* = 0, 1, 2, ⋯, then the temporal aggregate turns with increasing level of aggregation into an integrated (of order *r*) moving-average [MA] model, where the MA component is of order r with parameters independent of the underlying basic process. This classical result by Tiao [Biometrika 59 (1972):525-531] is generalized for a weak set of assumptions. The innovations driving the integrated processes are only required to be stationary with summable autocovariances. The derivation is settled in the frequency domain. In case of fractional integration it is further demonstrated that the order of integration is preserved with growing aggregation, too, under the same set of assumptions.

[Uwe Hassler, Goethe University Frankfurt (RuW), Grueneburgplatz 1, D-60323 Frankfurt, Germany; hassler@wiwi.uni-frankfurt.de]

18c7-Statistical Meta Analysis
December 18 (Sunday), 14:40 - 16:10, AC 4th Conference Room
Organizer: Philip Leung-ho Yu
Chair: Philip Leung-ho Yu

18c7-1 On Some Aspects of Statistical Meta-analysis with an Application

Bimal K. Sinha
Department of Mathematics and Statistics, University of Maryland, Baltimore County, U.S.A.

Statistical meta-analysis deals with meaningfully combining or integrating results from several independent studies all apparently with a common research objective. I will present an overview of statistical meta-analysis with applications. In particular, several tests for homogeneity of rare proportions will be discussed.

[Bimal K. Sinha, Department of Mathematics and Statistics, University of Maryland, Baltimore County, U.S.A.; sinha@umbc.edu]

18c7-2 On Statistical Inference for Direct and Indirect Multiple Treatment Comparisons

Guido Knapp
TU Dortmund University, Dortmund, Germany

Multiple or mixed treatment comparisons can be seen as a generalization of meta-analysis in which the information on more than two treatments are combined and, for estimating a treatment difference between two treatments, direct and/or indirect estimates may be available from independent multi-arm trials. In this paper, we discuss several models for multiple treatment comparisons. We show how to deal with possible heterogeneity and inconsistency between the observed treatment difference to end up with valid estimates for common or overall treatment differences. The standard approach for multiple treatment comparisons based only on indirect estimates yields too many unjustied significant results. We will discuss an improvement which better controls the nominal Type I error.

[Guido Knapp, Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany; guido.knapp@tu-dortmund.de]

18c7-3 A Comparison of Meta Analysis using Literature and using Individual Patient Data

Thomas Mathew
University of Maryland Baltimore County, Baltimore, Maryland, U.S.A.

Kenneth Nordstrom
University of Oulu, Finland

In the literature on meta-analysis, there is considerable work on summary versus individual patient data. However, recent articles in the medical literature indicate that there is still confusion and uncertainty as to the precision of an analysis based on aggregate data. In the talk we address this issue based on linear models for individual patient data. The setup, which allows for the presence of random effects and covariates in the model, is quite general. The one-way fixed-effects model and the two-way model without interaction and fixed or random study effects are all obtained as special cases. For this general model, we derive a condition for the estimator based on summary data to coincide with the one obtained from individual patient data. The implications of this result for some specific models are illustrated in detail, both theoretically and in terms of two data sets, and the role of balance is highlighted.

[Thomas Mathew, Regular mailing address: Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland 21250, U.S.A.; E-MAIL ADDRESS: mathew@umbc.edu]

18d1-Environmental Statistics
December 18 (Sunday), 16:30 - 18:00, HSS 1st Conference Room
Organizer: Hizir Sofyan
Chair: Hung-wen Cheng

18d1-1 Applied of Response Surface Method on Reverse Osmosis Process Using Polyurethane Membrane

Marlina
Sitti Saleha
Chemistry Department, Syiah Kuala University, Darussalam-Banda Aceh

Hizir Sofyan
Mathematics Department, Syiah Kuala University, Darussalam-Banda Aceh

Synthesis of polyurethane (PU) membrane from Jatropha Oil and methylene diisocyanate (MDI) had optimum condition at composition 25 : 5 (v/w), temperature 100 ℃ , time of polymerization 60 minutes, time of curring between 190 ℃ -195 ℃ for 72 hours. Membrane obtained was a brown, homogeneous, strong, and elastic. Characteristics membrane were :absorpsion of NHCOO (urethane group) at 3438.8 cm^{-1}, and CNH at 1220.9 cm^{-1}, temperature of glass transition at 161 ℃ ,

decomposition point at 614℃ , shear streght 93.33 kgf/mm^2 and elongation as 58.90%. Application this membrane on reverse osmosis (RO) process gave flux as 29.94 L m^{-2} h^{-1} and rejection factor as 12.5%, and showed no stability in flux after applied as long as 100 hours on RO process. Based on these characteristics, the membrane could not be applied on RO process (too high elongation and weakness). This result supported by applied of respon surface analysis at orde 1, where optimum condition on RO process were temperature = 2.6549 ℃ , pressure = 3.3586 atm, and filtration time =-0.3178 minutes, showed that the result not realigible for RO process.

[Marlina, Chemistry Department, Syiah Kuala University, Darussalam-Banda Aceh; marlina rachman@yahoo.com]

18d1-2 Seasional ARIMA Recurrent Neural Networks for Forcasting Short-term Maximum and Minimum Temperature

Suhartono
Department of Statistics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Sepuluh Nopember, Indonesia

Neural Network (NN) is one of the methods that usually used for forecasting non-linear time series data. NN which is mostly used in many researches is Feed-Forward Neural Network (FFNN) or Autoregressive Neural Network (AR-NN). AR-NN is unable to catch and represent the pattern of moving average (MA) order in time series data. This research focuses to study further another NN type for forecasting, i.e. Elman-Recurrent Neural Network (Elman-RNN) which is able to represent MA order, and to compare the forecast accuracy with Seasonal Autoregressive Integrated Moving Average (SARIMA) model. Maximum and minimum temperature data from one of Meteorology Stations in Indonesia are used as the case study. The result shows that the best ARIMA model for forecasting short-term maximum and minimum temperature is ARIMA([1,19],1,1)(0,0,1)365 and ARIMA([1,6,14],1,[1,5,11])(1,0,0)365, respectively. In this study, the inputs of Elman-RNN are determined based on the ARIMA lag inputs. There are three nets built for each maximum and minimum temperature data, i.e. input layer, one hidden layer with tangent sigmoid function, and one output layer with linear function. Based on the value of Mean Absolute Percentage Error (MAPE) in testing data, the results show that the best net for forecasting the maximum and minimum temperature are Elman-RNN(4,5,1) and Elman-RNN(11,9,1), respectively. Moreover, the results of forecast accuracy comparison show that both Elman-RNN models yield more accurate forecasted values than SARIMA model.

[Suhartono, Department of Statistics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Sepuluh Nopember, Indonesia; suhartono@statistika.its.ac.id]

18d1-3 Geometric Approach to Empirical Analysis of Landslide Disaster Prediction

Anton Abdulbasah Kamil
School of Distance Education, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.

This study presents a new mathematical geometric technique for the empirical analysis of disaster probability risk and stability of slopes using all hydrological and geological factors. Within our new static and dynamic frameworks, we propose the use of a new transformation and representation technique which named Landslide Disaster Prediction Index. Accuracy results of different prediction models like Multiple Discriminant Analysis (MDA), Logistic Analysis (LA) and Genetic Programming (GP) compared and results show the outperform of classifications by new geometrical index.

[Anton Abdulbasah Kamil, School of Distance Education, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.; anton@usm.my]

18d2-Analysis of Financial Time Series
December 18 (Sunday), 16:30 - 18:00, HSS 2st Conference Room
Organizer: Cathy W. S. Chen
Chair: Yasuhiro Omori

18d2-1 Time-Simultaneous Prediction Bands: A New Look at the Uncertainty Involved in Forecasting Mortality

Wai-Sum Chan
The Chinese University of Hong Kong, Hong Kong

Johnny Siu-Hang Li
The University of Waterloo, Canada

Conventionally, isolated (point-wise) prediction intervals are used to quantify the uncertainty in future mortality rates and other demographic quantities such as life expectancy. A pointwise interval reects uncertainty in a variable at a single time point, but it does not account for any dynamic property of the timeseries. As a result, in situations when the path or trajectory of future mortality rates is important, a band of pointwise intervals might lead to an invalid inference. To improve the communication of uncertainty, a simultaneous prediction band can be used. The primary objective of this paper is to demonstrate how simultaneous prediction bands can be created for prevalent stochastic models, including the Cairns-Blake-Dowd and Lee-Carter models. The illustrations in this paper are based on mortality data from the general population of England and Wales.

[Wai-Sum Chan, Department of Finance, The Chinese University of Hong Kong, Hong Kong; chanws@cuhk.edu.hk]

18d2-2 Predictive Approach for Model Selection on Econometric Models with Factor Augmented Predictors

Tomohiro Ando
Keio University, Japan

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

Econometric models with a factor augmented regressors provide a useful approach to forecasting when there are many predictors available. A properly chosen model can summarize relevant information about the variable of interest in a small number of indices and, hence, achieve substantial dimension reduction in the predictor space. The application of econometric models with a factor augmented regressors often involves two steps. In the first step, the common factors (or indices) are estimated via the principal component analysis. In the second step, the estimated common factors are used in conjunction with other pre-specified variables to build a model for forecasting. For successful applications of these models, selection of the best model is critical. In this paper, we propose a model selection criterion for generalized linear, quantile regression models that takes into account the uncertainty in estimated common factors. Results of real data analysis and Monte Carlo simulations demonstrate clearly that the proposed criterion performs well.

[Tomohiro Ando, Graduate School of Business Administration, Keio University 4-1-1 Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa,223-8526, Japan; andoh@kbs.keio.ac.jp]

18d2-3 Conditional Quantile Estimation in Heteroskedastic Time Series Models

Mike Ka Pui So
Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Hong Kong

This paper studies the statistical properties of a conditional quantile estimator under heteroskedastic time series models with the error distribution unspecified. The asymptotic distribution of the quasi-maximum likelihood estimators and the filtered empirical percentiles is derived. Three applications of the asymptotic result are considered. First, we construct interval estimator of the conditional quantile without any distributional assumption. Second, we develop a specification test for the error distribution. Finally, using the specification test, we propose methods to estimate the tail index of the error distribution, from which we can also construct a new conditional high quantile estimator. The asymptotic results and the applications are illustrated by simulations and real data analysis where we adopt our methods in analyzing daily and intraday financial return series.

[Mike Ka Pui So, Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Hong Kong; immkpso@ust.hk]

18d3-Environmental and Biological Statistics
December 18 (Sunday), 16:30 - 18:00, HSS Media Conference Room
Organizer: Yu-Shan Albert Shih
Chair: Yu-Shan Albert Shih

18d3-1 Pseudomartingale Estimating Equations for Modulated Renewal Process Models

Feng-Chang Lin
University of North Carolina at Chapel Hill, Chapel Hill, U.S.A.

We adapt martingale estimating equations based on gap time information to a general intensity model for a single realization of a modulated renewal process. The consistency and asymptotic normality of the estimators is proved under ergodicity conditions. Previous work has considered either parametric likelihood analysis or semiparametric multiplicative models using partial likelihood. The framework is generally applicable to semiparametric and parametric models, including additive and multiplicative specifications, and periodic models. It facilitates a semiparametric extension of a popular parametric earthquake model. Simulations and empirical analyses of Taiwanese earthquake sequences illustrate the methodologys practical utility. The current work was published in J. R. Statist. Soc. Ser. B 71, 3-23.

[Feng-Chang Lin, 160 N. Medical Dr. Brinkhous-Bullitt Building 2nd Floor, CB #7064, Chapel Hill, NC 27599; flin@bios.unc.edu]

18d3-2 Nonparametric Estimation and Test of Conditional Kendall's Tau between Failure and Truncation Times

Jin-Jian Hsieh
Department of Mathematics, National Chung Cheng University Chia-Yi, Taiwan, R.O.C.

In this article, we focus on estimation and test of conditional Kendall's tau under left-truncated data, double-truncation data, and bivariate left-truncation data. We apply the Inverse Probability Censoring Weighted (IPCW) technique to construct an estimator of conditional Kendall's tau, τ_c , and provide a Wald's type test statistic to test $H_0 : \tau_0 = \tau_c$, where $\tau_0 \in (-1, 1)$. When two random variables are quasi-independent, it implies $\tau_c = 0$. Thus, $H_0 : \tau_c = 0$ is a proxy for quasi-independence. Tsai (1990), and Martin and Betensky (2005) also considered the testing problem for quasi-independence. We compare the three test statistics for quasi-independence in simulation

studies. Furthermore, we provide the large sample properties for our proposed estimator. We also examine the finite-sample performance of the proposed estimator and the suggested test statistic via simulations. Finally, we provide a real data analysis for illustration.

[Jin-Jian Hsieh, Department of Mathematics, National Chung Cheng University Chia-Yi, Taiwan, R.O.C.; jjhsieh@math.ccu.edu.tw]

18d3-3 Multiple Point Genetic Association Studies with an Uncertainty-coding Matrix

C. K. Hsiao Y. H. Huang W. J. Chen
Department of Public Health and Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan, R.O.C.

M. H. Lee
Department of Mathematics and Computer Science Education, Taipei Municipal University of Education, Taipei, Taiwan, R.O.C.

Haplotype association studies based on family genotype data can provide more biological information than single marker analysis. Difficulties arise, however, in the inference of haplotype phase determination and transmission/non-transmission status. Incorporation of the uncertainty into regression models requires special care. This task can get even more complicated when the genetic region contains a large number of haplotypes. To avoid the curse of dimensionality, we employ a clustering algorithm based on the evolutionary relationship among haplotypes and retain for regression analysis only the ancestral core haplotypes identified by it. To integrate the three sources of variation, phase ambiguity, transmission status and ancestral uncertainty, we propose an uncertainty-coding matrix which combines these three types of variability simultaneously in a Bayesian regression model. Simulation studies and applications are presented and compared with other tools for family association studies. Our proposed method performs better and the implementation in R is freely available.

[Dr. C. K. Hsiao, Department of Public Health and Institute of Epidemiology and Preventive Medicine, National Taiwan University, No 17, Xu-Zhou Road, Taipei 100, Taiwan, R.O.C.; ckhsiao@ntu.edu.tw]

18d4-Applications of Clustering Techniques
December 18 (Sunday), 16:30 - 18:00, AC 1st Conference Room
Organizer: Donguk Kim
Chair: Hsiao-Yun Huang

18d4-1 Clustering of Climate Functional Data Using Mixture Models

Yongsung Joo
Department of Statistics, Dongguk University

In recent decades, we observe increase of annual mean temperature, accompanying unusual alternations of hot and cold years (large year-to-year variation). Also, the changes in temporal precipitation patterns is caused by the complex interaction among temperature change, the global water cycle, and other components of the Earth system. To analyze these temporal changes in mean and variance of temperature and precipitation, we propose the logistic mixture model for temporal heteroscedastic profiles and apply it to US climate data between 1900-2008. Our results show that the proposed model classifies the traditional climatology of US well with two important meteorological variables, temperature and precipitation, and identify the characteristics of each climate zone. Surface air temperature increased eminently in Northeast and Northern Midwest regions, of which climate is affected by arctic air. Increment of precipitation is also observed in Northeast region. On the other hand, climate in the Southeast region, of which climate is affected by tropical Atlantic ocean, is more stable than Northeast and Northern Midwest regions in terms of mean and variance of temperature and precipitation. Our analysis result supports that the polar warming proceeds more rapidly than other regions causing higher variability in precipitation.

[Yongsung Joo, Department of Statistics, Dongguk University; yongsungjoo@dongguk.edu]

18d4-2 Estimation of the Central Subspace Using Clustering

Jae Keun Yoo
Ewha Womans University, Seoul, Korea, Republic of

Popular sufficient dimension reduction methodologies such as sliced inverse regression, sliced average variance estimation and so on require linearity and coverage conditions for proper estimation of the central subspace. In this talk a new subspace to contain the central subspace is defined and is estimated, at least, not requiring linearity condition. For methodological implementation, first, all predictors are clustered by restricted *K*-means algorithm, which limits the minimum samples in each clusters. This approach generalizes existing clustered-based dimension reduction methods. Simulation studiesand data analysis are presented.

[Jae Keun Yoo, Department of Statistics, Ewha Womans University, 52, Ewhayeodaegil, Seodaemun-gu, Seoul 120-750, Korea, Republic of; peter.yoo@ewha.ac.kr]

18d4-3 A New Selection Model via Clustering Missingness Indicators

Hyekyung Jung
Samsungcard CO. LTD, Seoul, Korea

Joseph L. Schafer
Pennsylvania State University, University Park, U.S.A.

Byungtae Seo
Sungkyunkwan University, Seoul, Korea

The missing-at-random (MAR) assumption is very convenient for both modeling and computing when data contain missing values. However, in some cases, this assumption is questionable because missing values are often strongly related to unobserved values themselves. For this missing-not-at-random (MNAR) problem, we present a new method that uses a latent class model to cluster indicators of missingness. The proposed method can be also used as a tool to judge the validity of assumptions for the missing data mechanism.
[Byungtae Seo, ungkyunkwan University, Seoul, Korea; seobt@skku.edu]

18d5-Statistical Methods for Social Studies
December 18 (Sunday), 16:30 - 18:00, AC 2nd Conference Room
Organizer: Ting Hsiang Lin
Chair: Esher Hsu

18d5-1 A Study of Smoothing Methods on Small Area Population Projection

Shuoh Jin
Jack C. Yue
National Chengchi University

The population size plays a very important role in statistical estimation, and it is difficult to derive a reliable estimation for small areas. The estimation is even more difficult if the geographic and social attributes within the small areas vary widely. However, although the population aging and longevity risk are common phenomenon in the world, the problem is not the same for different countries. The aim of this study is to explore the population projection and mortality models for small areas, with the consideration of the small areas distinguishing characteristic.

The difficulties for small area population projection can be attributed into four directions: data quality, population size, number of base years, and projection horizon. The data quality is beyond the discussion of this study and the main focus shall be laid on the other three issues. The smoothing methods and coherent models will be applied to improve the stability and accuracy of small area estimation. In the study, the block bootstrap and the smoothing methods are combined to project the population to the small areas in Taiwan. Besides, the Lee-Cater and the age-period-cohort model are extended by the smoothing and coherent methods.

We found that the smoothing methods can reduce the fluctuation of estimation and projection in general, and the improvement is especially noticeable for areas with smaller population sizes.

To obtain a reliable population projection for small areas, we suggest using at least fifteen-year of historical data for projection and a projection horizon not more than twenty years. Also, for developing mortality models for small areas, we found that the smoothing methods have similar effects than those methods using more complicated models, such as the coherent models.

[Jack C. Yue, National Chengchi University; csyue@nccu.edu.tw]

18d5-2 Issues Related to Analyzing International Large-scale Assessment Data

Hak Ping Tam
National Taiwan Normal University

In recent years, such international large-scale assessment surveys as TIMSS and PISA have gained much attention from applied researchers around the world. These studies usually employed complex sampling designs (for example, two-stage stratified cluster sampling design) to select their research participants. For the sake of fairness, a large number of items were necessarily included for a reasonable coverage of the different curricula across countries. Hence intricate booklet designs were adopted through which items were systematically assigned to different booklets while each participant was required to answer just one booklet. The data structures from these studies are usually more complicated than those previously encountered by most applied researchers. Applications of the usual data analytical practices are inappropriate under most circumstances. This paper will discuss a number of issues that needed to be considered when undertaking secondary data analysis on such data sets, including the purposes of these large-scale assessment studies as well as some technical aspects that must be dealt with when working on these kinds of data. Several recommendations will be given at the end of the paper.

[Hak Ping Tam, National Taiwan Normal University; t45003@ntnu.edu.tw]

18d5-3 A Missing Treatment Adjustment for Nested Regression Analysis of DeviantBehavior – A TEPS Case Study

Hong-Long Wang
National Taipei University, New Taipei City, Taiwan, R.O.C.

Missing data might cause biases if completely delete the missing items cases when statistical analysis is involved. In this paper, through a study of on 2nd wave data of ”Taiwan Education Panel Survey” (in short as TEPS), we will explore possible biases of missing data when nested regression model applied on deviant behavior study. Missing treatments will be introduced to adjust these possible biases. We will also compare the missing treatments effect on the nested regression analysis with various missing proportions of data sets.

[Hong-Long Wang, National Taipei University, New Taipei City, Taiwan, R.O.C.; hlw@gm.ntpu.edu.tw]

18d6-Bayesian Statistics and Computing
December 18 (Sunday), 16:30 - 18:00, AC 3rd Conference Room
Organizer: Chul Eung Kim
Chair: Jia-hong Wu

18d6-1 Asymmetric Classification with Normal Populations Perturbed by Common Interval Screening

Hea-Jung Kim
Dongguk University-Seoul, Seoul, Korea

Statistical methods for asymmetric normal classification do not adapt well to the situations where the population distributions are perturbed by an interval screening scheme. This paper explores methods for providing optimal classification of future samples in this situation. Properties of the screened population distributions are considered and two optimal regions for classifying the future sample are obtained. These developments yield yet other rules for the interval-screened asymmetric normal classification. The rules are studied from several aspects such as probability of misclassification, robustness, and estimation of the rules. Investigation of the rule's performance as well as illustration of the screened classification idea (using a numerical example) are also considered.

[Hea-Jung Kim, Pil-Dong 3Ga, Chung-Ku, Department of Statistics, Dongguk University, Seoul 100-715, Korea; kim3hj@dongguk.edu]

18d6-2 A Bayesian Semiparametric Approach to Estimation of Median Incomes of Small Areas

Dal Ho Kim
Department of Statistics, Kyungpook National University, Korea

Estimation of median incomes of small areas is one of the principal targets of inference of the U.S Bureau of Census. Since these estimates are collected over time, they often possess an inherent logitudinal pattern. Taking proper account of this time varying pattern may result in better estimates for the current or future median household incomes for a particular state or county. In this study, we put forward a semiparametric modeling procedure for estimating the median household income for all the U.S states. Our models include a nonparametric functional part for accomodating any unspecified time varying income pattern and also a state specific random effect to account for the within-state correlation of the income observations. Model fitting and parameter estimation

is carried out in a hierarchical Bayesian framework using Markov chain Monte Carlo (MCMC) methodology.

[Dal Ho Kim, Department of Statistics, Kyungpook National University, Korea; dalkim@knu.ac.kr]

18d6-3 Species Sampling Model and Posterior Consistency

Gun Ho Jang
Jaeyong Lee
Sangyeol Lee
University of Pennsylvania, Seoul National University

We give a brief overview of the species sampling models, the nonparametric priors defined as the directing random probability measures of the species sampling sequences. We then will talk about the posterior consistency issue with the species sampling models under the simple iid setting nonparametric model. In particular, we present a result that in the class of Pitman-Yor process priors the only priors rendering posterior consistency are essentially the Dirichlet process priors. Under certain conditions, we also give a set of necessary and sufficient conditions for the posterior consistency for the general species sampling prior. Examples considered include the normalized inverse-Gaussian process, Poisson-Kingman partition and Gibbs partition.

[Jaeyong Lee, University of Pennsylvania, Seoul National University; leejyc@gmail.com;jylc@snu.ac.kr]

7th IASC-ARS 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