Session I-1

June 29, 13:30 – 14:50

理工一館 A310

魏慶榮論文獎場次

◆ **Organizer/Chair：**

潘建興 - 中央研究院 統計科學研究所

◆ **Speaker:**

1. 董亦賢 – 中央大學統計研究所

2. 邱詠惠 – 中央大學統計研究所

3. 潘彥碩 – 清華大學統計學研究所

The first-passage-time moments for the Hougaard process and its Birnbaum–Saunders approximation

Chien-Yu Peng (彭健育)[1], Yi-Shian Dong (董奕賢)*[2], Tsai-Hung Fan (樊采虹)[2]

[1]Institute of Statistical Science, Academia Sinica

[2]Graduate Institute of Statistics, National Central University

Abstract

Hougaard processes, which include gamma and inverse Gaussian processes as special cases, as well as the moments of the corresponding first-passage-time (FPT) distributions are commonly used in many applications. Because the density function of a Hougaard process involves an intractable infinite series, the Birnbaum–Saunders (BS) distribution is often used to approximate its FPT distribution. This article derives the finite moments of FPT distributions based on Hougaard processes and provides a theoretical justification for BS approximation in terms of convergence rates. Further, we show that the first moment of the FPT distribution for a Hougaard process approximated by the BS distribution is larger and provide a sharp upper bound for the difference using an exponential integral. The conditions for convergence coincidentally elucidate the classical convergence results of Hougaard distributions. Some numerical examples are proposed to support the validity and precision of the theoretical results.

Keywords: Characteristic function, Contour integration, Exponential dispersion model, Residue, Stirling numbers.

A Semiparametric Approach for Estimating Fixed Effects with Spatial Confounding

Yung-Huei Chiou（邱詠惠*）, Chun-Shu Chen（陳春樹）

Graduate Institute of Statistics, National Central University

Abstract

In spatial regression analysis, confounding between fixed effects and random effects adversely affects the estimates of regression coefficients. This paper proposes a novel estimation method for regression coefficients based on the fixed rank kriging approach which does not require specifying any parametric covariance structures for response variables and hence is more practical. The proposed method involves selecting the number of basis functions which affects the bias and the variance of estimators. Two methods based on resampling and error loss are respectively proposed to control the mean squared errors of estimators, resulting in the Bagging estimator and γ-estimator for estimating regression coefficients. Theoretical properties related to the proposed methodology are justified. The simulations under the settings of spatial regression models with spatial confounding reveal that the proposed estimation methods of regression coefficients perform well regardless ofthe underlying correlation structure is stationary, nonstationary, isotropic, or anisotropic. Finally, an application to precipitation data in Colorado is presented.

Keywords: Basis function, Bias reduction, Fixed rank kriging, Mean squared error, Restricted spatial regression

MPCGA: A Tree-Based Chebychev's Greedy Algorithm

Yan-Shuo Pan*(潘彥碩)[1], Ching-Kang Ing(銀慶剛)[1], Guan-Hua Huang(黃冠華)[2]

[1]Institute of Statistics, National Tsing Hua University

[2]National Yang Ming Chiao Tung University

Abstract

Prediction and feature selection are essential topics for statistical and machine learning (ML) methods when dealing with high-dimensional data. However, they come with limitations: statistical methods may exhibit lower predictive ability than ML methods, while ML methods are often criticized as black boxes. In this paper, we introduce a tree-based algorithm, the Multipath Chebyshev Greedy Algorithm (MPCGA), which enhances the predictive performance of statistical methods and feature selection capabilities under model misspecification. This algorithm extends the Chebychev's Greedy Algorithm (CGA) and High Dimensional Information Criterion (HDIC) into a tree-expanded structure, allowing for the simultaneous consideration of multiple models. MPCGA outperforms traditional statistical methods when models are misspecified, while maintaining high feature selection precision. Furthermore, we propose accelerated algorithms to boost the computational speed of MPCGA handling indicator features in binary outcome cases. The paper includes a case study on a lung cancer dataset, demonstrating that utilizing ML methods with the feature set selected by MPCGA leads to suitable results.

Keywords: Chebyshev's Greedy Algorithms, high-dimensional information criterion, machine learning, feature selection, logistic regression model