Waseda Lecture Series in Statistics

Venue: Waseda University, Nishi-Waseda Campus Building 63 - 1 Meeting Room

(Access map: http://www.sci.waseda.ac.jp/eng/access)

Date: March 18, 2019

(I)Speaker: Benjamin Kedem, University of Maryland, College Park

Emcee: Fumiya Akashi, Waseda University

March 18 (Mon.)

（1） 10:00--11:00 Time Series Analysis by Higher Order Crossings

（2） 11:00--12:00 Regression Models for Time Series Analysis

（3） 13:30--14:30 Statistical Data Fusion

(II) Speaker : Violetta Dalla, Department of Economic Sciences, National and Kapodistrian University of Athens (NKUA)

Emcee: Yuichi Goto, Waseda University

(4) 14: 45 – 16:00 Standard Testing Procedures for White Noise and Heteroskedasticity

Supported by “Introduction of General Causality to Various Data & Its Innovation of the Optimal Inference”

JSPS Kiban(S) (No. 18H05290) (Taniguchi, M. Waseda University)

(1) Title: Time Series Analysis by Higher Order Crossings

Abstract: Imagine a bank of filters applied to a finite signal or time series, indeed a pervasive engineering practice. Typically, the application of each filter from the bank changes the signal oscillation pattern and alters the zero-crossing count. Accordingly, the application of each member filter gives rise to a zero-crossing count, and the zero-crossing counts resulting from the application of some or all the filters are called higher order crossings, or HOC for short. This talk is about HOC and related ideas. It explores, for the most part, an approach to time series analysis based on the surprisingly fruitful connection between filtering and zero-crossing counts. In his important pioneering work in the 1920's, E.E. Slutsky used statistical and probabilistic tools to explain the effect of repeated linear filtering on the spectrum of a stationary time series. He recognized then what today is common knowledge, namely that conspicuous periodicities observed in a time record may be the result of linear filters applied to some underlying process, thus enhancing certain spectral bands. This program of repeated filtering, or more generally the application of filters from a designated family, is followed in this talk, the extra feature being the involvement of zero-crossing counts observed in the filtered data. Our message is that in addition to the spectrum and autocorrelation, HOC families and sequences too constitute viable tools which can be used to summarize the oscillatory content of time series. The oscillatory information contained in HOC families and sequences is useful for fast signal classification and spectral analysis, particularly frequency estimation in the presence of ambient noise

(2) Title: Regression Models for Time Series Analysis

Abstract: The generalized linear models (GLM) methodology has had a great impact on modern regression analysis. However, the methodology, introduced in a seminal paper by Nelder and Wedderburn in 1972, assumes independent data. In this talk we shall address the question of how to extend the generalized linear models methodology to time series, where the data are dependent,

possibly nonstationary, and the covariates are time dependent and also random. Our approach to the problem is to use partial likelihood in conjunction with temporal conditional inference with respect to a filtration generated by all that is known to the observer at the time of observation.

(3) Title: Statistical Data Fusion

Abstract: The density ratio model provides an inferential framework for semiparametric inference vis-a-vis fused data, such as meteorological satellite data fused with ground truth, fused data from several sensors, and fused case and control data. A concrete application where the density ratio model is used in connection with fused data is equi-distribution testing, this leading to a great generalization of the one-way ANOVA, obviating the normal assumption. Another application is time series prediction by predictive distributions. Yet another concrete example where the density ratio model is used is the estimation of small tail probabilities using numerous fusions (could be millions) of real and computer generated data in what nowadays is called augmented reality. In this talk we shall:

• Review the density ratio model and some of its basic underpinnings. • Discuss briefly a Bayesian extension applied to radar data. • Discuss time series prediction by out of sample fusion. • Argue that at times augmented reality is "better than real", a case in point is the estimation of small tail probabilities.

Regarding the estimation of small tail probabilities, often, it is required to estimate the probability that a quantity such as mercury, lead, toxicity level, plutonium, temperature, rainfall, damage, wind speed, risk, etc., exceeds an unsafe high threshold. The probability in question is then very small. To estimate such a probability, we need information about large values of the quantity of interest. However, in many cases, the data only contain values far below the designated threshold, let alone exceedingly large values, which ostensibly renders the problem insolvable. It is shown that by repeated fusion of the data with externally generated random data, more information about small tail probabilities is obtained with the aid of certain new statistical functions. This provides short, yet reliable interval estimates based on moderately large samples. A comparison of the approach with a method from extreme values theory (Peaks over Threshold, or POT), using both artificial and real data, points to the merit of repeated out of sample fusion.

(4) Standard Testing Procedures for White Noise and Heteroskedasticity

Violetta Dalla* co-authored by
L. Giraitis and P.C.B. Phillips

Commonly used tests to assess evidence for the absence of serial correlation between time series in applied work rely on procedures whose validity holds for i.i.d. data. When the series are not i.i.d., the size of correlogram and cumulative Ljung-Box tests can be significantly distorted. This paper adapts standard correlogram tests to accommodate hidden dependence and non-stationarities involving heteroskedasticity, thereby uncoupling these tests from limiting assumptions that reduce their applicability in empirical work. To enhance the Ljung-Box test for non i.i.d. data a new cumulative test is introduced. Asymptotic size of these tests is unaffected by hidden dependence and heteroskedasticity in the series. An extensive Monte Carlo study confirms good performance in both size and power for the new tests. Applications to real data reveal that standard tests frequently produce spurious evidence of serial correlation.