WASEDA

INTERNATIONAL SYMPOSIUM

- High Dimensional Statistical Analysis for Time Spatial Processes & Quantile Analysis for Time Series -

Organized by

Masanobu Taniguchi

(Research Institute for Science & Engineering, Waseda University)

- **Date:** October 24 - 26, 2016
- **Venue:** Waseda University
  Nishi-Waseda Campus Building 55S
  2nd Floor, Room 3
- **Supported by**
  - Kiban (A-15H02061)
    M. Taniguchi, Research Institute for Science & Engineering, Waseda University
  - Houga (26540015)
    M. Taniguchi, Research Institute for Science & Engineering, Waseda University

Waseda International Symposium

High Dimensional Statistical Analysis for Time Spatial Processes,
Quantile and Empirical Likelihood Analysis for Time Series

Date: October 24 - 26, 2016

Venue: Waseda University
Nishi-Waseda Campus Building 55S
2nd Floor, Room 3
(Access map: http://www.sci.waseda.ac.jp/eng/access/)

Organizer: Masanobu TANIGUCHI
(Research Institute for Science & Engineering, Waseda University)

Supported by
(1) Kiban (A-15H02061) M. Taniguchi, Research Institute for Science & Engineering, Waseda University
(2) Houga (26540015) M. Taniguchi, Research Institute for Science & Engineering, Waseda University

Program

(* Speaker)

October 24

Session (I): 13:30 -15:00: chaired by K. Chen

13:30 - 13:40: Masanobu Taniguchi (Waseda Univ.)
Opening

13:40 - 14:20: Yan Liu* (Waseda Univ.), Kun Chen, Ngai Hang Chan and Masanobu Taniguchi
A frequency domain bootstrap for irregularly spaced spatial data

14:20 - 15:00: Xiaoling Dou (Waseda Univ.)
An investigation of the Ochi estimator in the first order of ARCH model

15:00 - 15:10: Break

Session (II) (Invited): 15:10 -17:10: chaired by H. Künsch

15:10 - 16:10: Sangyeol Lee (Seoul National Univ.)
Recent developments in the change point test of integer-valued time series

16:10 - 17:10: Katsuto Tanaka (Gakushuin Univ.)
Estimation problems associated with the fractional O-U process in the ergodic and non-ergodic cases

October 25

Session (III) (Ph.D. session): 9:45 - 10:15: chaired by H. Nagahata

9:45 - 10:00: Yoshiyuki Tanida* (Waseda Univ.) and Masanobu Taniguchi
Asymptotic theory of Whittle functionals for high-dimensional time series

10:00 - 10:15: Yujie Xue* (Waseda Univ.), Yan Liu and Masanobu Taniguchi
Robust interpolation problems in L_p

10:15 - 10:30: Break

Session (IV) (Gene session): 10:30 - 12:00: chaired by F. Akashi

10:30 - 11:15: Yuji Chikashige (NICT)
Analysis of gene expression in growing fission yeast cells

11:15 - 12:00: Hiroto Tanaka (NICT)
An improved understanding of chemical recognition of E. coli with a data-driven

12:00 - 13:30: Lunch

Session (V) (High-Dim session): 13:30 -14:10 chaired by Y. Liu

13:30 - 14:10: Kazuyoshi Yata* (Univ. Tsukuba) and Makoto Aoshima
Reconstruction of a high-dimensional low-rank matrix and its applications

Session (VI) (Invited): 14:10 -17:20: chaired by S. Lee

14:10 - 15:10: Arup Bose (Indian Statistical Institute)
Large sample behaviour of high dimensional autocovariance matrices when the dimension grows at the same rate as the sample size

15:10-15:20: Break

15:20 - 16:20: Yosihiko Ogata* (ISM, ERI), Koich Katsura, Hiroshi Tsuruoka and Naoshi Hirata

Models for seismic activity beneath the Greater Tokyo Area

16:20 - 17:20: Hans R. Kunsch* (ETH Zurich), Andreas Papritz and Werner Stahel
Robust geostatistics

18:00 - Buffet Party

October 26

Session (VII) (Ph.D. session): 9:45 - 10:00 chaired by Y. Xue

9:45 - 10:00: Hideaki Nagahata* (Waseda Univ.) and Masanobu Taniguchi
Analysis of variance for multivariate time series

Session (VIII): 10:00 - 11:30: chaired by A. Bose

10:00 - 10:40: Koji Tsukuda (Univ. Tokyo)
Estimating the large mutation parameter of the Ewens sampling formula

10:40 - 10:50: Break

10:50 - 11:30: Fumiya Akashi (Waseda Univ.), Shuyang Bai and Murad S. Taqqu
Quantile regression-based self-normalized block sampling method for linear regression model with dependent errors

Session (IX) (Invited): 11:30 - 12:20 chaired by A. Bose

11:30 - 12:20: Kun Chen (South Western Univ.)
On Bartlett correction of Empirical Likelihood for Regularly Spaced Spatial Data

Abstract

(* Speaker)

Yan Liu

A frequency domain bootstrap for irregularly spaced spatial data

Abstract: In this talk, we consider the problem of bootstrapping irregularly spaced spatial data. Although there are many methods to bootstrap dependent data, we adopt a frequency domain bootstrap from computational time and its scope of application. The frequency domain bootstrap is a methodology to bootstrap periodogram based on Studentized periodogram ordinates. We take a specific increasing set of discrete frequencies for the bootstrap since the domain of the finite Fourier transform is not bounded for irregularly spaced data. We show that the frequency domain bootstrap is second-order correct for classes of ratio statistics under mixed increasing domain. The performance of the frequency domain bootstrap is shown in our simulation study. The method is also applied to some real examples. (Joint work with Kun Chen, Ngai Hang Chan and Masanobu Taniguchi)

Xiaoling Dou

An investigation of the Ochi estimator in the first order of ARCH model

Abstract: Ochi (1983) proposed an estimator for the first order of autoregression (AR) model with two constants for the end points of the process. Classical estimators of the parameter correspond to special choices of the constants in the Ochi estimator. Writing the autoregressive conditional heteroskedasticity model of order 1, ARCH(1), into a similar form of AR(1), the least square estimator, Burg's estimator and Yule-Walker's estimator can be found are special cases of the Ochi estimator. With a simulation, we investigate the Ochi estimator for ARCH(1) with different parameters and different sample sizes.

Sangyeol Lee

Recent developments in the change point test of integer-valued time series

In this study, we consider the problem of testing for parameter change inzero-inflated generalized Poisson autoregressive models and bivariate Poisson autoregressive models. We verify that the considered processes are stationary and ergodic and that the conditional maximum likelihood estimators are strongly consistent and asymptotically normal. Then, based on these results, we construct CMLE- and residual-based cumulative sum tests and show that their limiting null distributions are a function of independent Brownian bridges.

Katsuto Tanaka

Estimation problems associated with the fractional O-U process in the ergodic and non-ergodic cases

Abstract: We discuss some inference problems associated with the fractional Ornstein-Uhlenbeck (fO-U) process driven by the fractional Brownian motion (fBm). In particular, we are concerned with the estimation of the drift parameter, assuming that the Hurst index is known. The discussion is divided into three cases, depending on the sign of the drift parameter. The one is the ergodic case, another the non-ergodic case, and the other the boundary case. Under this setting, we compute the distributions of the maximum likelihood estimator (MLE) and the minimum contrast estimator (MCE) for the drift parameter, and explore their distributional properties by paying attention to the influence of the Hurst index and the sampling span. We shall also derive the asymptotic distributions of the two estimators as the sampling span becomes large. We further deal with the ordinary least squares estimator (OLSE) and examine the asymptotic relative efficiency. It is shown that the MCE is asymptotically efficient, while the OLSE is inefficient, unlike the case of the ordinary O-U process. We also consider the unit root testing problem in the fO-U process and compute the powers of the tests based on the MLE and MCE.

Yoshiyuki Tanida

Asymptotic theory of Whittle functionals for high-dimensional time series

Abstract: Recently, in many fields involving electrical and genome engineering, high-dimensional and low sample size data are often observed, and the various methods have been investigated to deal with such data appropriately. The most of classical results were proved in the situation where the model has Gaussianity and the sample is independent and identically distributed. However, it is often sever to assume Gaussianity for real time series data and the methods are required to be extended. In this paper, we develop the asymptotic theory for Whittle functional of high-dimensional non-Gaussian dependent observations. In particular, thresholded-periodogram matrix is proposed, and asymptotic normality of the proposed for the functional is shown. (Joint work with Masanobu Taniguchi)

Yujie Xue

Robust interpolation problems in L_p

Abstract: We consider the interpolation problem in L_p under the condition that the spectral density of a stationary process concerned is vaguely known. For a zero mean stationary process, the linear extrapolation problem can be considered as a distance problem from the constant 1 to the closed linear hull of the set generated by the Fourier basis without 1. The robust interpolator in L_2 is found in Taniguchi (1981). In this talk, we discuss the case of $p \geq 1$. The interpolation error in L_p is shown and the closed form of the interpolator is found under $p > 1$. The construction of the robust interpolator in L_p (and under further general conditions) is discussed. (Joint work with Yan Liu and Masanobu Taniguchi)

Yuji Chikashige

Analysis of gene expression in growing fission yeast cells

Abstract: Free living cells maintain proliferation as much as possible in the given environment. The gene expression profile is thought to change according to the given environment to bring the best proliferation at each environment. In this report, we will describe the changes of gene expression profile in the proliferating cells based on the empirical measurements of using the fission yeast.

Hiroto Tanaka

An improved understanding of chemical recognition of E. coli with a data-driven

Abstract: Detection and recognition of chemical substances are primitive but vital for living organisms to survive. Organisms can detect and recognize various chemicals by using a relatively small number of receptors, integrate the information, and take actions. We focus on the organisms' ability of flexible recognition of environmental chemicals, and characterize the ability in a data-driven manner. We observe chemotactic behavior of *E. coli* against several chemical attractants, and analyze relationship between chemical stimuli and chemotactic response with a statistical framework. Our study reveals that *E. coli* recognize chemical species during chemotaxis.

Kazuyoshi Yata

Reconstruction of a high-dimensional low-rank matrix and its applications

Abstract: In this talk, we consider the problem of recovering a low-rank signal matrix in high-dimensional situations. The main issue is how to estimate the signal matrix in the presence of huge noise. We introduce the power spiked model to describe the structure of singular values of a huge data matrix. We first consider the conventional PCA to recover the signal matrix and show that the estimation of the signal matrix holds consistency properties under severe conditions. The conventional PCA is heavily subjected to the noise. In order to reduce the noise we apply the noise-reduction (NR) methodology by Yata and Aoshima (2012, JMA) and propose a

new estimation of the signal matrix. We show that the proposed estimation by the NR method holds the consistency properties under mild conditions and improves the error rate of the conventional PCA effectively. Finally, we demonstrate the reconstruction procedures by using a microarray data set. (Joint work with Makoto Aoshima)

Arup Bose

Large sample behaviour of high dimensional autocovariance matrices when the dimension grows at the same rate as the sample size

Abstract: Consider a sample of size n from a linear process of dimension p where $n, p \to \infty$, $p/n \to y \in (0, \infty)$. Let $\hat{\Gamma}_u$ be the sample autocovariance of order u. The existence of limiting spectral distribution (LSD) of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$, the symmetric sum of the sample autocovariance matrix $\hat{\Gamma}_u$ of order u, is known in the literature under appropriate (strong) assumptions on the coefficient matrices. Under significantly weaker conditions, we prove, in a unified way, that the LSD of any symmetric polynomial in these matrices such as $\hat{\Gamma}_u + \hat{\Gamma}_u^*$, $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $\hat{\Gamma}_u \hat{\Gamma}_u + \hat{\Gamma}_k \hat{\Gamma}_k^*$ exist. Our approach is through the more intuitive algebraic method of free probability that is applicable after an appropriate embedding, in conjunction with the method of moments. Thus, we are able to provide a general description for the limits in terms of some freely independent variables. All the previous results follow as special cases. We suggest statistical uses of these LSD and related results in problems such as order determination and white noise testing.

Yosihiko Ogata

Models for seismic activity beneath the Greater Tokyo Area

Abstract: Development of point-process models for the seismicity in 3D space (longitude, latitude and depth) beneath the Kanto area down to 100km depth is more required than for seismicity in the rest of the world. We consider the

location-dependent space-time ETAS (epidemic-type aftershock sequence) models. In addition, the impact of the 2011 Tohoku-Oki earthquake of M9.0 to the seismicity beneath the Kanto region has been so large that we need a space-time intensity function for representing the amount of the induced seismicity beneath the Kanto Plain by this giant earthquake. For forecasting future large earthquakes beneath the Kanto region, we further need to estimate the location-dependent Gutenberg-Richter law of magnitude frequency. (Joint work with Koich Katsura, Hiroshi Tsuruoka and Naoshi Hirata)
(ISM = Institute of Statistical Mathematics, ERI = Earthquake Research Institute, University of Tokyo)

Hans R. Kunsch
Robust geostatistics

Abstract: We consider a Gaussian random field $(Z(s); s \in R^2)$ whose mean depends linearly on known predictors and whose covariance function is stationary and continuous. Observations of $Z(s)$ are subject to observation noise which is independent at different locations and possibly long-tailed. Besides the regression coefficients β, the model has additional parameters denoted by θ which describe the covariance function of Z and the scale of the observation errors. The goal is to estimate β and θ and to predict the Gaussian random field Z at locations without observations (kriging), together with an estimate of the prediction variance. In the Gaussian case, the parameter θ is usually estimated by restricted maximum likelihood (REML) and then plugged into the generalized leastsquares estimator of β and in the universal kriging estimator for predictions. Our method is based on a robust modification of the estimating equations for Gaussian REML estimates. I will explain the main ideas behind this approach, discuss open theoretical problems and show some results from simulations and from the analysis of a data set on soil contamination by heavy metals. (Joint work with Andreas Papritz and Werner Stahel)

Hideaki Nagahata

Analysis of variance for multivariate time series

Abstract: For independent observations, analysis of variance (ANOVA) has been enoughly tailored. Recently there has been much demand for ANOVA of dependent observations in many fields. For example it is important to analyze differences among industry averages of financial data. However ANOVA for dependent observations has been immature. In this paper, we study ANOVA for dependent observations. Specifically, we show the asymptotics of classical tests proposed for independent observations and give a sufficient condition for them to be asymptotically χ^2 distributed. If the sufficient condition is not satisfied, we suggest a likelihood ratio test based on Whittle likelihood and derive an asymptotic χ^2 distribution of our test. Some numerical examples for simulated and real financial data are given as applications of these results. This presentation opens a new aspect for the analysis of variance for time series. (Joint work with Masanobu Taniguchi)

Koji Tsukuda

Estimating the large mutation parameter of the Ewens sampling formula

Abstract: The Poisson-Dirichlet distribution, a probability distribution on an infinite-dimensional simplex, can be derived as the stationary distribution of descending order allele frequencies assuming the infinitely-many neutral allele diffusion model, and the Ewens sampling formula (ESF) gives the distribution of the allelic partition of a sample from a random population following the Poisson-Dirichlet distribution. In this talk, asymptotic properties of estimators for the parameter θ in the ESF and some applications are presented. The results are as follows: (1) under several asymptotic situations that the parameter grows with the sample size, we derive some new limit theorems associated with the total number of alleles; (2) we show the consistency and the asymptotic normality of the maximum likelihood estimator for large θ; (3) we apply the results of (2) to some statistical problems in the field of genetics.

Fumiya Akashi

Quantile regression-based self-normalized block sampling method for linear regression model with dependent errors

Abstract: Self-normalized block sampling approach formulated by Bai et al. (2016) provides a unified view of robust inference approach for the population mean of time series models with possibly infinite variance and long-range dependent error terms. Motivated by Bai et al. (2016), this talk extends the classical approach toward the parameter estimation problem of a linear regression model with dependent error terms. In particular, we propose quantile regression-based self-normalized block sampling approach, and construct an estimation procedure for the unknown parameter of the model without any priori estimation of nuisance quantities. The model and methodology are nicely generalized by the quantile regression approach, and the proposed approach is shown to have robustness against infinite variance, infinite mean and long-range dependence of the model. Some simulation experiments are also reported, and we observe that the proposed method also works well under the finite sample situation. (Joint work with Shuyang Bai and Murad Taqqu)

Kun Chen

On Bartlett correction of Empirical Likelihood for Regularly Spaced Spatial Data

Abstract: Bartlett correction is one of the desirable features of empirical likelihood that allows constructions of confidence regions for parameters with improved coverage probabilities. In this paper we study Bartlett correction of spatial frequency domain empirical likelihood based on general spectral estimating functions for regularly spaced spatial data. This general formulation can be applied to testing and estimation problems in spatial data, for example testing covariance isotropy, testing covariance separability as well as estimating parameters of spatial models. We also show that the spatial frequency domain empirical likelihood is Bartlett correctable. In particular, the improvement of coverage errors of the Bartlett corrected C.I. depends on the underlying asymptotic structures.