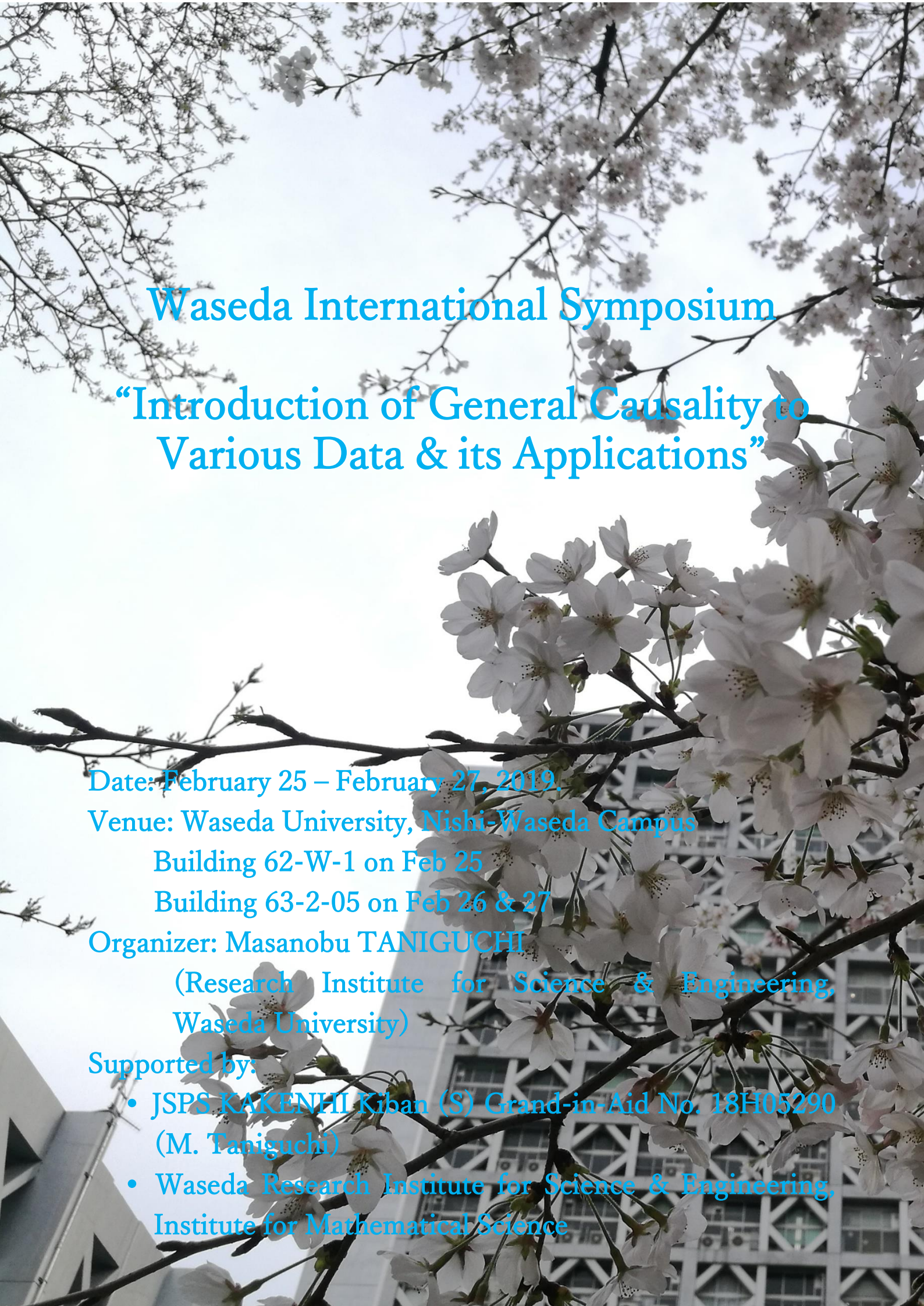

Waseda International Symposium

“Introduction of General Causality to Various Data & its Applications”

Date: February 25 – February 27, 2019.

Venue: Waseda University, Nishi-Waseda Campus
Building 62-W-1 on Feb 25
Building 63-2-05 on Feb 26 & 27

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

Supported by:

- JSPS KAKENHI Kiban (S) Grand-in-Aid No. 18H05290 (M. Taniguchi)
- Waseda Research Institute for Science & Engineering, Institute for Mathematical Science

Waseda International Symposium
"Introduction of General Causality to Various Data & its Applications"

Date: February 25–27, 2019

Venue: Waseda University, Nishi-Waseda Campus
Building 62-W-1 on Feb 25,
Building 63-2-05 on Feb 26 & 27
(Access map: https://www.waseda.jp/fsci/en/access/)

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

This workshop is supported by:

- JSPS KAKENHI Kiban (S) Grand-in-Aid No. 18H05290 (M. Taniguchi)
- Waseda Research Institute for Science & Engineering, Institute for Mathematical Science

Program (*speaker)

February 25

Session I (13:00–15:10): chaired by Tommaso Proietti

13:00–13:10: Opening
Masanobu Taniguchi

13:10–13:50:
Fréchet-Hoeffding copula bounds for circular data
Hiroaki Ogata* (Tokyo Metropolitan University)

13:50–14:30:
A measure for comparing upper and lower tail probabilities of bivariate distributions
Shogo Kato* (Institute of Statistical Mathematics, Japan), Toshinao Yoshiba, Shinto Eguchi

14:30–15:10:
Non-Sparse Modeling for High-Dimensional Data
Makoto Aoshima (Tsukuba University)

15:10–15:25: Coffee break

Session II (15:25–17:25): chaired by Hormann Siegfried

15:25–16:05:
Predictability, Real Time Estimation, and the Formulation of Unobserved Components Models
Tommaso Proietti* (University of Rome, "Tor Vergata")

16:05–16:45:
Inference on Time Series with Changing Mean and Variance
V Dalla, Liudas Giraitis* (Queen Mary University of London), PM Robinson

16:45–17:25:
Smallest singular value and limit eigenvalue distribution of a class of non-Hermitian random matrices with statistical application

Arup Bose* (Indian Statistical Institute, Kolkata)

February 26

Session III (9:30–10:35): chaired by Murad Taqqu

9:30–10:00:
Discriminant Analysis based on Binary Time Series
Yuichi Goto* (Waseda University) and Masanobu Taniguchi

10:00–10:35:
Robust causality test of infinite variance processes
Fumiya Akashi* (Waseda University), Masanobu Taniguchi & Anna Clara Monti

10:35–10:45: Coffee break

Session IV (10:45–12:00): chaired by Arup Bose

10:45–11:20:
Tests of high-dimensional mean vectors and its application under the SSE model
Aki Ishii* (Tokyo University of Science), Kazuyoshi Yata (University of Tsukuba), Makoto Aoshima (University of Tsukuba)

11:20–12:00:
Delayed reporting of faults in warranty claims
Richard Arnold, Stefanka Chukova and Yu Hayakawa* (Waseda University)

12:00–13:00: Lunch

Session V (13:00–15:00): chaired by Thomas Mikosh

13:00–13:40:
Three topics in nonparametric statistics
Yoichi Nishiyama* (Waseda University)

13:40–14:20:
The Laguerre expansion of the ruin probability under L¥'evy insurance risks with statistical inference
Yasutaka Shimizu* (Waseda University)

14:20–15:00
Periodicity tests for functional time series
Siegfried Hörmann* (Graz University of Technology)

15:00–15:15: Coffee break

VI Keynote Session (15:15–17:30): chaired by Liudas Giraitis

15:15–16:00:
Testing independence of random elements with the distance covariance
Thomas Mikosch* (University of Copenhagen)

16:00–16:45:
Behavior of the generalized Rosenblatt process at extreme critical exponent values
Murad S. Taqqu* (Boston University)

16:45–17:30:
Functional data analysis in the Banach space of continuous functions
Holger Dette* (Ruhr-Universitaet Bochum)

18:00– Buffet Party

February 27

Session VII (9:00–10:10): chaired by Ngai Hang Chan

9:00–9:35:
Generalized maximum composite likelihood estimators for determinantal point processes
Kou Fujimori* (Waseda University), Sota Sakamoto (Waseda University) and Yasutaka Shimizu (Waseda University)

9:35–10:10:
The asymptotic variance estimators of the correlation estimator between latent processes
Akitoshi Kimura* (Waseda University)

10:10–10:20: Coffee break

Session VIII (10:20–11:40): chaired by Holger Dette

10:20–11:00:
Spectral models for locally stationary processes
Alessandra Luati* (University of Bologna)

11:00–11:40:
Inference for Spatial Trends
Ngai Hang Chan* (Department of Statistics, The Chinese University of Hong Kong)

Session IX (11:50–12:30) Special Session: chaired by Alessandra Luati

11:50–12:30:
The vocal performance of OPERA
Yuya Harada* (Tokyo Univ. of the Arts, Baritone singer) and Yu Tomita

Abstracts

Title: Fréchet-Hoeffding copula bounds for circular data
Author: Hiroaki Ogata* (Tokyo Metropolitan University)
Abstract: We propose the simple extension of Fréchet-Hoeffding copula bounds for circular data. The copula is powerful tool for describing the dependency of random variables. In two dimension, Fréchet-Hoeffding upper (lower) bound indicates the perfect positive (negative) dependence between two random variables. However, for circular random variables, the usual concept of dependency is not accepted because of their periodicity. In this talk, we redefine Fréchet-Hoeffding bounds and consider modified Fréchet and Mardia families of copulas for modelling the dependency of two circular random variables. Simulation studies are also given to see the behavior of the model.
[Back to talk list]

Title: A measure for comparing upper and lower tail probabilities of bivariate distributions
Author: Shogo Kato* (Institute of Statistical Mathematics, Japan), Toshinao Yoshiba (Bank of Japan, Japan), Shinto Eguchi (Institute of Statistical Mathematics, Japan)
Abstract: It is well known that the lack of fit in tails of probability distributions leads to erroneous results in statistical analysis. In this study we propose a measure to compare upper and lower tail probabilities of bivariate distributions. It is seen that the expression for the proposed measure can be simplified if bivariate distribution functions are represented using copulas. With this representation, some properties of the proposed measure are investigated. It is shown that the limit of the proposed measure as a tuning parameter goes to zero can be expressed in a simple form under certain conditions on copulas. A sample analogue of the proposed measure is given and its asymptotic normality is shown. A nonparametric test of symmetry in upper and lower tails based on the sample analogue of the measure is presented. As an illustrative example, the presented measure is applied to stock daily returns of Nikkei225 and S&P500.
[Back to talk list]

Title: Non-Sparse Modeling for High-Dimensional Data
Author: Makoto Aoshima（Tsukuba University）
Abstract: High-dimensional data is often much more non-sparse than you imagine. Non-sparsity about intrinsic/necessary information is useful to improve the accuracy of inference. However, non-sparsity about noise/unnecessary information is troublesome for inference because the intrinsic information is covered with huge noise and the accuracy of inference of

the necessary information is heavily damaged by the huge noise. Aoshima and Yata (Methodol. Comput. Appl., 2018) created a non-sparse modeling approach for high-dimensional classification. In this talk, I emphasize how non-sparse modeling is essential in high-dimensional statistical analysis. My messages are as follows: (1) Analyze the pattern of high-dimensional data. Key tools are the dual space and the geometric representation of high-dimensional data. (2) As for handling huge intrinsic information, powerful tools are new PCAs by Yata and Aoshima: the noise-reduction methodology (J. Multivariate Anal., 2012; Electron. J. Stat., 2016) and the cross-data-matrix methodology (J. Multivariate Anal., 2010). (3) As for handling huge noise, a novel technique is the data transformation by Aoshima and Yata (Statist. Sinica, 2018; Ann. Inst. Statist. Math., 2019). With transformed data, one can create a new statistic which can ensure high accuracy of inferences by using asymptotic normality even under the huge noise. In this talk, I will introduce the new PCAs and non-sparse modeling for inferences on multiclass mean vectors, correlation tests and classification problems.

[Back to talk list]

Title: Predictability, Real Time Estimation, and the Formulation of Unobserved Components Models

Author: Tommaso Proietti* (University of Rome, “Tor Vergata”)

Abstract: The formulation of unobserved components models raises some relevant interpretative issues due to the existence of alternative observationally equivalent specifications differing for the timing of the disturbances and their covariance matrix. In this paper we show that predictability and, when the model is formulated in contemporaneous form, real time estimation, are invariant to the specification of the disturbances covariance matrix. We illustrate these findings with reference to unobserved components models with ARMA(m,m) reduced form, performing the decomposition of the series into an ARMA(m,q) signal, $q < m$, and a noise component. For both the future and contemporaneous forms we characterize the set of covariance structures that are observationally equivalent, showing that only for the canonical model the noise converges in mean square to the innovation process. The set also nests the noninvertible case, which corresponds to a signal-noise decomposition with collinear disturbances, with the noise resulting from applying a Blaschke filter to the innovations. We also discuss the case when the correlation between the signal disturbance and the noise is identified.

Keywords: ARMA models. Steady State Kalman filter. Correlated Components. Blaschke Factors.

[Back to talk list]

Title: Inference on Time Series with Changing Mean and Variance
Author: V Dalla, Liudas Giraitis* (Queen Mary University of London), PM Robinson
Abstract: The paper develops point estimation and large sample statistical inference with respect to a semiparametric model for time series with moving mean and unconditional heteroscedasticity. These two features are modelled nonparametrically, whereas autocorrelations are described by a short memory stationary parametric time series model. We first study the usual least squares estimate of the coefficient of the first-order autoregressive model based on constant but unknown mean and variance. Allowing for both the latter to vary in a general way we establish its probability limit and a central limit theorem for a suitably normed and centred statistic, giving explicit bias and variance formulae. As expected mean variation is the main source of inconsistency and heteroscedasticity the main source of inefficiency, though we discuss circumstances in which the estimate is consistent for, and asymptotically normal about, the autoregressive coefficient, albeit inefficient. We then consider standard Whittle estimates of a more general class of short memory parametric time series model, under otherwise more restrictive conditions.
When the mean is correctly assumed to be constant, estimates that ignore the heteroscedasticity are again found to be consistent for the dependence parameters, and asymptotically normal with parametric rate, and inefficient. Allowing a slowly time-varying mean we resort to trimming out of low frequencies to achieve the same outcome. Returning to finite order autoregression, nonparametric estimates of the varying mean and variance are given asymptotic justification, and forecasting formulae developed. Finite sample properties are studied by a small Monte Carlo simulations, and an empirical example is also included.
[Back to talk list]

Title: Smallest singular value and limit eigenvalue distribution of a class of non-Hermitian random matrices with statistical application
Author: Arup Bose* (Indian Statistical Institute, Kolkata)
Abstract: Suppose X is an $N ¥times n$ complex matrix whose entries are centered, independent, and identically distributed random variables with variance $1/n$ and whose fourth moment is of order ${¥mathcal O}(n^{-2})$. We first consider the non-Hermitian matrix $X A X^* - z$, where A is a deterministic matrix whose smallest and largest singular values are bounded below and above respectively, and $z¥neq 0$ is a complex number. Asymptotic probability bounds for the smallest singular value of this model are obtained in the large dimensional regime where N and n diverge to infinity at the same rate.
Then we consider the special case where $A = J = [¥1_{|i-j| = 1¥mod n}]$ is a circulant matrix. Using the result of the first part, it is shown that the limit eigenvalue distribution of $X J X^*$ exists in the large dimensional regime, and we determine this limit explicitly. A

statistical application of this result devoted towards testing the presence of correlations within a multivariate time series is considered. Assuming that X represents a high domensional complex-valued time series which is observed over a time window of length n, the matrix $X J X^*$ represents the one-step sample autocovariance matrix of this time series. Guided by the result on the limit spectral measure of this matrix, a whiteness test against an MA correlation model on the time series is introduced. Numerical simulations show the excellent performance of this test.
This is joint work with Walid Hachem.
[Back to talk list]

Title: Discriminant Analysis based on Binary Time Series
Author: Yuichi Goto* (Waseda Univ.) and Masanobu Taniguchi
Abstract: Binary time series is the time series converted into 0 and 1. In this talk, we discuss the discriminant analysis and propose a classification method based on binary time series. First, we will show that the misclassification probability tends to zero when the number of observation tends to infinity. Next, we evaluate the asymptotic misclassification probability when the two categories are contiguous. Finally, we show that our classification method based on binary time series has a good robustness when the process is contaminated by an outlier, that is, our classification method is insensitive to the outlier. However, the classical method based on smoothed periodogram is sensitive to the outlier.
[Back to talk list]

Title: Robust causality test of infinite variance processes
Author: Fumiya Akashi* (Waseda University), Masanobu Taniguchi & Anna Clara Monti
Abstract: This talk develops a robust causality test for time series models with infinite variance innovation processes. First, we introduce a measure of dependence for vector nonparametric linear processes and derive the limit distribution of the test statistic by Taniguchi et al. (1996) in infinite variance case. Second, we construct a weighted-version of the generalized empirical likelihood (GEL) test statistic, called the self-weighted GEL statistic in time domain. The limit distribution of the self-weighted GEL test statistic is shown to be a standard chi-squared one regardless of whether the model has finite variance or not. Some simulation experiments illustrate desired finite sample performance of the proposed method.
[Back to talk list]

Title: Tests of high-dimensional mean vectors and its application under the SSE model
Author: Aki Ishii* (Tokyo University of Science), Kazuyoshi Yata (University of Tsukuba),

Makoto Aoshima (University of Tsukuba)

Abstract: We discuss inference problems on high-dimensional mean vectors under the strongly spiked eigenvalue (SSE) model. First, we consider one-sample test. In order to avoid huge noise, we derive a new test statistic by using a data transformation technique. We show that the asymptotic normality can be established for the new test statistic. We give an asymptotic size and power of a new test procedure. We apply the findings to the construction of confidence regions on the mean vector under the SSE model. We further discuss multi-sample problems under the SSE models. Finally, we demonstrate the new test procedure by using actual microarray data sets.

[Back to talk list]

Title: Delayed reporting of faults in warranty claims

Author: Richard Arnold, Stefanka Chukova and Yu Hayakawa* (Waseda University)

Abstract: When a complex system is operated, it may experience multiple faults. If the system is operating under warranty these faults may be claimed for and repaired at zero or minimal cost to the consumer. However, if the faults do not lead to system failure the user may find it inconvenient to claim for each repair as it occurs, and may instead delay making a report or claim until a sufficiently large number of faults has accumulated.

In this talk, we will present a model for the delayed reporting of faults: multiple non-fatal faults are accumulated and then simultaneously reported and repaired. The reporting process is modelled as a stochastic process dependent on the underlying stochastic process generating the faults. The joint distribution of the reporting times and numbers of reported faults is derived. We will also present a few extensions of the above model, which deal with multiple fault types, planned preventative maintenance and customer rush.

[Back to talk list]

Title: Three topics in nonparametric statistics

Author: Yoichi Nishiyama* (Waseda University)

Abstract: The talk consists of some small results for three topics in nonparametric statistics. First, we prove the impossibility of weak convergence of kernel density estimators to any non-degenerate law in $L^2(R^d)$ space. Second, we show asymptotic normality and efficiency of some nonparametric estimators for invariant distribution and density of ergodic diffusion processes based on high-frequency data. Last, we propose a rank-type statistic for nonparametric k-sample and change point problems.

[Back to talk list]

Title: The Laguerre expansion of the ruin probability under L¥'evy insurance risks with

statistical inference
Author: Yasutaka Shimizu* (Waseda University)
Abstract: A statistical inference for ruin probability from a certain discrete sample of the surplus is discussed under a Levy insurance risk. The surplus model is a spectrally negative Levy process with diffusion terms and possibly infinite activity jumps. We assume that discrete records of the surplus every equi-distant time interval and ``large'' jumps, each of which corresponds to a claim size larger than a certain threshold. We consider the Laguerre series expansion of the ruin probability, and give an estimator of the arbitrary partial sum, which is an approximation of the ruin probability in the mean squared sense. We show the consistency and the asymptotic normality of the proposed estimator.
[Back to talk list]

Title: Periodicity tests for functional time series
Author: Siegfried Hörmann* (Graz University of Technology)
Abstract: Periodicity is one of the most important characteristics of time series, and tests for periodicity go back to the very origins of the field.
We consider the two situations where the potential period of a functional time series (FTS) is known and where it is unknown. For both problems we develop fully functional tests and work out the asymptotic distributions. When the period is known we allow for dependent noise and show that our test statistic is equivalent to the functional ANOVA statistic. The limiting distribution has an interesting form and can be written as a sum of independent hypoexponential variables whose parameters are eigenvalues of the spectral density operator of the FTS.
When the period is unknown our test statistic is based on the maximal norm of the functional periodogram over fundamental frequencies. The limiting distribution of this object is rather delicate: it requires a central limit theorem for vectors of functional data, where the number of components increases proportional to the sample size.
The talk is based on joint work with Piotr Kokoszka (Colorado State University) and Gilles Nisol (ULB) and Clément Cerovecki (ULB).
[Back to talk list]

Title: Testing independence of random elements with the distance covariance
Author: Thomas Mikosch* (University of Copenhagen)
Abstract: This is joint work with Herold Dehling (Bochum), Muneya Matsui (Nagoya), Gennady Samorodnitsky (Cornell) and Laleh Tafakori (Melbourne). Distance covariance was introduced by Sz¥'ekely, Rizzo and Bakirov (2007) as a measure of dependence between vectors of possibly distinct dimensions. Since then it has attracted attention in various fields

of statistics and applied probability. The distance covariance of two random vectors X,Y is a weighted L^2 distance between the joint characteristic function of (X,Y) and the product of the characteristic functions of X and Y. It has the desirable property that it is zero if and only if X,Y are independent. This is in contrast to classical measures of dependence such as the correlation between two random variables: zero correlation corresponds to the absence of linear dependence but does not give any information about other kinds of dependencies. We consider the distance covariance for stochastic processes X,Y defined on some interval and having square integrable paths, including L¥'evy processes, fractional Brownian, diffusions, stable processes, and many more. Since distance covariance is defined for vectors we consider discrete approximations to X,Y. We show that sample versions of the discretized distance covariance converge to zero if and only if X,Y are independent. The sample distance covariance is a degenerate V-statistic and, therefore, has rate of convergence which is much faster than the classical $¥sqrt{n}$-rates. This fact also shows nicely in simulation studies for independent X,Y in contrast to dependent X,Y.

[Back to talk list]

Title: Behavior of the generalized Rosenblatt process at extreme critical exponent values
Author: Murad S. Taqqu* (Boston University)
Abstract: The generalized Rosenblatt process is obtained by replacing the single critical exponent characterizing the Rosenblatt process by two different exponents living in the interior of a triangular region.

What happens to that generalized Rosenblatt process as these critical exponents approach the boundaries of the triangle? We show that on each of the two symmetric boundaries, the limit is non-Gaussian. On the third boundary, the limit is Brownian motion. The rates of convergence to these boundaries are also given. The situation is particularly delicate as one approaches the corners of the triangle, because the limit process will depend on how these corners are approached. All limits are in the sense of weak convergence in C[0,1].

This is joint work with Shuyang Bai.

[Back to talk list]

Title: Functional data analysis in the Banach space of continuous functions
Author: Holger Dette* (Ruhr-Universitaet Bochum)
Abstract: Functional data analysis is typically conducted within the $L2$-Hilbert space framework. There is by now a fully developed statistical toolbox allowing for the principled application of the functional data machinery to real-world problems, often based on dimension reduction techniques such as functional principal component analysis. At the

same time, there have recently been a number of publications that sidestep dimension reduction steps and focus on a fully functional $L2$-methodology. This paper goes one step further and develops data analysis methodology for functional time series in the space of all continuous functions. The work is motivated by the fact that objects with rather different shapes may still have a small $L2$-distance and are therefore identified as similar when using an $L2$-metric. However, in applications it is often desirable to use metrics reflecting the visualization of the curves in the statistical analysis. The methodological contributions are focused on developing two-sample and change-point tests as well as confidence bands, as these procedures appear do be conducive to the proposed setting. Particular interest is put on relevant differences; that is, on not trying to test for exact equality, but rather for pre-specified deviations under the null hypothesis.
The procedures are justified through large-sample theory. To ensure practicability, non-standard bootstrap procedures are developed and investigated addressing particular features that arise in the problem of testing relevant hypotheses. The finite sample properties are explored through a simulation study and an application to annual temperature profiles.
[Back to talk list]

Title: Generalized maximum composite likelihood estimators for determinantal point processes
Author: Kou Fujimori* (Waseda University), Sota Sakamoto (Waseda University) and Yasutaka Shimizu (Waseda University)
Abstract: The maximum composite likelihood estimator for stationary and isotropic parametric models of determinantal point processes will be discussed. Since the joint intensities of these point processes are given by determinant of positive definite kernels, we have the explicit form of the joint intensities for every order. This fact enables us to consider the generalized maximum composite likelihood estimator for every order.
[Back to talk list]

Title: The asymptotic variance estimators of the correlation estimator between latent processes
Author: Akitoshi Kimura* (Waseda University)
Abstract: In this talk, we treat a model in which the finite variation part of a two-dimensional semi-martingale is expressed by time-integration of latent processes. We propose a correlation estimator between the latent processes and show its consistency and asymptotic mixed normality. Moreover, we propose two types of estimators for asymptotic variance of the correlation estimator and show their consistency in a high frequency setting. Our model includes doubly stochastic Poisson processes whose intensity processes are correlated It¥^o

processes. We compare our asymptotic variance estimators based on the simulation of the doubly stochastic Poisson processes.
[Back to talk list]

Title: Spectral models for locally stationary processes
Author: Alessandra Luati* (University of Bologna)
Abstract: A class of models for the time-varying spectrum of a locally stationary process is introduced. The models are specified in the frequency domain and the class depends on a power param- eter that applies to the spectrum so that it can be locally represented by a finite Fourier polynomial. The coefficients of the polynomial have an interpretation as generalised auto- covariances whose smooth time variation is determined according to a linear combination of logistic transition functions of the time index. Estimation is carried out in the frequency domain based on the generalised Whittle likelihood. Parametric inference is developed. (Joint work with Tommaso Proietti and Stefano Grassi)
[Back to talk list]

Title: Inference for Spatial Trends
Author: Ngai Hang Chan* (Department of Statistics, The Chinese University of Hong Kong)
Abstract: Non-stationary spatial models are widely applicable in diverse disciplines, ranging from bio-medical sciences to geophysical studies. In many of theses applications, testing for structural changes in the trend and testing the specific form of the trend are highly relevant. A novel statistics based on a discrepancy measure over small regions is proposed in this paper. Such a measure can be used to construct tests for structural trends and to identify change boundaries of the trends. By virtue of the m-dependence approximation of a stationary random field, asymptotic properties and limit distributions of these tests are established. The method is illustrated by simulations and data analysis.
[Back to talk list]

Title: The vocal performance of OPERA
Author: Yuya Harada* (Tokyo Univ. of the Arts, Baritone singer) and Yu Tomita
Abstract: (i) Play multiple operas with different age of different performance expressions. (ii)Listen to the performance and consider whether statistical consideration can be done.
[Back to talk list]