Risk and Statistics, 3rd Tohoku-ISM-UUlm Joint Workshop

Program and Abstracts

October 12–14, 2022

Tohoku Forum for Creativity

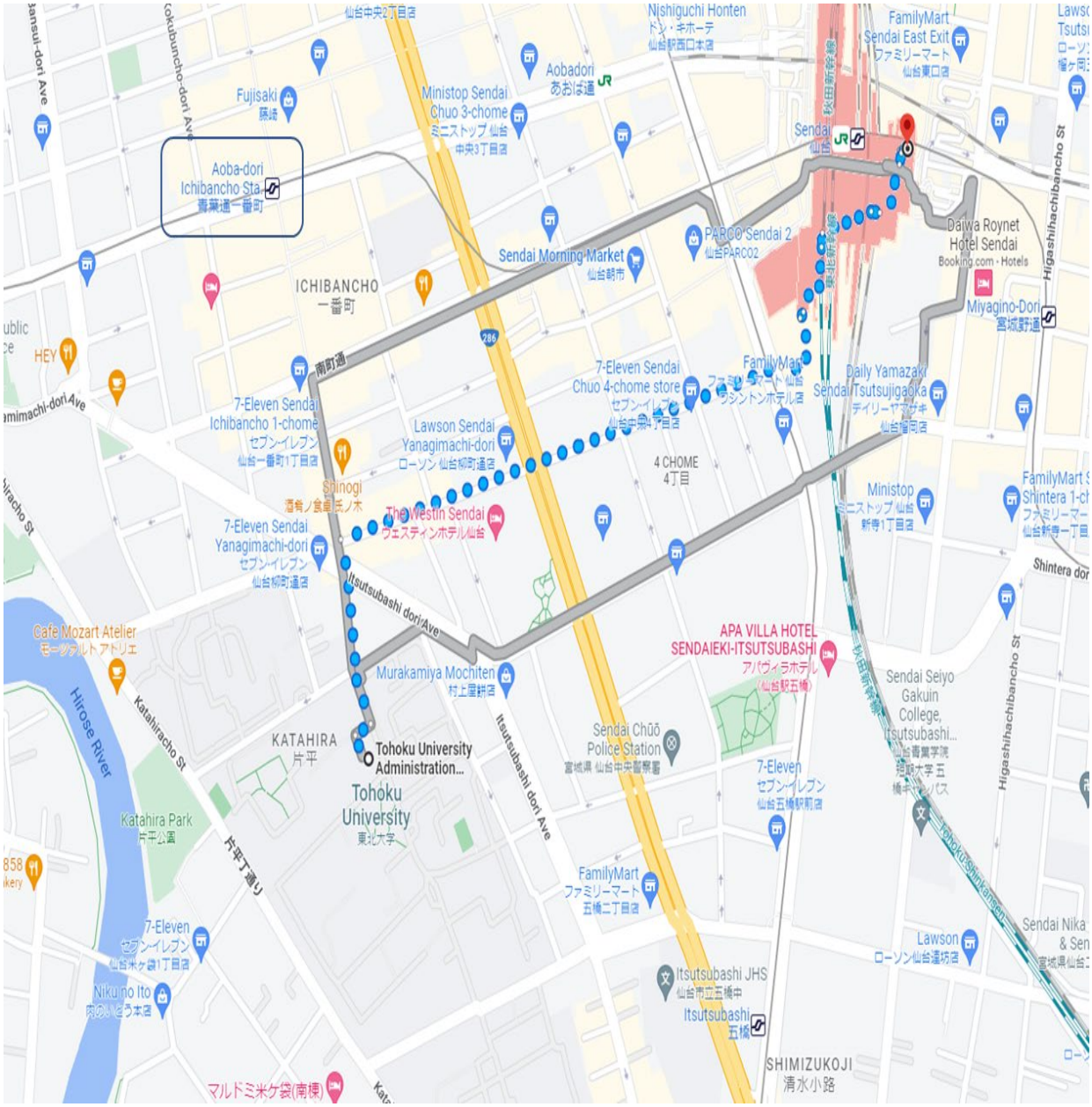
知の館　３F

Graduate School of Economics and Management

Tohoku University

Banquet place information

restaurant on 3F at Hotel Metropolitan East, 18:30-20:00, Oct. 13th



Visit the hotel located at East Exit in Sendai Station. It takes 20 mins on foot from the venue. You can use “Aobadori-Ichibancho station” in TOZAI subway line as the nearest station from the workshop venue.

Program

October 12, 2022

9:30 **Li, Yingxing (Tohoku Univ.)**
Product Embedding with Receipt context for Disaggregated Sales Data

10:00 **Nomura, Shunichi (Waseda Univ.)**
Hierarchical Topic Model for Tensor Data and Extraction of Weekly and Daily Activity Patterns

10:30 **Wu, Stephen (ISM)**
Performance-Based Earthquake Early Warning for Regional Seismic Risk Mitigation

Break

11:15 **Yoshiba, Toshinao (Tokyo Metropolitan Univ.)**
Portfolio Risk Valuation Using Asymmetric Copulas

11:45 **Ogihara, Teppei (Univ. Tokyo)**
Local Asymptotic Normality for Jump-Diffusion Processes with Discrete Observations

Lunch

13:30 **Lindner, Alexander (Univ. Ulm)**
Quasi-Infinitely Divisible Distributions

14:00 **Matsui, Muneya (Nanzan Univ.)**
Subexponentialiy of Densities of Infinitely Divisible Distributions

Break

14:45 **Berger, David (Univ. Dresden)**
Existence and Distributional Properties of Solutions for SPDEs Driven by Lévy White Noise

15:15 **Kurisu, Daisuke (Yokohama National Univ.)**
Nonparametric Regression for Locally Stationary Random Fields on R^d

Break

16:00 **Zhang, Fangyuan (EURECOM)**

Intergenerational Risk Sharing in Defined Contribution Pension Systems: Analysis with Bayesian Optimization

16:30 **Klicnarová, Jana (Univ. South Bohemia)**

On Modified Interdirections

October 13, 2022

AI Yotta Special Session

9:30 **Shioiri, Satoshi (Tohoku Univ.)**

Extracting Valuable Information based on Human Behavior

10:00 **Sakai, Nobuyuki (Tohoku Univ.)**

Measurement and Statistics of Taste and Deliciousness

10:30 **Ishigaki, Tsukasa (Tohoku Univ.)**

Recommendation Systems with Network Structure and Big Data

Break

11:15 **Chen, An (Univ. Ulm)**

Non-Concave Optimization Under Risk Constraints

11:45 **Araki, Yuko (Tohoku Univ.)**

Statistical Modeling for the Nonlinear Structure of Quantiles in Hierarchical Data via Regularization

Lunch

13:30 **Spodarev, Evgeny (Univ. Ulm)**

Prediction of Heavy-Tailed Random Functions

14:00 **Kuriki, Satoshi (ISM)**

Expected Euler Characteristic Heuristic for Smooth Gaussian Random Fields with Inhomogeneous Marginals

Break

14:45 **Kanagawa, Motonobu (EURECOM)**
Counterfactual Mean Embeddings
15:15 **Koike, Yuta (Univ. Tokyo)**
Central Limit Theorems in High-Dimensions: Recent Developments

Break

16:00 **Kariya, Takeaki (Nagoya Univ. of Commerce & Business)**
A Modelling Framework for Regression with Collinearity
16:30 **Miura, Ryozo (Hitotsubashi Univ.)**
Occupation Time, Quantiles and Rank on Vasicek Process with Applications to Exotic Options

Banquet

October 14, 2022

9:30 **Sawada, Masayuki (Hitotsubashi Univ.)**
Spatial Regression Discontinuity Designs
10:00 **Ishihara, Takuya (Tohoku Univ.)**
Shrinkage Methods for Treatment Choice
10:30 **Ko, Stanley Iat-Meng (Tohoku Univ.)**
Forecasting Stock Returns with Conditional Quantile-level Dependence

Break

11:15 **Robinson, Peter M (LSE)**
Issues in Spatial Processes with Long Range Dependence
11:45 **Yajima, Yoshihiro (Univ. Tokyo)**
On Estimation of Fractional Brownian Fields and Sheets
12:15 **Taniguchi, Masanobu (Waseda Univ.)**
Joint Circular Distributions in View of Higher Order Spectra of Time Series and Copula

Product Embedding with Receipt context for Disaggregated Sales Data

Yinxing Li and Nobuhiko Terui

Marketing data are expanding in several modes nowadays, as the number of variables explaining customer behavior has greatly increased, a Product Embedding for Large-Scale Disaggregated Sales Data Yinxing Li and Nobuhiko Terui Marketing data are expanding in several modes nowadays, as the number of variables explaining customer behavior has greatly increased, and automated data collection in the store has also led to the recording of customer choice decisions from large sample sizes. Thus, high-dimensional models have recently gained considerable importance in several areas, including marketing. Although some models, such as Prod2Vec, involve various marketing variables such as price and customer demographic data, the role of the variables in forecasting is still not discussed. In light of the limitations mentioned above, our study not only aims to propose a model with better forecasting precision but also to reveal how customer demographics affect customer behavior. we propose a Bayesian Word2Vec based framework that incorporates marketing variables and environment by considering following situations in our model. When considering the market basket, our study incorporates the receipt vector into the model as the prior information of each purchased product in a basket, which means a preferred purchasing pattern for a certain shopping. It assumes that the customer will consider the whole purchasing context before choosing a product. We also assume a state space model for the receipt vector through the trips for each customer. We use the weekday, promotion information as the data for the state space model for higher interpretability of the prior structure such as the purchasing scenarios of each customer. Besides, we consider the purchasing probability of a certain product conditional on an existing market basket is influenced by the following three factors - 1) The compatibility with the marketing basket, which is represented by the inner product of product vectors, 2) customer utility for the product, which incorporate the customer heterogeneity structure, and 3) Thinking ahead algorithm, which represents one-step ahead forecasting before purchasing the product. Our proposed model contributes both to higher precision for forecasting by incorporating the marketing environment and customer heterogeneity into the model, and better interpretability. We use receipt data from a retailer for our empirical analysis, containing the information of customer demographic, promotion and other marketing information. We show not only the effectiveness of marketing environment for the forecasting by using the Hit Rate@K for the hold-out sample comparing to the several benchmark models , but also the high interpretability of our proposed model in the empirical studynd automated data collection in the store has also led to the recording of customer choice decisions from large sample sizes. Thus, high-dimensional models have recently gained considerable importance in several areas, including marketing. Although some models, such as Prod2Vec, involve various marketing variables such as price and customer demographic data, the role of the variables in forecasting is still not discussed. In light of the limitations mentioned above, our study not only aims to propose a model with better forecasting precision but also to reveal how customer demographics affect customer behavior. we propose a Bayesian Word2Vec based framework that incorporates marketing variables and environment by considering following situations in our model.

When considering the market basket, our study incorporates the **receipt vector** into the model as the prior information of each purchased product in a basket, which means a preferred purchasing pattern for a certain shopping. It assumes that the customer will consider the whole purchasing context before choosing a product. We also assume a **state space model** for the receipt vector through the trips for each customer. We use the weekday, promotion information as the data for the state space model for higher interpretability of the prior structure such as the purchasing scenarios of each customer.

Besides, we consider the purchasing probability of a certain product conditional on an existing market basket is influenced by the following three factors - 1) **The compatibility with the marketing basket**, which is represented by the inner product of product vectors, 2) **customer utility** for the product, which incorporate the customer heterogeneity structure, and 3) **Thinking ahead** algorithm, which represents one-step ahead forecasting before purchasing the product.

Our proposed model contributes both to higher precision for forecasting by incorporating the marketing environment and customer heterogeneity into the model, and better interpretability. We use receipt data from a retailer for our empirical analysis, containing the information of customer demographic, promotion and other marketing information. We show not only the effectiveness of marketing environment for the forecasting by using the Hit Rate@K for the hold-out sample comparing to the several benchmark models , but also the high interpretability of our proposed model in the empirical study.

Hierarchical Topic Model for Tensor Data and Extraction of Weekly and Daily Activity Patterns

Shunichi Nomura (Waseda University)

Latent Dirichlet allocation (LDA) is a popular topic model for extracting common patterns from discrete datasets. It is extended to the pachinko allocation model (PAM) with a hierarchical topic structure. This study presents a combination meal allocation (CMA) model, which is a further enhanced topic model from the PAM that has both hierarchical categories and hierarchical topics. We consider count datasets in multiway arrays, i.e., tensors, and introduce a set of topics to each mode of the tensors. The topics in each mode are interpreted as patterns in the topics and categories in the next mode. Despite there being a vast number of combinations in multilevel categories, our model provides simple and interpretable patterns by sharing the topics in each mode. Latent topics and their membership are estimated using Markov chain Monte Carlo (MCMC) methods.

We apply the proposed model to step-count data recorded by activity monitors to extract some common activity patterns exhibited by the users. Our model identifies four daily patterns of ambulatory activities (commuting, daytime, nighttime, and early-bird activities) as sub-topics, and six weekly activity patterns as super-topics. We also investigate how the amount of activity in each pattern dynamically affects body weight changes.

This is joint work with Michiko Watanabe at Rissho University and Yuko Oguma at Keio University.

Performance-Based Earthquake Early Warning for Regional Seismic Risk Mitigation

Stephen Wu, The Institute of Statistical Mathematics, Japan
Keisuke Yano, The Institute of Statistical Mathematics, Japan
Masumi Yamada, DPRI, Kyoto University, Japan
Koji Tamaribuchi, Meteorological Research Institute, Japan
Sarah Minson, US Geological Survey, USA
Elizabeth S. Cochran, US Geological Survey, USA

Talk summary:

Earthquake early warning system (EEWS) aims to provide few seconds to a minute of shaking alert before strong shaking of an on-going seismic event reaches a target site. Developing a reliable EEWS typically involves rapidly solving a highly ill-posed seismic source inversion problem using only few seconds of seismic wave data. Such a problem is inherently highly uncertain, which means that false and missed alarm is unavoidable. Therefore, a performance-based approach for the system design is needed in order to develop a practically useful EEWS. Here, we propose a fully probabilistic framework that covers three important aspects of EEWS – real-time ground motion prediction, spatially correlated warning criteria, and temporally correlated warning criteria. In each aspect, exploiting relevant prior knowledge can help reducing the uncertainty of warnings, thus, developing a more robust EEWS. This talk will mainly focus on discussion of the ideology and design strategy of such practically useful EEWS from a statistical perspective.

Portfolio risk valuation using asymmetric copulas*

Toshinao Yoshiba†

October 12, 2022

The multivariate Student-t copula is frequently used in financial portfolio risk management and other statistical areas when there is tail dependence in the data. It often is a good-fitting copula but can be improved on when there is tail asymmetry. We propose to use Generalized Hyperbolic (GH) and Azzalini–Capitanio (AC) skew-t copulas to incorporate asymmetric tail dependence of risk factors using the numerical implementation for maximum likelihood estimation proposed in Yoshiba (2018). The GH skew-t copula was proposed by Demarta and McNeil (2005).

The GH skew-t random vector $\boldsymbol{X}$ is constructed by using the Gamma random variable $V \sim G(\nu/2, \nu/2)$ and the d-variate normal random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^\top \sim N_d(\boldsymbol{0}, \Psi)$ with the correlation matrix Ψ as:

$$\boldsymbol{X} = \boldsymbol{\gamma} V^{-1} + \frac{\boldsymbol{Z}}{\sqrt{V}}, \tag{1}$$

where $\boldsymbol{\gamma} \in \mathbb{R}^d$ is the skewness paramemter vector. On the other hand, AC skew-t copula implicit in Azzalini and Capitanio (2003) multivariate skew-t distribution was proposed by Joe (2006). The AC skew-t random vector $\boldsymbol{X}$ is given as:

$$\boldsymbol{X} = \frac{\boldsymbol{Y}}{\sqrt{V}}, \tag{2}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ is the skew Normal random vector proposed by Azzalini and Dalla Valle (1996), which is constructed as:

$$Y_j = \delta_j |Z_0| + \sqrt{1 - \delta_j^2} Z_j, \ j = 1, \ldots, d, \quad Z_0 \sim N(0,1), \tag{3}$$

where $\boldsymbol{\delta} \equiv (\delta_1, \ldots, \delta_d)^\top \in (-1, 1)^d$ is the skewness paramemter vector and $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^\top \sim N_d(\boldsymbol{0}, \Psi)$ as the GH skew-t random vector.

We compare the parameters of AC skew-t, GH skew-t, Student-t, Normal copulas using Akaike and Bayesian information criteria for the two groups of daily stock returns in stress and in peacetime. For skew-t copulas, we assume equi-skewness settings, that is, $\delta_1 = \cdots = \delta_d = \delta$ for AC and $\gamma_1 = \cdots = \gamma_d = \gamma$ for GH. Each group is given by the equally weighted portfolio which consists of three indices from TOPIX33 Sector Indices. The first portfolio consists of financial sectors portfolio with high correlation including bank, insurance, and securities sectors. The second consists of bank, air transportation, electricity sectors with low correlation.

In the empirical analyses, we show the skewness δ, γ of skew-t copulas are significantly negative for both the unfiltered returns and the filtered returns by EGARCH(1,1). We also show that the more sophisticated the marginal model, the more significant the negativity. The tendency holds both in stress and in peacetime.

With validating the skewness, we investigate the behavior of the value-at-risk and expected shortfall of the two types of stock portfolio by employing several backtesting methods. The backtesting methods include the unconditional coverage test of Kupiec (1995), the independence test, and the conditional coverage test of Christoffersen (1998) for value-at-risk, and the discrepancy measurement of Embrechts, Kaufmann and Patie (2005) for expected shortfall.

*This research is supported by JSPS KAKENHI Grant Number JP21K01581.

†Tokyo Metropolitan University / Institute of Statistial Mathematics (E-mail: tyoshiba@tmu.ac.jp)

The unconditional coverage tests for the value-at-risk with unfiltered return are rejected in stress for both the financial sectors portfolio and the low-correlated portfolio. On the other hand, the tests with EGARCH filtered return are not rejected.

The independence tests for the value-at-risk with unfiltered return are rejected in peacetime for both the financial sectors portfolio and the low-correlated portfolio. On the other hand, the tests with EGARCH filtered return are not rejected.

The discrepancies of expected shortfall by skew-t copulas are small especially in stress. Even in peacetime, the discrepancies of expected shortfall by skew-t or Student-t copulas are small.

We conclude that the value-at-risk with high confidence level and the expected shortfall of a stock portfolio are well captured by skew-t copulas with filtered return.

References

Azzalini, A. and Dalla Valle, A. (1996) "The Multivariate Skew-Normal Distribution," *Biometrika*, **83**(4), 715–726.

Azzalini, A. and Capitanio, A. (2003) "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society Series B*, **65**(2), 367–389.

Christoffersen, P. F. (1998) "Evaluating Interval Forecasts," *International Economic Review*, **39**(4), 841–862.

Demarta, S. and McNeil, A. J. (2005) "The t Copula and Related Copulas," *International Statistical Review*, **73**(1), 111–129.

Embrechts, P., Kaufmann, R., and Patie, P. (2005) "Strategic Long-Term Financial Risks: Single Risk Factors," *Computational Optimization and Applications*, **32**(1), 61–90.

Joe, H. (2006) "Discussion of "Copulas: Tales and facts", by Thomas Mikosch," *Extremes*, **9**(1), 37–41.

Kupiec, P. H. (1995) "Techniques for Verifying the Accuracy of Risk Measurement Models," *The Journal of Derivatives*, **3**(2), 73–84.

Yoshiba, T. (2018) "Maximum likelihood estimation of skew-t copulas with its applications to stock returns," *Journal of Statistical Computation and Simulations*, **88**(13), 2489–2506.

Local asymptotic normality for jump-diffusion processes with discrete observations

Teppei Ogihara (University of Tokyo) and Yuma Uehara (Kansai University)

We study local asymptotic normality (LAN) of jump-diffusion processes with discrete observations. Local asymptotic normality (LAN) is an important property in asymptotic statistical theory because it enables us to discuss asymptotic efficiency of parameter estimators for parametric models. Let X_t^α be a parametrized m-dimensional stochastic process satisfying

$$X_t^\alpha = x_0 + \int_0^t a(X_s^\alpha, \theta)ds + \int_0^t b(X_s^\alpha, \sigma)dW_s + \sum_{j=1}^{N_t} Y_i, \tag{1}$$

where $a, b : \mathbb{R}^m$- and $\mathbb{R}^m \otimes \mathbb{R}^m$-valued continuous functions, N_t is a Poisson process with intensity $\lambda(\theta)$, W_t: an m-dim standard Brownian motion, $(Y_i)_{i=1}^\infty$ is an i.i.d. sequence with the density function F_θ, and $\alpha = (\sigma, \theta) \in \mathbb{R}^d$ is a parameter to be estimated. Let $\alpha_0 = (\sigma_0, \theta_0)$ be the true value of the parameter, and let $X_t = X_t^{\alpha_0}$. We observe $(X_{t_k})_{k=0}^n$, where $t_k = kh_n$, $h_n \to 0$ and $nh_n \to \infty$.

Let $\{P_{\alpha,n}\}_{\alpha,n}$ be a family of probability measures generated by the observation $(X_{t_k}^\alpha)_{k=0}^n$. $\{P_{\alpha,n}\}_{\alpha,n}$ is said to satisfy local asymptotic mixed normality (LAMN) at $\alpha = \alpha_0$ if there exist $d \times d$ random, positive definite matrices Γ_n and Γ, d-dim random vectors $\mathcal{N}_n$ and $\mathcal{N}$, and a positive definite matrix ϵ_n such that

$$\log \frac{dP_{\alpha_0+\epsilon_n u,n}}{dP_{\alpha_0,n}} - \left(u^\top \sqrt{\Gamma_n}\mathcal{N}_n - \frac{1}{2}u^\top \Gamma_n u \right) \to 0$$

in $P_{\alpha_0,n}$-probability, $\mathcal{N} \sim N(0, I_d)$, $\mathcal{N}$ and Γ are independent, and $\mathcal{L}((\mathcal{N}_n, \Gamma_n)|P_{\alpha_0,n}) \to \mathcal{L}((\mathcal{N}, \Gamma))$ as $n \to \infty$ for any $u \in \mathbb{R}^d$. If Γ is nonrandom, $\{P_{\alpha,n}\}_{\alpha,n}$ satisfies the LAN property. Under LAMN, any regular estimator $\{V_n\}$ satisfies the inequality

$$\liminf_{n\to\infty} E_{\alpha_0,n}[l(|\epsilon_n^{-1}(V_n - \alpha_0)|)] \geq E[l(|\Gamma^{-1/2}\mathcal{N}|)]$$

for any increasing function $l : [0, \infty) \to \mathbb{R}$ with $l(0) = 0$ (Proposition 3 in Jeganathan [3]). An estimator which attains the lower bound of the above inequality is called asymptotically efficient.

To show the LAN property, we need to specify the limit of $\log(dP_{\alpha_0+\epsilon_n u,n}/dP_{\alpha_0,n})$. It is difficult to deal with transition density ratio for two different jump-diffusion processes. In the proof of the LA(M)N properyt for diffusion processes in Gobet [1, 2], he used Aronson estimates

$$C_1 G_1(x, y) \leq p_k(x, y) \leq C_2 G_2(x, y)$$

for transition density p_k of the diffusion process to control transition density ratios, where G_1, G_2 are Gaussian density functions and C_1, C_2 are positive constants. However, it is difficult to obtain Aronson-type estimates for jump-diffusion processes.

Instead, we use a scheme with the so-called L^2 regularity condition developed in Jeganathan [3]. Let $p_{k,\alpha}(x_{k-1}, x_k)$ be the transition density function. We assume that $p_{k,\alpha} \in C^2(\Theta)$ and the zero points of $p_{k,\alpha}$ do not depend on the parameter α. A simple version of the L^2 regularity condition is the condition of the following convergence:

$$\sum_{k=1}^n E_{\alpha_0}\left[\int \left(\sqrt{p_{k,\alpha_u}} - \sqrt{p_{k,\alpha_0}} - \frac{u^\top \epsilon_n \partial_\theta p_{k,\alpha_0}}{2\sqrt{p_{k,\alpha_0}}} \right)^2 (x_{k-1}, x_k)dx_k \right] \to 0,$$

where $\alpha_u = \alpha_0 + \epsilon_n u$. The integrand in the left-hand side of the above equation can be rewritten as follows.

$$\int_0^1 (1-s) \int \left\{ u^\top \epsilon_n \left(\frac{\partial_\alpha^2 p_{k,\alpha_{su}}}{2p_{k,\alpha_{su}}} - \frac{\partial_\alpha p_{k,\alpha_{su}} \partial_\alpha p_{k,\alpha_{su}}^\top}{4(p_{k,\alpha_{su}})^2} \right) \epsilon_n u \right\}^2 p_{k,\alpha_{su}} dx_k ds.$$

In the last expression, only the transition density at α_{su} appears, and hence we do not need Aronson-type estimates for transition density functions. Jeganathan [3] showed LAMN for Markov processes under the L^2 regularity condition and some conditions for $\partial_\alpha^l \log p_{k,\alpha_0}$ $(l \in \{1,2\})$. The original scheme cannot be applied for jump-diffusion processes because of their fat-tailed behaviors. Therefore, we extend the scheme so that it can be applied to jump-diffusion processes.

Another problem to show the LAN property for jump-diffusion processes is that the transition probability for no jump is quite different from that for the presence of jumps. This fact makes it difficult to identify the asymptotic behavior of the likelihood function. To deal with this problem, we approximate the original likelihood function by using a thresholding likelihood function that detects the existence of jumps. By using these techniques, we obtain the LAN property for jump-diffusion processes. The matrices ϵ_n and Γ in the definition of the LAMN property is given by

$$\epsilon_n = \begin{pmatrix} n^{-1/2} I_{d_1} & 0 \\ 0 & (nh_n)^{-1/2} I_{d_2} \end{pmatrix} \quad \text{and} \quad \Gamma = \begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{pmatrix},$$

where $\pi(dx)$ is the limit distribution of X_t, $f_\theta(z) = \lambda(\theta) F_\theta(z)$, $S(x,\sigma) = bb^\top(x,\sigma)$,

$$[\Gamma_1]_{ij} = \frac{1}{2} \int \mathrm{tr}(\partial_{\sigma_i} S S^{-1} \partial_{\sigma_j} S S^{-1})(x, \sigma_0) d\pi(x),$$

$$[\Gamma_2]_{ij} = \int (\partial_{\theta_i} a)^\top S^{-1} (\partial_{\theta_j} a)(x, \alpha_0) d\pi(x) + \int \frac{\partial_{\theta_i} f_{\theta_0} \partial_{\theta_j} f_{\theta_0}}{f_{\theta_0}}(z) dz.$$

These ϵ_n and Γ are the same as the ones associated with the quasi-maximum-likelihood and Bayes-type estimators proposed in Shimizu and Yoshida [5] and Ogihara and Yoshida [4], and then we can show these estimators are asymptotically efficient in this model.

参考文献

[1] E. Gobet. Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli*, 7(6):899–912, 2001.

[2] E. Gobet. LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5):711–737, 2002.

[3] P. Jeganathan. On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Ser. A*, 44(2):173–212, 1982.

[4] T. Ogihara and N. Yoshida. Quasi-likelihood analysis for the stochastic differential equation with jumps. *Stat. Inference Stoch. Process.*, 14(3):189–229, 2011.

[5] Y. Shimizu and N. Yoshida. Estimation of parameters for diffusion processes with jumps from discrete observations. *Stat. Inference Stoch. Process.*, 9(3):227–277, 2006.

Quasi-infinitely divisible distributions

Alexander Lindner

Infinitely divisible probability distributions on $\mathbb{R}^d$ constitute a well-studied class of probability distributions. By definition, a probability distribution μ on $\mathbb{R}^d$ is *infinitely divisible* if and only if it has convolution roots of all orders, i.e. if for every $n \in \mathbb{N}$ there exists some probability measure μ_n on $\mathbb{R}^d$ such that μ_n^{*n}, the n-fold convolution of μ_n with itself, is equal to μ. By the famous Lévy-Khintchine formula, a probability measure μ on $\mathbb{R}^d$ is infinitely divisible if and only its characteristic function $\widehat{\mu}$, defined by $\widehat{\mu}(z) = \int_{\mathbb{R}^d} e^{izx}\,\mu(dx)$ for $z \in \mathbb{R}^d$, has a Lévy–Khintchine representation, i.e. if and only if there are $\gamma \in \mathbb{R}^d$, a symmetric non-negative definite matrix $A \in \mathbb{R}^{d\times d}$ and a Lévy measure ν on $\mathbb{R}^d$ such that

$$(1) \quad \widehat{\mu}(z) = \exp\left\{ i\langle z,\gamma\rangle - \frac{1}{2}\langle z, Az\rangle + \int_{\mathbb{R}^d} (e^{i\langle z,x\rangle} - 1 - i\langle z,x\rangle)\,\nu(dx)\right\} \quad \forall\, z \in \mathbb{R}^d;$$

recall that a *Lévy measure* is a measure on $\mathbb{R}^d$ such that $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^d}(1 \wedge |x|^2)\,\nu(dx) < \infty$. The triplet (A,ν,γ) is called the characteristic triplet of ν and is known to be unique; further, to any given triplet (A,ν,γ) there exists an infinitely divisible distribution having this triplet as characteristic triplet. The importance of infinitely divisible distributions lies in their one-to-one correspondence to Lévy processes (in law), i.e. stochastic processes with independent and increments that start in 0 and have càdlàg paths. Properties and examples of infinitely divisible distributions are well-studied. Normal distributions, compound Poisson distributions, stable distributions, geometric distributions and many other distributions are known to be infinitely divisible. On the other hand, probability distributions with compact support or not infinitely divisible unless they are Dirac measures, and by the Lévy–Khintchine representation, the characteristic function of an infinitely divisible distribution must be zero-free.

The goal of the present talk is to extend the class of infinitely divisible distributions to a larger class. Inspired by the Lévy–Khintchine formula, we call a probability distribution μ on $\mathbb{R}^d$ *quasi-infinitely divisible*, if its characteristic function has a Lévy–Khintchine representation as in (1), however with a 'signed Lévy measure' and some symmetric, but not necessarily non-negative definite, matrix $A \in \mathbb{R}^{d\times d}$. Here, by a signed Lévy measure we mean a 'signed measure' that can be written as a difference of two Lévy measures, which will henceforth be called a *quasi-Lévy measure*. The

corresponding triplet (A, ν, γ) can again be shown to be unique and is again called the *characteristic triplet* of the quasi-infinitely divisible distribution.

We will study quasi-infinitely divisible distributions to some extent and present examples. In particular, we will see that not every triplet gives rise to a probability distribution, and even more, that the matrix A must be positive semi-definite. The quasi-Lévy measure however can be truly signed, although not all quasi-Lévy measures can appear. We then study certain subclasses of distributions. E.g., it can be shown that a probability distribution that has an atom of mass greater than $1/2$ must be quasi-infinitely divisible. Further, we study probability distributions on the grid $\mathbb{Z}^d$ and obtain that such a distribution is quasi-infinitely divisible if and only if its characteristic function is zero-free. We deduce that the class of quasi-infinitely divisible distributions is dense with respect to weak convergence in dimension $d = 1$, but fails to be dense in higher dimensions. We also mention some further results regarding characterisations of quasi-infinite divisibility for subclasses of probability distributions, such as discrete distributions on $\mathbb{R}$ or probability distributions that have a non-trivial absolutely continuous part and non-trivial discrete part but no continuous singular part, as discussed in [1, 2, 3, 5]. The talk is based on [4, 6, 7], which are (joint) works with/of Berger, Kutlu, Pan and Sato.

References

[1] Alexeev, I.A. and Khartov, A.A. (2021) Spectral representations of characteristic functions of discrete probability laws. *Bernoulli*, to appear.

[2] Berger, D. (2019) On quasi-infinitely divisible distribuitons with point mass. *Math. Nachr.* **292**, 1674–1684.

[3] Berger, D. and Kutlu, M. (2022) Quasi-infinite divisibility of a class of distributions with discrete part. *Proc. Amer. Math. Soc.*, to appear.

[4] Berger, D. and Lindner, A. (2022) A Cramér–Wold device for infinite divisibility of $\mathbb{Z}^d$-valued distributions. *Bernoulli* **28**, 1276–1283.

[5] Khartov, A.A. (2022) A criterion of quasi-infinite divisibility for discrete laws. *Stat. Probab. Lett.* **185**, Paper No. 109436, 4 pp.

[6] Kutlu, M. (2021) On a denseness result for quasi-infinitely divisible distributions. *Stat. Probab. Lett.* **176**, Paper No. 109139, 4 pp.

[7] Lindner, A., Pan, L., and Sato, K. (2018) On quasi-infinitely divisible distributions. *Trans. Amer. Math. Soc.* **370**, 8483-8520.

Subexponentiality of densities of infinitely divisible distributions

Nanzan University Muneya Matsui

Abstract

We show the equivalence of three properties for an infinitely divisible distribution: the subexponentiality of the density, the subexponentiality of the density of its Lévy measure and the tail equivalence between the density and its Lévy measure density, under monotonic-type assumptions on the Lévy measure density. The key assumption is that tail of the Lévy measure density is asymptotic to a non-increasing function or is almost decreasing. Our conditions are natural and cover a rather wide class of infinitely divisible distributions. Several significant properties for analyzing the subexponentiality of densities have been derived such as closure properties of [convolution, convolution roots and asymptotic equivalence] and the factorization property. Moreover, we illustrate that the results are applicable for developing the statistical inference of subexponential infinitely divisible distributions which are absolutely continuous.

Introduction

Let f, g be probability density functions on $\mathbb{R}$ and denote by $f * g$ the convolution of f and g:

$$f * g(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy,$$

and denote by f^{*n} the nth convolutions with itself. Throughout the paper, for functions $\alpha, \beta : \mathbb{R} \to \mathbb{R}_+$, $\alpha(x) \sim \beta(x)$ means that $\lim_{x\to\infty} \alpha(x)/\beta(x) \to 1$. We study the following characteristics for densities.

Definition 0.1. (i) *f is (right-side) long-tailed, denoted by $f \in \mathcal{L}$, if there exists $x_0 > 0$ such that $f(x) > 0$, $x \geq x_0$ and for any fixed $y > 0$ $f(x+y) \sim f(x)$.*
(ii) *f is (right-side) subexponential on $\mathbb{R}$, denoted by $\mathcal{S}$, if $f \in \mathcal{L}$ and $f^{*2}(x) \sim 2f(x)$.*
(iii) *f with dist. F is weakly (right-side) subexponential on $\mathbb{R}$, denoted by $\mathcal{S}_+$, if $f \in \mathcal{L}$ and the function $f_+(x) = \mathbf{1}_{\mathbb{R}_+}(x)f(x)/\overline{F}(0)$, $x \in \mathbb{R}$ is subexponential, i.e. $f_+ \in \mathcal{S}$. Here $\overline{F}(x) = 1 - F(x)$.*

Definition 0.2. (i) *We say that a density $f : \mathbb{R} \to \mathbb{R}_+$ is asymptotic to a non-increasing function (a.n.i. for short) if f is locally bounded and positive on $[x_0, \infty)$ for some $x_0 > 0$, and*

$$\sup_{t\geq x} f(t) \sim f(x) \quad \text{and} \quad \inf_{x_0 \leq t \leq x} f(t) \sim f(x). \tag{0.1}$$

(ii) *We say that a density $f : \mathbb{R} \to \mathbb{R}_+$ is almost decreasing (al.d. for short) if there exists $x_0 > 0$ and $K > 0$ such that*

$$f(x+y) \leq Kf(x), \quad \text{for all } x > x_0,\ y > 0.$$

Notice that the al.d. property includes the a.n.i. property, and the latter is satisfied by the regularly varying functions with negative indices.

We will investigate properties of the above sort, particularly on infinitely divisible distributions μ on $\mathbb{R}$. The characteristic function (ch.f.) of μ is

$$\widehat{\mu}(z) = \exp\Big\{ \int_{-\infty}^{\infty} (e^{izy} - 1 - izy\mathbf{1}_{\{|y|\leq 1\}})\nu(dy) + iaz - \frac{1}{2}b^2z^2 \Big\}, \tag{0.2}$$

where $a \in \mathbb{R}$, $b \geq 0$ and ν is the Lévy measure satisfying $\nu(\{0\}) = 0$ and $\int_{-\infty}^{\infty}(1 \wedge x^2)\nu(dx) < \infty$. Throughout this paper, we always assume that the Lévy measure ν of μ has a density, and we denote by $\mathrm{ID}(\mathbb{R})$ the class of all infinitely divisible distributions on $\mathbb{R}$.

Main contents

We separate the cases depending on weather $\nu(\mathbb{R}) < \infty$ or $\nu(\mathbb{R}) = \infty$. Note that we use notation g also for the (non-proper) density of a Lévy measure.

Theorem 0.3. *Let $\mu \in \mathrm{ID}(\mathbb{R})$ with $\nu(dx) = g(x)dx$ such that $\nu(\mathbb{R}) < \infty$. Denote the non-Gaussian part μ', which is a γ-shifted compound Poisson given by*

$$\mu'(dx) = e^{-\lambda}\delta_\gamma(dx) + (1 - e^{-\lambda})f(x - \gamma)dx, \quad \gamma \in \mathbb{R},$$

where δ_γ is Dirac measure at γ, f is a proper density and $\lambda > 0$ is the Poisson parameter. Then the following are equivalent.

(i) $f \in \mathcal{S}_+$ *and* f *is al.d.*

(ii) $g \in \mathcal{S}_+$ *and* g *is al.d.*

(iii) $g \in \mathcal{L}$, g *is al.d. and* $\lim_{x\to\infty} f(x)/g(x) = \lambda/(1 - e^{-\lambda})$.

Theorem 0.4. *Let $\mu \in \mathrm{ID}(\mathbb{R})$ with $\nu(dx) = g(x)dx$ such that $\nu(\mathbb{R}) = \infty$. Suppose that $g_1(x) = \mathbf{1}_{\{x>1\}}g(x)/\nu((1,\infty))$ is bounded. For a density f of μ we consider the following properties.*

(i) $f \in \mathcal{S}_+$ *and* f *is al.d.*

(ii) $g_1 \in \mathcal{S}_+$

(iii) $g_1 \in \mathcal{L}$ & $\lim_{x\to\infty} f(x)/g_1(x) = \nu((1,\infty))$.

(*a*) *If g is a.n.i., then we can choose f such that* (i), (ii) *and* (iii) *are equivalent.*
(*b*) *If g is al.d., then we can choose f such that* (ii) $\Leftrightarrow$ (iii) *implies* (i).

Since the a.n.i. property includes regular variation, the following is immediate.

Corollary 0.5. *Let $\mu \in \mathrm{ID}(\mathbb{R})$ with $\nu(dx) = g(x)dx$ such that $\nu(\mathbb{R}) = \infty$. Suppose that $g_1(x) = \mathbf{1}_{\{x>1\}}g(x)/\nu((1,\infty))$ is bounded. Then, $g_1(x)$ is a.n.i. and we can choose a density f of μ such that it is regularly varying if and only if g is regularly varying, and in this case $f(x) \sim g(x)$.*

We apply our results to the consistency proof of the maximum likelihood estimation (MLE for short) for $\mu \in \mathrm{ID}(\mathbb{R})$ which is absolutely continuous. For simplicity we put $a = b = 0$ in $\widehat{\mu}(z)$ of (0.2) and assume that $\widehat{\mu}(z)$ is absolutely integrable.

Let $f(x;\theta)$ be the density of μ with θ a parameter vector and $g(x;\theta)$ be a density of the corresponding Lévy measure ν. Let $(X_1, \dots, X_n)$ be a random sample from $f(x;\theta_0)$ with $\theta_0 \in \Theta$ where Θ is a compact parameter space. Define the likelihood function

$$M_n(\theta) = n^{-1}\sum_{i=1}^{n} \log f(X_i;\theta).$$

MLE $\hat{\theta}_n$ maximizes the function $\theta \mapsto M_n(\theta)$. We say that a function $\alpha(x;\theta)$ is identifiable if $\alpha(\cdot\,;\theta) \neq \alpha(\cdot\,;\theta')$ every $\theta \neq \theta' \in \Theta$, i.e. $\alpha(x;\theta) \overset{a.e.}{=} \alpha(x;\theta')$ does not hold. For convenience, we only consider the symmetric or positive-half case, but we can easily generalize the result in the non-symmetric two-sided case. We use the function g_1 defined in Theorem 0.4.

Proposition 0.6. *Let $\mu \in \mathrm{ID}(\mathbb{R})$ given by* (0.2) *with $a = b = 0$ such that $\widehat{\mu}(z)$ is absolutely integrable. Let $g(x;\theta)$ be a symmetric or positive-half density of ν. Suppose* (i) : *$g(x;\theta)$ is identifiable, $\theta \mapsto g(x;\theta)$ is continuous in θ for every x, and $\int(\sup_{\theta\in\Theta} |\log g_1(x;\theta)|)g_1(x;\theta_0)dx < \infty$ with Θ a compact set such that $\theta_0 \in \Theta$. Suppose* (ii) : *$g_1(x;\theta)$ is bounded and a.n.i., and $g_1 \in \mathcal{S}$. Then MLE $\hat{\theta}_n$ satisfies $\hat{\theta}_n \xrightarrow{p} \theta_0$.*

References

[1] Matsui, M. (2022) Subexponentialiy of densities of infinitely divisible distributions, *arXiv:2205.02074*.

Existence and distributional properties of solutions for SPDEs driven by Lévy white noise

David Berger

In this talk we are discussing solutions of linear stochastic partial differential equations (SPDEs) driven by so-called Lévy white noise, where we are interested in distributional solutions, i.e. our solutions are generalized stochastic processes on the space of infinitely differentiable functions with compact support $\mathcal{D}(\mathbb{R}^d)$. Let us shortly recall the definition of a generalized process.

Definition 0.1 (see [1])**.** A generalized random process is a linear and continuous function $s : \mathcal{D}(\mathbb{R}^d) \to L^0(\Omega)$. The linearity means that, for every $\varphi_1, \varphi_2 \in \mathcal{D}(\mathbb{R}^d)$ and $\gamma \in \mathbb{R}$,

$$s(\varphi_1 + \gamma\varphi_2) = s(\varphi_1) + \gamma s(\varphi_2) \text{ almost surely.}$$

The continuity means that if $\varphi_n \to \varphi$ in $\mathcal{D}(\mathbb{R}^d)$, then $s(\varphi_n) \to s(\varphi)$ in $L^0(\Omega)$.

A generalized process s can also be seen as a random variable on the space of distributions $\mathcal{D}'(\mathbb{R}^d)$, which then gives rises to define the so-called characteristic functional $\mathcal{P}_s(\varphi) := \int\limits_{\mathcal{D}'(\mathbb{R}^d)} e^{i\langle u,\varphi\rangle} dP_s(u)$ for $\varphi \in \mathbb{D}(\mathbb{R}^d)$. The characteristic functional describes uniquely the probability measure generated by the generalized random process. We define a Lévy white noise as a generalized random process with a special characteristic functional.

Definition 0.2. A Lévy white noise $\dot{L}$ is a generalized random process, where the characteristic functional is given by

$$\widehat{\mathcal{P}}_{\dot{L}}(\varphi) = \exp\left(\int\limits_{\mathbb{R}^d} \psi(\varphi(x))\lambda^d(dx)\right)$$

for every $\varphi \in \mathcal{D}(\mathbb{R}^d)$, where $\psi : \mathbb{R} \to \mathbb{C}$ is given by

$$\psi(z) = i\gamma z - \frac{1}{2}az^2 + \int\limits_{\mathbb{R}} (e^{ixz} - 1 - ixz\mathbf{1}_{|x|\leq 1})\nu(dx)$$

where $a \in \mathbb{R}^+$, $\gamma \in \mathbb{R}$ and ν is a Lévy-measure, i.e. a measure such that $\nu(\{0\}) = 0$ and $\int\limits_{\mathbb{R}} \min(1, x^2)\nu(dx) < \infty$. We say that $\dot{L}$ has the characteristic triplet (a, γ, ν).

We are interested in solutions of the equation

$$p(D)s = q(D)\dot{L}, \tag{0.1}$$

where $p(D)$ and $q(D)$ are linear partial differential operators. We say that a generalized process s solves (0.1) if

$$\langle s, p(D)^*\varphi\rangle = \langle \dot{L}, q(D)^*\varphi\rangle \text{ for every } \varphi \in \mathcal{D}(\mathbb{R}^d).$$

Constructing an extension of the Lévy white noise to a greater domain we show the following.

Theorem 0.3 (see [1]). *Let p and q be polynomials such that the rational function $q(i\cdot)/p(i\cdot)$ has a holomorphic extension in a strip $\{z \in \mathbb{C}^d : \|\Im z\| < \varepsilon\}$ for some $\varepsilon > 0$ and let $\dot{L}$ be a Lévy white noise with characteristic triplet (a, γ, ν) such that*

$$\int_{\mathbb{R}} \mathbf{1}_{|r|>1} \log(|r|)^d \nu(dr).$$

Then there exists a stationary solution s of (0.1), which we call a stationary $CARMA(p, q)$ generalized process.

We also discuss distributional properties of the solution s.

Proposition 0.4 (see [1]). *Let $\dot{L}$ have existing β-moment ($\beta > 0$) and let p and q be polynomials satisfying the condition from Theorem 0.3. Then the stationary $CARMA(p, q)$ generalized process s constructed in Theorem 0.3 has existing β-moment, too.*

References

[1] D. Berger, *Lévy driven CARMA generalized processes and stochastic partial differential equations*, Stoch. Process. Their Appl. (2020), 130 (10), 5865-5887.

Nonparametric regression for locally stationary random fields on $\mathbb{R}^d$

Daisuke Kurisu (Yokohama National University)

1 Introduction

In this study, we develop an asymptotic theory of nonparametric regression for locally stationary random fields (LSRFs) $\{\boldsymbol{X}_{\boldsymbol{s},A_n} : \boldsymbol{s} \in R_n\}$ in $\mathbb{R}^p$ observed at irregularly spaced locations in $R_n = [0, A_n]^d \subset \mathbb{R}^d$. We first derive the uniform convergence rate of general kernel estimators, followed by the asymptotic normality of an estimator for the mean function of the model. Moreover, we consider additive models to avoid the curse of dimensionality arising from the dependence of the convergence rate of estimators on the number of covariates. Subsequently, we derive the uniform convergence rate and joint asymptotic normality of the estimators for additive functions. We also introduce *approximately m_n-dependent* RFs to provide examples of LSRFs. We find that these RFs include a wide class of Lévy-driven moving average RFs.

2 Model

In this study, we consider the following model:

$$Y_{\boldsymbol{s}_j,A_n} = m\left(\frac{\boldsymbol{s}_j}{A_n}, \boldsymbol{X}_{\boldsymbol{s}_j,A_n}\right) + \epsilon_{\boldsymbol{s}_j,A_n},\ \boldsymbol{s}_j \in R_n,\ j = 1, \ldots, n, \tag{1}$$

where $E[\epsilon_{\boldsymbol{s},A_n}|\boldsymbol{X}_{\boldsymbol{s},A_n}] = 0$ and $R_n = [0, A_n]^d \subset \mathbb{R}^d$ is a sampling region with $A_n \to \infty$ as $n \to \infty$. Here, $Y_{\boldsymbol{s}_j,A_n}$ and $\boldsymbol{X}_{\boldsymbol{s}_j,A_n}$ are random variables of dimensions 1 and p, respectively. We assume that $\{\boldsymbol{X}_{\boldsymbol{s},A_n} : \boldsymbol{s} \in R_n\}$ is a locally stationary random field on $R_n \subset \mathbb{R}^d$ $(d \geq 2)$.

3 Main Results

The objectives of this study are to (i) derive the uniform convergence rate of kernel estimators for the density function of $\boldsymbol{X}_{\boldsymbol{s},A_n}$ and the mean function m in the model (1) over compact sets; (ii) derive the asymptotic normality of the estimators at a specified point; and (iii) provide examples of locally stationary random fields on $\mathbb{R}^d$ with a detailed discussion of their properties. To attain the first and second objectives, we first derive the uniform convergence rate of the important general kernel estimators; the result is crucial for demonstrating our main results. As general estimators include a wide range of kernel-based estimators such as the Nadaraya-Watson estimators, the general results are of independent interest. Although these results are general, the estimators are affected by dimensionality because their convergence rate depends on the number of covariates. Hence, we consider additive models and derive the uniform convergence rate and joint asymptotic normality of kernel estimators for additive functions based on the backfitting method developed by Mammen et al. (1999) and Vogt (2012). Our results are extensions of the results for time series in Vogt (2012) to random fields with irregularly spaced observations, which include irregularly spaced time series as a special case.

To attain the third objective of our study, we discuss examples of locally stationary random fields on $\mathbb{R}^d$

that satisfy our regularity conditions. For this, we introduce the concept of *approximately m_n-dependent* locally stationary random fields ($m_n \to \infty$ as $n \to \infty$) and we extend continuous autoregressive and moving average (CARMA)-type random fields developed in Brockwell and Matsuda (2017) to locally stationary CARMA-type random fields. CARMA random fields are characterized by solutions of (fractional) stochastic partial differential equations (cf. Berger (2020)) and are known as a rich class of models for spatial data (cf. Brockwell and Matsuda (2017)).

References

Berger, D. (2020). Lévy driven CARMA generalized processes and stochastic partial differential equations. *Stochastic Process. Appl.* **130**, 5865-5887.

Brockwell, P.J. and Matsuda, Y. (2017). Continuous auto-regressive moving average random fields on $\mathbb{R}^n$. *J. Roy. Statist. Soc. Ser. B Stat. Methodol.* **79** 833-857.

Kurisu, D. (2022). Nonparametric regression for locally stationary random fields under stochastic sampling design. *Bernoulli* **28**, 1250-1275.

Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27** 1443-1490.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601-2633.

Intergenerational Risk Sharing in a Defined Contribution Pension System: Analysis with Bayesian Optimization

An Chen[1], Motonobu Kanagawa[2], and Fangyuan Zhang[2]

[1]Institute of Insurance Science, Ulm University
[2]Data Science Department, EURECOM

The talk at the 3rd Tohoku-ISM-UULM workshop in Japan

Abstract

We study a fully funded, collective defined-contribution (DC) pension system with multiple overlapping generations. We investigate whether the welfare of participants can be improved by intergenerational risk sharing (IRS) implemented with a realistic investment strategy (e.g., no borrowing) and without an outside entity (e.g., share holders) that helps finance the pension fund. To implement IRS, the pension system uses an automatic adjustment rule for the indexation of individual accounts, which adapts to the notional funding ratio of the pension system. The pension system has two parameters that determine the investment strategy and the strength of the adjustment rule, which are optimized by expected utility maximization using Bayesian optimization. The volatility of the retirement benefits and that of the funding ratio are analyzed, and it is shown that the trade-off between them can be controlled by the optimal adjustment parameter to attain IRS. Compared with the optimal individual DC benchmark using the life-cycle strategy, the studied pension system with IRS is shown to improve the welfare of risk-averse participants, when the financial market is volatile.

On Modified Interdirections

Jana Klicnarová*

The high-dimensional data analysis is still more and more required; therefore, simple, robust, and powerful nonparametric tests for high-dimensional data are in demand. One of the possible concepts of how to handle such data is based on hyperplane signs (see [Randles, 1989] and ranks (see [Hettmansperger et al., 1999], [Oja and Paindaveine, 2005]). These concepts became popular due to their simplicity, geometric interpretation, affine invariance, weak moment assumptions, and robustness to both – radial and angular – outliers. Many authors were interested in such approaches, see [Randles and Peters, 1990], [Jan and Randles, 1996], [Gieser and Randles, 1997], [Hallin and Paindaveine, 2002]. On the other hand, the original concepts suffer from high computational demands, which led to falling out of the fashion of these technics. In the talk, we will discuss the possibilities of these concept modifications, which reduce computational effort, and, if possible, preserve the advantages of these methods.

We focus on statistics based on hyperplanes in more detail. More precisely, on statistics based on interdirections [Randles, 1989] and lift-interdirections [Hettmansperger et al., 1999], [Oja and Paindaveine, 2005]. Let us suppose p dimensional sample of size n. The classical tests based on interdirections and lift-interdirections are computationally feasible only for small data dimensions p and small sample sizes n. If we take a closer look at the statistics, we realize that we handle with complete U-statistics. It means that the computational demand arises exponentially with the dimension p. Therefore, Hudecova et al. ([Hudecová et al., 2020]) discussed the possibility of using incomplete U-statistics instead of complete ones, i.e., they define so-called incomplete lift-interdirections and incomplete interdirections. They showed, for interdirections, that it is possible, in test statistics given by Randles ([Randles, 1989]), to use incomplete interdirections based on subsets chosen randomly with or without replacement if the cardinality of the subsets ($m_{n,p-1}$) satisfies: $m_{n,p-1}/n \to 0$ as $n \to \infty$ (n is a size of sample). Also in the test statistics proposed by Oja and Paindaveine [Oja and Paindaveine, 2005], it is possible to replace classical interdirections and lift-interdirections by incomplete ones if design sets are chosen independently and randomly with or without replacement, satisfying that $m_{n,p-1}/n \to 0$ as $n \to \infty$ and $m^s_{n,p}/n \to 0$ as $n \to \infty$, where $m^s_{n,p}$ is a cardinality of design set for lift-interdirections.

Such results allow us to reduce the computational demand of tests. In the case of incomplete interdirections, the number of considered hyperplanes does not grow with the dimension of observation; hence we can also easily handle samples from $p = 100$ dimensional spaces. On the other hand, for lift-interdirections, the number of needed hyperplanes depends on the dimension by the factor 2^p. I.e., for $p > 10$, the computation of the test is already quite difficult.

*Faculty of Economics, University of South Bohemia, Studentská 13, CZ-370 05, České Budějovice, Czech Republic
email: klicnarova@ef.jcu.cz

Another possibility of allowing a smaller number of hyperplanes is to apply randomized lift-interdirections. In such a case, the test statistics asymptotics will be affected by randomization; on the other hand, it allows us to use tests based on this concept in higher dimensions. In the paper [Hudecová and Šiman, 2022], Hudecová and Šiman study such an approach, give theoretical results and provide a small simulation study. Such randomized lift-interdirections allow to run of the test for high dimensional data ($p > 100$) and large samples ($n > 1000$) in a short time (the computational demand does not depend on p); on the other hand, there is some uncertainty to the statistical inference, given by randomization. Still, it does not affect the mean p-value too much; moreover, it quickly becomes reasonably small with a growing number of chosen hyperplanes.

References

[Gieser and Randles, 1997] Gieser, P. W., Randles, R. H. (1997). A nonparametric test of independence between two vectors. Journal of the American Statistical Association, 92(438), 561-567.

[Hallin and Paindaveine, 2002] Hallin, M., Paindaveine, D., 2002. Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Annals of Statistics* 30, 1103–1133.

[Hettmansperger et al., 1999] Hettmansperger, TP., Möttönen, J. and Oja, H., 1999. The geometry of the affine invariant multivariate sign and rank methods. *Journal of Nonparametr. Stat.* 11, 271–285.

[Hettmansperger and Randles, 2002] Hettmansperger, T. P., Randles, R. H., 2002. A practical affine equivariant multivariate median. *Biometrika* 89, 851–860.

[Hudecová and Šiman, 2022] Hudecová, Š., Šiman, M. (2022). Multivariate ranks based on randomized lift-interdirections. *Computational Statistics & Data Analysis*, 172, 107480.

[Hudecová et al., 2020] Hudecová, Š., Klicnarová, J., and Šiman, M. (2020). Incomplete interdirections and lift-interdirections. *Journal of Nonparametric Statistics*, 32(1), 93-108.

[Jan and Randles, 1996] Jan, S.-L., Randles, R. H., 1996. Interdirection tests for simple repeated-measures designs. *Journal of the American Statistical Association* 91, 1611–1618.

[Oja and Paindaveine, 2005] Oja, H., Paindaveine, D., 2005. Optimal signed-rank tests based on hyperplanes. *Journal of Statistical Planning and Inference* 135, 300–323.

[Randles, 1989] Randles, R. H., 1989. A distribution-free multivariate sign test based on interdirections. *Journal of the American Statistical Association* 84, 1045–1050.

[Randles and Peters, 1990] Randles, R. H., Peters, D., 1990. Multivariate rank tests for the two-sample location problem. *Communications in Statistics - Theory and Methods* 19, 4225–4238.

Extracting Valuable Information based on Human Behavior

Satoshi Shioiri
Advanced Institute of Yotta Informatics & Research Institute of Electrical Communication, Tohoku University

Accumulation of information is essential for human knowledge production, and information technology has accelerated the speed of data accumulation. The momentum of expansion of its production volume is growing more and more in this century. Based on a forecast, it will exceed 1 Yottabytes in 2030 [1, 2]. We introduce an approach for solution of these problems [3, 4] with an example of research along the approach, in order to solve the problem of quantity and quality of information: that is, which information deserves to be kept, what information should be left for knowledge production, which information should be discarded in order to secure storage capacity. This can be recognized as evaluation of data value. It is difficult to define and evaluate the qualitative aspects of information compared to the quantitative aspects of information. Our research group proposed a new approach based on information quality and value concepts as "quali-informatics" [4]. Since evaluation of information is inherently relative to the evaluator's subjectivity and to the context, the target is a specific data set for individual person, organization or society. Critical factor of quali-informatics is, therefore, evaluation by human [5].

In this presentation, I introduce three topics related to evaluation of value of information: attention as a selection process, learning effect for selection, and estimation of subjective judgments by machine learning. First is selecting information by attention. The brain receives a vast amount of input from the sensory organs such as eyes and ears each second and the brain must decide what to process by selecting the appropriate information. The function of selective attention is indeed what is required in society with Yottabyte scale data. The human attention process could provide clues for evaluation method of information values. The visual system, for example, can attend to several aspects of stimulation such as space, time, features, objects, link to body or action and so on [6,7]. This indicates that the brain selects information processing a limited number of aspects instead of any possible information from the retina, such as signals from individual photoreceptors or any possible combination of them. In other words, aspects that can be attended is along the line of value estimation of the visual system. Value of information is evaluated based on processing of the information in the brain.

Second is learning effect for selection. It is not surprising that a person can find things in a familiar place like own office easily than in a place novel to s/he. Either/both explicit or/and implicit learning generates representation of a place through spending time there, with which s/he can find things there efficiently. If the value of information from the environment, such as arrangement of things in a room, is considered for finding things in a place, objects identified as landmarks and global feature that identify the room are likely valuable. Since their meaning is only based on previous experience, we can say that knowledge plays important roles for evaluation of information values. Tsuchiai et al. reported that object arrangements in a space is memorized implicitly in three dimensional representations throughout repeated excursion of visual search [8]. The study compared the transfer of the contextual cueing effect between cases with and without self-motion by using visual search displays for 3D objects. The contextual cueing effect is a learning effect of spatial layout in visual search displays. The contextual cueing effect was obtained with self-motion but disappeared when the display changed without self-motion. This indicates that there is an implicit learning effect in spatial coordinates and suggests that the spatial representation of object layouts or scenes can be obtained and updated implicitly.

Third is the use of image features and facial expressions to evaluate the value of information. To evaluate value of information, it is fundamental to find features that is relevant to the usage of the information. In the condition where the purpose to decide whether to keep or delete pictures in a mobile device, one method is image features can be used to evaluate human preference to each picture. A computer can make the decision by finding the link between image features and preference judgements. Since human judgements are subjective, collection of preference judgements results are required. Similar to image features, facial expressions during judgements can also be used to evaluate value of the pictures. Shioiri et al reported that both image features and facial expressions are useful to estimate preference judgments. In the study, participants was asked to judge whether they like or dislike (prefer/not prefer) each picture retrieved from Instagram, which is a photo-sharing social network service. Stimulus pictures were obtained with a hashtag to collect a certain type of pictures (#luchbox and #landscape). The face was video recorded while participants were thinking for the preference, and the facial expressions were analyzed later. Pictures used were also analyzed to extract image features using a trained network (Resnet-50). Facial features were extracted using open source software (OpenFace). Image and face features were used to predict participants' judgments with a machine learning method (LightGBM) to predict the participants preference from image and facial features. The results of the analyses showed that facial expression features are useful for predicting the participants' judgments of preference. However, it was

also found that prediction from the facial expression is likely limited for the same individuals since prediction accuracy is poor when training data and the test data were from different participants.

Evaluation of information values is critical in Yottabyte scale society. I believe that approaches with consideration of human sciences will play important roles there as described in this presentation.

[1] DC2012 THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East, https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

[2] L. Floridi, the 4TH Revolution: How the Infosphere is reshaping Human Reality, Oxford University Press, 2014.

[3] V. Smil, Data World: Racing toward Yotta. IEEE Spectrum 56(7), p. 20, 2019

[4] H. Muraoka, G. Gyoba, Y. Suzuki, S. Shioiri, M. Nakao, M. Nihei & H. Shoji, Gigantic Amount Information and Storage Technology : Challenge to Yotta-Byte-Scale Informatics (in Japanese), IEICE technical report : 116(441), pp. 27-32, 2017 (Invited)

[5] S. Shioiri, Y. Sato, Y. Horaguchi, H. Muraoka and M. Nihei, Quali-informatics in the society with yotta scale data, IEEE International Symposium on Circuits and Systems 2021, Pages: 1-4

[6] S. Shioiri, H. Honjyo, Y. Kashiwase, K. Matsumiya and I. Kuriki, Visual attention spreads broadly but selects information locally, Scientific Reports 2016 Vol. 6 Issue ARTN 35513, Accession Number: WOS:000385953900001

[7] S. Shioiri, T. Sasada and R. Nishikawa, Visual attention around a hand location localized by proprioceptive information, Cereb Cortex Commun 2022 Vol. 3 Issue 1 Pages tgac005, Accession Number: 35224493

[8] S. Shioiri, H. Honjyo, Y. Kashiwase, K. Matsumiya and I. Kuriki, Visual attention spreads broadly but selects information locally, Scientific Reports 2016 Vol. 6 Issue ARTN 35513, Accession Number: WOS:000385953900001

Measurement and statistics of taste and deliciousness.

Nobuyuki Sakai (SAL, AI Yotta and RIEC, Tohoku University)

The amount of data from our daily lives is too large to be handled by humans. The largeness of the data sometimes direct humans toward the wrong conclusion. To avoid this problem, the Advanced Institute of Yotta Informatics aims to proceed the research on the "data triage" system. In the Advanced Institute of Yotta Informatics, we are studying on the cognitive and brain mechanisms about human evaluation in our daily lives, especially in our eating behaviors. In this talk, I will present following four topics;

1. What is the TASTE?
2. How can we measure the TASTE?
3. Can the DELICIOUSNESS be measured?
4. What is the DELICIOUSNESS?

1. *What is the TASTE?*

We feel TASTEs during eating the foods and/or drinking the beverages. We think TASTEs come from the oral cavity and tongue. But there are some mistakes in this thought.

First, researchers in psychology and physiology define the taste sensation as "the sensation evoked by the activities of taste buds responding to the chemical substances delivered from the foods and beverages". The taste sensation is named as gustation, and consists of five basic tastes: sweet, umami, salty, sour, and bitter. The sweet taste is a signal for an energy source (=sugar), the umami taste is for body components (amino acids and nucleic acid), the salt taste is for mineral, the sour taste is for unripen fruit or for putrefied foods, and the bitter taste is for neuroactive toxin. Thus, gustation is a very important sense to maintain our lives.

But can you describe the TASTE of your breakfast with these five tastes? As this question indicates, the gustation is only a part of the TASTEs. ***The main part of the TASTEs is olfaction, the odor sense.*** The tactile sensation in oral cavity is also an important component of TASTE, ex. the hot taste is tactile sensation evoked by an activation of a kind of TRP V1 receptor.

2. *How can we measure the TASTE?*

The discipline of measuring human senses is called as "Sensory Studies". The Sensory Studies are highly developed in the field of food science; we can say ***the Sensory Studies are specially equipped to measure the TASTEs.*** The methods to measure the TASTEs are grouped into two parts; the quantitative measurements and the qualitative one. The former can be measured with the 5~11 points-scale or the **visual analogue scale (VAS)**. We can obtain numerical data about how strong humans perceive the intensity of the

TASTE. Currently, the temporal changes of TASTE intensities can be measured with the **Time-Intensity methods (TI)**.

The latter can be measured computer-based tools, such as the **Temporal Dominance of Sensation (TDS)**, the **Check All That Apply method (CATA)**, etc. We can obtain numerical pattern or distributions of properties about the TASTEs.

Also we can apply the neuroscientific tools for measuring the TASTEs such as fMRI, ERP, and fNIRS. However, the data obtained with those tools include not only the sensory data but also cognitive data. Thus, we should pay attention to interpretation of the neuronal data without the behavioral data.

3. *Can DELICIOUSNESS be measured?*

   As the proverb "there is no accounting for taste" says, we intuitively think measurements of DELICIOUSNESS are meaningless.

   However, there are oppressive needs for them by food industries. Some methods derived from the TASTE measurements are developed. The methods presuppose the following hypothesis: Deliciousness is the sensory information evoked by the chemical component of the foods.

   Other methods to measure DELICIOUSNESS are derived from the animal experiment. In these methods, the amounts of eaten by participants are considered as indices of DELICIOUSNESS. These methods presuppose the following hypothesis: People eat more when the food is delicious.

   There are some successful studies showing the palatability can be measured with these methods. But our experiments showed that ***both hypotheses are null***. The ratings and the amount eaten are affected not only by the DELICIOUSNESS but also by the context, such as back ground music, order, tableware, choice, misattribution, etc. Thus, we should pay attention to interpretation of the ratings and the amount without cognitive responses to the context.

4. *What is the DELICIOUSNESS?*

   Our studies showed that humans do not make evaluations of DELICIOUSNESS irrationally. The evaluations are understandable with the cognitive and the behavioral data. The real DELICIOUSNESS is not the components of the foods, but the cognitive functions and the emotions related to them. ***The real DELICIOUSNESS is not in the foods, but in our brains.*** Thus, we should attach importance to the cognitive and the behavioral data rather than the chemical compound and the sensory properties of the foods.

Conclusion:

We should develop the "Data Triage" system to understand TASTEs and DELICIOUSNESS.

Recommendation Systems with Network Structure and Big Data

Tsukasa ISHIGAKI (Tohoku University)

Recommendation system supports to provide some valued items for each user from huge number of alternatives and it are used in a variety of industries, including retail, restaurant, news, trip, music, movie and media. The market of recommendation systems was valued at USD 1.2 billion in 2020 [Mordor Intelligence 2021] and is forecast to reach USD 16.13 Billion by 2026 [Reports and Data 2021]. Recommendation system has a crucial role in today's business.

In this talk, I will talk about some novel recommendation methods using network structures and big data. First, I will provide some introductory topics about motivations, properties of data, problem settings and basic methods on recommendation system. Then, three topics that we have proposed in [1-3] will be talked. The methods have been realized using deep learning with implicit feedback or knowledge graph. The results show that the proposed methods have a high performance in terms of accuracy or novel recommendation in some experiments using big data.

[1] Linh Nguyen, Tsukasa Ishigaki, D2D-TM: A Cycle VAE-GAN for Multi-Domain Collaborative Filtering, Proceedings of IEEE International Conference on Big Data (IEEE Big Data 2019), pp.1175-1180, 2019

[2] Kachun Lo, Tsukasa Ishigaki, PPNW: Personalized Pairwise Novelty Loss Weighting for Novel Recommendation, Knowledge and Information Systems, Vol.63, No.5, pp.1117-1148, 2021

[3] Kachun Lo, Tsukasa Ishigaki, X-2ch: Quad-Channel Collaborative Graph Network over Knowledge-Embedded Edges, Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2021), pp.2076-2080, 2021

Non-concave optimization under risk constraints

An Chen*, Mitja Stadje* and Fangyuan Zhang‡

August 20, 2022.

* Institute of Insurance Science, Ulm University, Helmholtzstr. 20, 89069 Ulm, Germany.
E–Mails: an.chen@uni-ulm.de, mitja.stadje@uni-ulm.de
‡ EURECOM, E–Mail: fangyuan.zhang0069@gmail.com

In the literature, a non-concave maximization problem under a Value-at-Risk (VaR) or an Average Value-at-Risk (AVaR) constraint has been studied. For example, one can find the optimal solution for a non-concave maximization problem arising from a non-linear payoff function under a VaR constraint in Nguyen and Stadje (2020), and the optimal solution with an S-shape utility function under a VaR constraint in Dong et al. (2019). In addition, a non-linear payoff function under an AVaR constraint is studied in De Franco and Tankov (2011), and an S-shape utility function under an AVaR constraint is discussed in Armstrong and Brigo (2019). However, both of these discussions under an AVaR constraint are restricted to some special cases. Despite these studies, the closed form solutions for a non-concave utility maximization under an Expected Shortfall or an Expected Discounted Shortfall constraint are missing. Moreover, a thoroughly comparative analysis of the risk constraints on a non-concave maximization problem based on closed form solutions is also missing in the literature. Our paper fill this gap.

In this work, we consider a non-concave maximization problem under a risk constraint (Expected Shortfall, Expected Discounted Shortfall, Value-at-Risk, Average Value-at-Risk). We provide closed-form solutions by the Lagrangian approach. Due to the non-concave utility function and the risk constraint, the constructed Lagrangian for solving the optimization problem is highly non-concave and discontinuous. We show that by decomposing the highly non-concave Lagrangian into local *concave* or *affine* functions on *disjoint* sets, the global maximum is obtained by comparing local maximums of piece-wise Lagrangian, which reduces to finding the zero roots of a sequence of conjugate

functions. However, it is non-trivial to determine the global maximum as the conjugate functions do *not* necessarily have zero roots, which implies that the existence of the optimal solution is *not* fully ensured. Note that in traditional concave utility maximization problems under risk constraints, the existence of the optimal solution is guaranteed by the existence of the Lagrangian multipliers. But in the non-concave maximization problem under risk constraints, the optimal solution exists if the Lagrangian multipliers as well as the zero roots of the conjugate functions both exist, which is more challenging to prove. In this paper, we work out the conditions under which the zero roots of the conjugate functions exist and obtain full analytical solutions for the non-concave utility maximization problems under risk constraints.

References

Armstrong, J. and Brigo, D. (2019). Risk managing tail-risk seekers: Var and expected shortfall vs s-shaped utility. *Journal of Banking and Finance*, 101:122–135.

De Franco, C. and Tankov, P. (2011). Portfolio insurance under a risk-measure constraint. *Insurance: Mathematics and Economics*, 49(3):361–370.

Dong, Y., Wu, S., Lv, W., and Wang, G. (2019). Optimal asset allocation for participating contracts under the var and pi constraint. *Scandinavian Actuarial Journal*, pages 1–26.

Nguyen, T. and Stadje, M. (2020). Nonconcave optimal investment with value-at-risk constraint: An application to life insurance contracts. *SIAM Journal on Control and Optimization*, 58(2):895–936.

Statistical modeling for the nonlinear structure of quantiles in hierarchical data via regularization

Yuko Araki
Graduate School of Information Sciences
Tohoku University

The purpose of this study was to develop a model that can estimate the transition of quantile points of a response variable from longitudinal data with a nonlinear structure when the nonlinear function cannot be identified in advance. To achieve this goal, we consider a basis expansion method for both fixed- and random-effects terms in quantile mixed-effects models. Since models with basis expansions can be expressed in the same form as a linear model, they can be treated as a special case of the linear quantile mixed-effects model, such as those proposed by Geraci and Bottai (2007) and Geraci and Bottai (2014).

To overcome the expensive computational cost of this approach, we proposed the Bayesian nonparametric quantile mixed-effects models (BNQMs, Tanabe, Araki et al. 2022). In BNQMs, Bayesian regularization can be performed by assuming a specific distribution of the prior for the coefficient parameters of the basis function. The priors or hyperpriors are assumed hierarchically for the parameters of the priors for the coefficient parameters, and regularization parameters can be estimated simultaneously by the MCMC method. We also proposed the use of a Gaussian process (GP) prior for the coefficient parameter vectors of the basis functions. We showed that the smoothness can be appropriately adjusted when the number of basis functions is excessive by using a GP prior for Bayesian regularization.

We take the observation of the i-th individual at the j-th measurement time t_{ij}, $\{(t_{ij}, y_{ij}); i = 1, ..., N, j = 1, ..., n_i\}$. Then, the τ-th quantile of y_{ij} at t_{ij} can be modeled as

$$\begin{aligned} Q_{y_{ij}}(\tau|t_{ij}) &= \sum_{k=1}^{p} \beta_{\tau k}\phi_k(t_{ij}) + \sum_{l=1}^{q} b_{\tau il}\psi_l(t_{ij}) \\ &= \boldsymbol{\beta}_\tau^\top \boldsymbol{\phi}(t_{ij}) + \boldsymbol{b}_{\tau i}^\top \boldsymbol{\psi}(t_{ij}), \end{aligned} \tag{1}$$

where $\tau \in (0,1)$, $Q_{y_{ij}}(\cdot) \equiv F^{-1}(\cdot)$, and $\boldsymbol{\phi}(t) = (\phi_1(t), \cdots, \phi_p(t))^\top$ and $\boldsymbol{\psi}(t) = (\psi_1(t), \cdots, \psi_q(t))^\top$ are vectors of the basis functions in the fixed- and random-effects terms, respectively, $\boldsymbol{\beta}_\tau = (\beta_{\tau 1}, \cdots, \beta_{\tau p})^\top$ is a $p \times 1$ coefficient parameter vector of $\boldsymbol{\phi}(t)$, and $\boldsymbol{b}_{\tau i} = (b_{\tau i1}, \cdots, b_{\tau iq})^\top$ is a $q \times 1$ coefficient parameter vector of $\boldsymbol{\psi}(t)$, where $\boldsymbol{b}_{\tau i} \sim N(\boldsymbol{0}, \boldsymbol{\Gamma}_\tau)$ is assumed. Here, $\boldsymbol{\Gamma}_\tau$ is a $q \times q$ positive-definite covariance matrix.

We assume that the conditional distribution of y_{ij} is an asymmetric Laplace distribution. Here, the conditional distribution of y_{ij} is written as $y_{ij}|\boldsymbol{\beta}, \boldsymbol{b}_i, \sigma \sim AL(\mu_{ij}, \sigma, \tau)$; therefore, its probability density function can be written as

$$p(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b}_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y_{ij} - \mu_{ij}}{\sigma}\right)\right\}, \tag{2}$$

where $\mu_{ij} = \boldsymbol{\beta}^\top \boldsymbol{\phi}(t_{ij}) + \boldsymbol{b}_i^\top \boldsymbol{\psi}(t_{ij})$ is the location parameter, $-\infty < \mu_{ij} < \infty$, $\sigma > 0$ is the scale parameter, $0 < \tau < 1$ is the skewness parameter, and $\rho_\tau(u) = u(\tau - I(u < 0))$ is the loss function with indicator function $I(\cdot)$. For further details of the model and estimation, including GP prior regularization for basis expansion method and prior setting and posterior of BNQM, please refer to Tanabe, Araki et al. (2022).

The performance of BNQMs was evaluated by a Monte Carlo simulation. The proposed BNQMs showed the highest estimation accuracy for each data structure and were shown to be useful as a quantile regression technique in hierarchical data with a nonlinear structure. Then, a BNQM was applied to longitudinal data of cortisol in infants. The results suggested that the cortisol secretion rhythm in infancy is bimodal, and the magnitude of the amplitude increases as the cortisol level itself increases.

References

[1] Geraci M, Bottai M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*. **8**(1), 140-54.

[2] Geraci, M., Bottai, M. (2014). Linear quantile mixed models. *Stat Comput* . **24**, 461–479.

[3] Tanabe, Y., Araki, Y., Kinoshita, M. et al. (2022). Bayesian nonparametric quantile mixed-effects models via regularization using Gaussian process priors. *Jpn J Stat Data Sci.*, **5**, 241–267.

Prediction of Heavy-Tailed Random Functions

Vitalii Makogin*, Evgeny Spodarev†

September 9, 2022

This talk is based on a recent paper [3]. The R extrapolation code for heavy tailed time series can be found in [2].

Let $Y : \Omega \to \mathbb{R}$ be a square integrable random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G} \subset \mathcal{F}$ be an observable sub–σ–algebra.The classical L^2–theory of prediction of random variables states that the best unbiased predictor of Y with respect to $\mathcal{G}$ is given by the conditional expectation $\mathbf{E}(Y|\mathcal{G})$. But as far as Y has no finite moments, no unified widely accepted prediction theory exists. Our contribution creates such theory which also applies to the finite variance case. Its main idea is the following. Let $u \in \mathbb{R}$ be an excursion level chosen according to a finite measure $m(\cdot)$ on $\mathbb{R}$. For any two random variables $Y_1, Y_2 : \Omega \to \mathbb{R}$ introduce the *excursion pseudo-metric*

$$E_m(Y_1, Y_2) := \int_{\mathbb{R}} \mathbb{P}(\{Y_1 > u\} \triangle \{Y_2 > u\})\, m(du),$$

which is an m-weighted average probability of symmetric difference of excursions of Y_1 and Y_2 over $u \in \mathbb{R}$. Then, we say that Z is a prediction of a random variable Y onto the σ–algebra $\mathcal{G}$ introduced above if

$$Z = \operatorname{argmin}_{Y_0} E_m(Y, Y_0),$$

whenever this minimum (taken over all $\mathcal{G}$-measurable random variables Y_0) exists and is unique. Sometimes we add more constraints to the geometry of our projection space saying that, additionally to $\mathcal{G}$-measurability, $Y_0 \stackrel{d}{=} Y$ or that Y_0 is a linear combination of observables. Apparently, the above solution Z, its existence and uniqueness may heavily depend on the choice of measure $m(\cdot)$. A natural candidate for this would be the distribution of Y. We will show that E_m is a metric on the space of random variables whenever the distribution function of $m(\cdot)$ is strictly increasing.

The intuition behind the use of the new metric is the following. Assume that a stationary heavy–tailed time series $\{Y_t, t \in \mathbb{R}\}$ is observed at locations $t_1, \dots, t_n$ in a compact window $W \subset \mathbb{R}$. As proposed in [1], the linear predictor $\widehat{Y}_t = \sum_{j=1}^{n} \lambda_j Y_{t_j}$, $t \notin \{t_1, \dots, t_n\}$, is a minimizer of the functional

$$\int_{\mathbb{R}} \mathbf{E}\left[v_1\left(A_Y(u) \Delta A_{\widehat{Y}}(u)\right)\right]\, m(du) = \int_W E_m(Y_t, \widehat{Y}_t)\, dt$$

with respect to the choice of weights $\lambda_1, \dots, \lambda_n$ subject to the constraint $\widehat{Y}_t \stackrel{d}{=} Y_t$. The left hand side term is the mean length of the symmetric difference of excursion sets $A_Y(u) := \{t \in W : Y_t > u\}$ and $A_{\widehat{Y}}(u) := \{t \in W : \widehat{Y}_t > u\}$ averaged over the levels $u \in \mathbb{R}$ picked up according to the measure $m(\cdot)$. Here $v_1(\cdot)$ is the Lebesgue measure on $\mathbb{R}$.

Let U be a random variable with probability law m which is independent of Y_1, Y_2. It can be interpreted as a random excursion level which we choose to build the metric E_m. We show that E_m coincides with the so–called *separation (pseudo) metric* [4] whenever m is a probability measure.

Lemma 1. *Let m be a probability measure on $\mathbb{R}$ with c.d.f. $F_U(x) = \int_{-\infty}^{x} m(dy)$, $x \in \mathbb{R}$. Then*

$$E_m(Y_1, Y_2) = \mathbf{E}|F_U(Y_2-) - F_U(Y_1-)|.$$

*Institute of Stochastics, Ulm University, Germany; vitalii.makogin@uni-ulm.de
†Institute of Stochastics, Ulm University, Germany; evgeny.spodarev@uni-ulm.de

Theorem 1. *Let X_S be the space of random variables with support $S \subseteq \mathbb{R}$. If F_U is strictly increasing on S, then E_m is a metric on $X_S \times X_S$.*

Restricted to the space of random variables Y with the same absolutely continuous distribution F, the metric E_F turns to be distribution–free depending only on bivariate copulas. We call this metric a *Gini metric*. It properties are investigated.

Consider the prediction of a value of random variable X with continuous distribution function F_X based on the set $\mathcal{X}_n := (X_1, \dots, X_n)$ of realizations of X via the excursion metric metric E_{F_X}. Namely, we propose a predictor $\widehat{X}_\lambda := g(\lambda, \mathcal{X}_n)$, where $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda \subset \mathbb{R}^n$ is deterministic and $g : \mathbb{R}^n \times \Lambda \to \mathbb{R}$, $n \in \mathbb{N}$, is a continuous measurable function such that the excursion metric is minimal:

$$\hat{\lambda} := \arg\min_{\lambda \in \Lambda} E_{F_X}(X, \widehat{X}_\lambda) = \arg\min_{\lambda \in \Lambda} \left[2\mathbf{E}F_X(X \vee \widehat{X}_\lambda) - \mathbf{E}F_X(\widehat{X}_\lambda) \right].$$

Here, the set of admissible parameters Λ as well as the analytic form of $g(\lambda, \mathcal{X}_n)$ depend on the law $\mathbb{P}_X$, e.g., $g(\lambda, \mathcal{X}_n) = \sum_{j=1}^n \lambda_j X_j$ for infinitely divisible laws of X or $g(\lambda, \mathcal{X}_n) = \max_{j=1,\dots,n} \lambda_j X_j$ for max–stable X. Since g and F_X are continuous, the constraint $\widehat{X}_\lambda \stackrel{d}{=} X$ is equivalent to $F_X(\widehat{X}_\lambda) \stackrel{d}{=} U \sim U(0,1)$. Denote $\Lambda_g := \{\lambda \in \mathbb{R}^n : F_X(\widehat{X}_\lambda) \stackrel{d}{=} U\}$. Should our prediction be law-preserving, the above optimization problem rewrites

$$\hat{\lambda} := \arg\min_{\lambda \in \Lambda_g} \left\{ \mathbf{E}F_X(X \vee \widehat{X}_\lambda) \right\}.$$

If the analytic form of Λ is given explicitly but Λ_g is hardly available, we modify the minimization functional by adding a term which penalizes a difference between the law of $F_X(\widehat{X}_\lambda)$ and $U(0,1)$:

$$\hat{\lambda} := \arg\min_{\lambda \in \Lambda} \left\{ 2\mathbf{E}F_X(X \vee \widehat{X}_\lambda) - \mathbf{E}F_X(\widehat{X}_\lambda) + \gamma \left[\mathbf{E}F_X^2(\widehat{X}_\lambda) - \mathbf{E}[F_X(\widehat{X}_\lambda) \vee Y] \right] \right\},$$

where $\gamma > 0$ and Y is an independent copy of $F_X(\widehat{X}_\lambda)$, which allows for an equivalent reformulation

$$\hat{\lambda} := \arg\min_{\lambda \in \Lambda} \left\{ 2\mathbf{E}F_X(X \vee \widehat{X}_\lambda) - \mathbf{E}F_X(\widehat{X}_\lambda) + \gamma \int_0^1 F_{Y_1}(y) \left[F_{Y_1}(y) - 2y \right] dy \right\}.$$

Existence of the solution and consistency of the above predictors are discussed as well. A special case of extrapolating heavy-tailed time series is considered in detail. Numerical examples predicting Gaussian, α–stable and autoregressive heavy–tailed stationary time series round up the talk. We explore several advantages of our excursion predictors using theoretical results and computational studies. Namely, they are computationally fast, weakly consistent for stochastically continuous random fields, and work for random fields without finite moments. In addition, they are robust with respect to the statistical estimation of the c.d.f. F_X. These results reveal a great potential for many real world applications.

References

[1] A. Das, V. Makogin, and E. Spodarev. Extrapolation of stationary random fields via level sets. *Theory of Probability and Mathematical Statistics*, 106:85–103, 2022.

[2] V. Makogin. Prediction of random time series via the excursion metric. R code, August 2022. `https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/forschung/Software/ExtrapolationGradient3.R`.

[3] V. Makogin and E. Spodarev. Prediction of random variables by excursion metric projections. Preprint, arXiv:2207.00447v2, September 2022. `https://doi.org/10.48550/arXiv.2207.00447`.

[4] M. D. Taylor. Separation metrics for real-valued random variables. *Internat. J. Math. Math. Sci.*, 7(2):407–408, 1984.

Expected Euler characteristic heuristic for smooth Gaussian random fields with inhomogeneous marginals

Satoshi Kuriki
The Institute of Statistical Mathematics
10-3 Midoricho, Tachikawa, Tokyo 190-8562, Japan
kuriki@ism.ac.jp

Abstract

Expected Euler characteristic (EC) heuristic is a method for approximating the tail probability of the maximum of a Gaussian random field. In this talk, we provide an expected Euler characteristic formula for the approximate tail probability and its relative approximation error when the index set M is a closed manifold and the mean and variance of the marginal distribution are not necessarily constant. When the variance is constant, [TTA05] proved that the relative approximation error is exponentially small in a general setting where the index set M is a stratified manifold. When the variance is not constant, it is shown that only the subset $M_{\rm supp}$ of M, referred to as the supporting index set, contributes to the maximum tail probability. The proposed tail probability formula is an integral of the Euler characteristic density over $M_{\rm supp}$, and its relative approximation error is proven to be exponentially small as in the case of constant variance. These results are generalizations of [KTT22], who addressed a restricted case of finite Karhunen-Loève expansion by the volume-of-tube method. As an example, the tail probability formula for the largest eigenvalues of noncentral Wishart matrices $\mathcal{W}_p(\nu, \Sigma; \Phi)$ and its relative approximation error are obtained. Numerical experience supports the high accuracy of the expected Euler characteristic formulas regardless of whether the marginals are homogeneous or inhomogeneous.

Keywords: Borel's inequality, Kac-Rice formula, noncentral Wishart distribution, volume-of-tube method, Weyl's tube formula.

References

[KTT22] Satoshi Kuriki, Akimichi Takemura, and Jonathan E. Taylor, *The volume-of-tube method for gaussian random fields with inhomogeneous variance*, Journal of Multivariate Analysis **188** (2022), 104819.

[TTA05] Jonathan E. Taylor, Akimichi Takemura, and Robert J. Adler, *Validity of the expected Euler characteristic heuristic*, Ann. Probab. **33** (2005), no. 4, 1362–1396.

Counterfactual Mean Embeddings

Motonobu Kanagawa
EURECOM, Sophia Antipolis, France

Joint work with
Krikamol Muandet (MPI Tuebingen), Sorawit Saengkyongam (U. Copenhagen),
Sanparith Marukatat (NECTEC)

To make a rational decision, a decision maker must be able to anticipate the effects of a decision to the outcomes of interest, before committing to that decision. For instance, before building a certain facility in a city, *e.g.*, a dam, policymakers and citizens must seek to understand its environmental effects. In medicine, a doctor has some prior knowledge about the effects a certain drug will have on a patient's health, before actually prescribing it. In business, a company needs to understand the effects of a certain strategy of advertisement to its revenue. One approach to addressing these questions is *counterfactual inference*.

Counterfactual inference we consider in this work consists of the following three main ingredients. Suppose that there exists a hypothetical subject (*e.g.*, a patient in medical treatment), and let X be *covariates* representing the features of the subject (*e.g.*, age, weight, medical record, etc.), T be a *treatment indicator* representing the treatment assigned to the subject (a drug of interest or a placebo), and Y be the *observed outcome* representing the post-treatment quantity of interest (*e.g.*, whether the patient is recovered or not). Given certain realizations of these variables $\{(\mathbf{x}_i, t_i, \mathbf{y}_i)\}_{i=1}^n$, in which each index i represents the identity of a subject, an analyst wishes to know how the treatment affects the outcome.

This problem is called counterfactual since for each subject i, we only observe the outcome $\mathbf{y}_i$ resulting from the assigned treatment t_i and can never observe the outcome (say $\mathbf{y}_i'$) that would have been realized under an alternative treatment $t_i' \neq t_i$. For example, if a patient receives an active treatment (*e.g.*, a drug of interest), we can never observe the outcome from the same patient under a control treatment (*e.g.*, a placebo). This is known as the fundamental problem of causal inference [9] and also as *bandit feedback* in the bandit literature [5]. One way to partially address this issue is a randomized experiment [6], in which treatments are randomly assigned to subjects. Although considered a gold standard, in practice randomization can be too expensive, time-consuming, or unethical. In most cases, therefore, analysis about treatment effects needs to be done on the basis of observational data $\{(\mathbf{x}_i, t_i, \mathbf{y}_i)\}_{i=1}^n$ in which the treatment assignment t_i may depend on covariates $\mathbf{x}_i$ and possibly on some hidden confounders; this setting is commonly known as *observational studies* [14, 17].

A fundamental framework for observational studies is the *potential outcome framework* [13, 16]. It provides a clear notation for *potential outcomes*, *i.e.*, the outcomes that would have been observed under different treatments, and elucidates the conditions required for making a valid inference about treatment effects. The framework has been studied extensively in statistics, and has a wide range of applications in biomedical and social sciences; see, *e.g.*, [10]. Moreover, important applications of machine learning such as off-policy evaluation for online advertisement and recommendation systems can be reformulated under this framework [18, 11].

In this work, we propose a novel approach to counterfactual inference that addresses the above challenges, which we term *counterfactual mean embedding* (CME). Our approach is built on kernel mean embedding [1, 19, 12], a framework for representing probability distributions as elements in a reproducing kernel Hilbert space (RKHS), so that each element representing a distribution maintains all of its information. We define an element representing a counterfactual distribution, for which we propose a nonparametric estimator. Notable advantages of the proposed approach are summarized as follows:

1. The proposed estimator can be computed based only on linear algebraic operations involving kernel matrices. Being a kernel method, it can be applied to not only standard domains (such as the Euclidean space), but also more complex and structured covariates and/or outcomes such

as images, sequences, and graphs, by using off-the-shelf kernels designed for such data [7]; this widens possible applications of counterfactual inference in general. Thus our work offers more flexibility than the existing approaches by [15] and [4], who focused on estimating the cumulative distribution functions of counterfactual distributions by assuming real-valued outcomes.

2. The proposed estimator can be used for computing a distance between the counterfactual and controlled distributions, thereby providing a way of quantifying the effect of a treatment to the distribution of outcomes; we define this distance as the maximum mean discrepancy (MMD) [2, 8] between the counterfactual and controlled distributions. It also provides a way to sample points from a counterfactual distribution based on kernel herding [3], a kernel-based deterministic sampling method.

3. The proposed estimator is nonparametric, and has theoretical guarantees. Specifically, we prove the consistency of the proposed estimator under a very mild condition, and derive its convergence rates under certain regularity assumptions involving kernels and underlying distributions. Both results hold without assuming any parametric assumption.

References

[1] Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, 2004.

[2] Karsten Borgwardt, Arthur Gretton, Malte Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):49–57, 2006.

[3] Y. Chen, M. Welling, and A. Smola. Super samples from kernel-herding. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, pages 109–116. AUAI Press, 2010.

[4] Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013.

[5] Miroslav Dudík, John Langford, and Lihong Li. Doubly Robust Policy Evaluation and Learning. In *Proceedings of the 28th International Conference on Machine Learning*, pages 1097–1104. Omnipress, 2011.

[6] R.A. Fisher. *The Design of Experiments*. Oliver and Boyd, 1935.

[7] Thomas Gärtner. A survey of kernels for structured data. *SIGKDD Explor. Newsl.*, 5(1):49–58, 2003.

[8] Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

[9] Paul Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[10] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York, NY, USA, 2015.

[11] Nathan Kallus and Angela Zhou. Policy evaluation and optimization with continuous treatments. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1243–1251, 2018.

[12] Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1-2):1–141, 2017.

[13] Jerzy Neyman. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes. Master's thesis, 7 1923. Excerpts reprinted in English, Statistical Science, Vol. 5, pp. 463–472. (D. M. Dabrowska, and T. P. Speed, Translators.).

[14] Paul Rosenbaum. *Observational Studies*. Springer Series in Statistics. Springer-Verlag, New York, 2nd edition, 2002.

[15] C. Rothe. Nonparametric estimation of distributional policy effects. *Journal of Econometrics*, 155:56–70, 2010.

[16] D.B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

[17] Donald Rubin. Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[18] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. Recommendations as treatments: Debiasing learning and evaluation. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48, pages 1670–1679, 2016.

[19] Alexander J. Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A Hilbert space embedding for distributions. In *Proceedings of the 18th International Conference on Algorithmic Learning Theory (ALT)*, pages 13–31. Springer-Verlag, 2007.

Central Limit Theorems in High-Dimensions: Recent Developments

Yuta Koike

The University of Tokyo and CREST JST

Let $X_1, \ldots, X_n$ be independent random vectors in $\mathbb{R}^d$ with mean 0 and covariance matrix Σ. Set

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

It is well-known by the central limit theorem (CLT) that the law of S_n can be approximated by $\mathrm{N}(0, \Sigma)$, the d-dimensional normal distribution with mean 0 and covariance matrix Σ, under some regularity assumptions, particularly when d is fixed. This approximation is ubiquitous in statistics because it gives a basis of asymptotically valid construction of confidence intervals and hypothesis testing. Recently, motivated by the growing interest in high-dimensional data analysis, it is of great interest to develop CLTs for S_n in high-dimensional settings such as $d = d_n$ tends to infinity as $n \to \infty$.

For statistical application, we need the following convergence for a sufficiently rich class $\mathcal{A}$ of Borel sets in $\mathbb{R}^d$:

$$\rho_n(\mathcal{A}) := \sup_{A \in \mathcal{A}} |\mathrm{P}(S_n \in A) - \mathrm{P}(Z \in A)| \to 0,$$

where $Z \sim \mathrm{N}(0, \Sigma)$. If $\mathcal{A} = \mathcal{C}_d$ is the class of convex Borel sets in $\mathbb{R}^d$, it is known that such convergence is generally impossible unless $d = o(n)$. In the pathbreaking work of Chernozhukov *et al.* (2013, 2017), they showed that one can prove $\rho_n(\mathcal{A}) \to 0$ under mild moment assumptions even when $d \gg n$ if $\mathcal{A} = \mathcal{R}_d$ is the class of rectangles in $\mathbb{R}^d$. Moreover, the class $\mathcal{R}_d$ is sufficiently rich in the sense that the convergence $\rho_n(\mathcal{R}_d) \to 0$ justifies many uniform inference procedures; see Chernozhukov *et al.* (2013), Belloni *et al.* (2018) and Chernozhukov *et al.* (2022a).

In the original work of Chernozhukov *et al.* (2017), the authors have shown that

$$\rho_n(\mathcal{R}_d) = O\left(\left(\frac{\log^7(dn)}{n}\right)^{1/6}\right)$$

under some moment assumptions, and this convergence rate was conjectured to be optimal with respect to the dependence on n. However, the recent work of Chernozhukov *et al.* (2022b) has showed that this rate can be improved to

$$\rho_n(\mathcal{R}_d) = O\left(\left(\frac{\log^5(dn)}{n}\right)^{1/4}\right) \tag{1}$$

under the same assumptions. This improvement poses the question of whether further improvement of this rate is possible, especially whether we can achieve the Berry–Esseen rate $n^{-1/2}$ while keeping the poly-log dependence on d. In this talk, we will review the recent progress in this problem.

The convergence rate (1) is obtained without assuming the invertibility of Σ, and it is known that a Berry–Esseen type bound is generally not achievable when Σ is degenerate (cf. Senatov (1986)). Therefore, it is natural to assume the invertibility of Σ in order to improve (1). In this direction, Chernozhukov *et al.* (2021) have shown that

$$\rho_n(\mathcal{R}_d) = O\left(\sqrt{\frac{\log^3 d}{n}}\log n\right)$$

when the smallest eigenvalue of Σ is lower bounded and coordinates of X_i are uniformly bounded. It is known that the rate $\sqrt{\frac{\log^3 d}{n}}$ gives a lower bound for the minimax optimal convergence rate for $\rho_n(\mathcal{R}_d)$. In the meantime, the recent work of Fang & Koike (2022) found that, if $X_1, \ldots, X_n$ are i.i.d. and log-concave, the rate $n^{-1/2}$ is achievable while keeping the poly-log dependence on d without any restriction on Σ. To be precise, they have shown that

$$\rho_n(\mathcal{R}_d) = O\left(\psi_d\sqrt{\frac{\log^3(dn)}{n}}\right),$$

where ψ_d is the so-called KLS constant, which is known to be (at most) of order $O(\log^{3.3336} d)$ and conjectured to be bounded by a universal constant (see Jambulapati *et al.* (2022)).

References

Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C. & Kato, K. (2018). High-dimensional econometrics and regularized GMM. Working paper. arXiv: 1806.01888.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41**, 2786–2819.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45**, 2309–2353.

Chernozhukov, V., Chetverikov, D., Kato, K. & Koike, Y. (2022a). High-dimensional data bootstrap. *Annu. Rev. Stat. Appl. (forthcoming)* .

Chernozhukov, V., Chetverikov, D., Kato, K. & Koike, Y. (2022b). Improved central limit theorem and bootstrap approximation in high dimensions. *Ann. Statist. (forthcoming)* .

Chernozhukov, V., Chetverikov, D. & Koike, Y. (2021). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab. (forthcoming)* .

Fang, X. & Koike, Y. (2022). Sharp high-dimensional central limit theorems for log-concave distributions. Working paper. Available at arXiv: `https://arxiv.org/abs/2207.14536`.

Jambulapati, A., Lee, Y. T. & Vempala, S. S. (2022). A slightly improved bound for the KLS constant. Working paper. Available at arXiv: `https://arxiv.org/abs/2208.11644`.

Senatov, V. V. (1986). Four examples of lower estimates in the multi-dimensional central limit theorem. *Theory Probab. Appl.* **30**, 797–805.

A Modelling Framework for Regression with Collinearity

Takeaki KARIYA , Nagoya University of Commerce and Business,
The corresponding author: thekariya70@gmail.com
Hiroshi KURATA[2] University of Tokyo, Graduate School of Arts and Sciences,
Takaki HAYASHI[3] Keio University, Graduate School of Business Administration,

This study addresses a fundamental, yet overlooked, gap between the standard theory and empirical practices in the OLS regression . To fill it, introducing a new concept "accommodation", this paper formulates a novel conceptual framework for developing our own model selection process in empirical modelling for given ($\boldsymbol{X}$) with collinearity in $\boldsymbol{X}$. With no use of **,** the new process enables us to find a class of effective and collinearity-resilient models. In fact, it directly controls not only the sampling variance of each OLSE, which includes Variance Inflation Factor, but also the individual power property of each *t*-test on regression coefficient, which includes what we call "Power Deflation Factor" as a collinearity factor. And to materialize our model selection process, two computational algorithms are proposed.

Consequently, it will provide an advance model-screening process and serve as an empirical platform for pre-selecting a class of effective models that well accommodate with both collinearity and inefficiency controlled in advance. In such a class of models we can freely use such statistical measures and procedures with use of as OLS estimation, *t*-value, coefficient of determination, stepwise model selection, etc.

More specifically, within the traditional OLS (Ordinary Least Squares) framework in the linear regression model,

(1.1) with and ,

we aim to formulate a conceptual framework for developing our model selection process, called XMOSEP, by connecting the OLS sampling theory with methodology of empirically effective modelling. Here, is assumed to be given, contain all possible variables for analysis and to be used as they are. As its vast literature shows, itself does not have a capacity to identify the collinearity structure of and avoid the ill-effect even if is really generated from it, though hybrid remedies have been provided. Our XMOSEP stands for X-based model selection process and with no use of **,** it enables us to obtain a class of empirically effective and collinearity-resilient models, which is called " -accommodating class of models", where $\boldsymbol{y}$ here simply denotes "dependent variable" symbolically. In the class $\boldsymbol{y}$ will come into its own capacity to enable empirically effective and reliable modelling via frequently used variable and model selection procedures (VMOSEPs) that use .

These VMOSEPs include such measures and procedures as those using *t*-value, *F*-value, adjusted coefficient of determination and so on. In fact, without controlling strong collinearity in advance, those procedures will not lead us to empirically effective models in practice within the OLS framework. While, whether or not is used in the model selection, VMOSEPs and XMOSEP, which are called commonly MOSEPs, are regarded as a process of replacing in (1.1) by a "better" submatrix where data ($\boldsymbol{X}$) is given. Our XMOSEP connects the sampling variances of individual OLSEs in estimation and the power performances of individual *t*-tests with a practical process of

making empirically effective modelling, so that standard errors of estimates and power of *t*-tests are controlled against collinearity and inefficiency, where inefficiency is implied by large individual sampling variances (IndSVs).

In the OLS framework, as is well known, if the initial model in (1.1) is supposedly "true" with rank($\boldsymbol{X}$) =K, is the best linear unbiased estimator (BLUE) in the nonnegative definite ordering, in which no shrinkage-type estimator will beat the OLS estimator in risk matrix. This optimality holds no matter how strong the collinearity in $\boldsymbol{X}$ may be, implying no collinearity effect on this basic optimality of . However, in empirical modelling, it is often the case that a final model is selected via a MOSEP together with various diagnostics including *t*- or *F*- tests and variance analysis. Then is not only different from $\boldsymbol{X}$ (1.1) but also it may depend on analyst's view. The final OLS model thus selected faces a conflict in claiming its BLUE-ness and effectiveness of the diagnostic results, because is not $\boldsymbol{X}$. In Section 3, this problem will be overcome by defining the model in (1.1) as a "bundle model". After all, a finally selected model via a MOSEP will have to be regarded as the model having generated $\boldsymbol{y}$ in any empirical analysis, so long as $\boldsymbol{y}$ is regressed on .

Key words: OLS, model selection process, collinearity effect, empirically modelling, *t*-test

Kariya and Hayashi' s portion of this work was supported by JSPS KAKENHI Grant Number 21K01431. Kurata's portion is partially supported by JSPS KAKENHI Grant Number 19K11853.

Risk and Statistics,3rd Tohoku-ISM-UUlm Joint Workshop.　October 13. 2022

Occupation time, Quantiles and Rank on Vasicek process with applications to Exotic options.

Ryozo Miura.　　Professor Emeritus, Hitotsubashi University

Abstract.

A simple Ornstein -Uhlenbeck (OU)process

$$dU_t = \lambda(\theta - U_t)dt + \sigma dW_t, \quad U_0 = a$$

is called Vasicek process in Quantitative Finance. It is used to describes the behavior of an interest rate.

Occupation time was a well-known quantity since earlier. Quantiles for Geometric Brownian Motion process were defined by Miura in (2) and Ranks by Fujita and Miura in (8).Since then, they were studied in several papers (3) (4) (6) (9) and others not listed here. The three quantities correspond to the empirical distribution functions, the order statistics and the ranks in the theory of mathematical statistics. From the same viewpoint, Brownian motion is a continuous time version of CUSUM (cumulative sums of observations).

Now the stochastic process we work on is a simple OU process and will present our trial to obtain the probability distribution of the three quantities for Vasicek process X_t.

$$F(K) = \frac{1}{T}\int_0^T I\{Xu \le K\}du\,, \qquad m(\alpha,:[0,T],Xt) = \inf\{x: \alpha \le F(x)\}$$

R(t, X)=$\frac{1}{T}\int_0^T I\{Xu \leqq Xt\}$du, t∈[0, T]

The quantile m(α) and Rank (t) are directly related to occupation time F(K) in such a way: **F(m(α))=α, hence {F(K)> α }={m(α)<K} and $F(X_t)$=R(t),** so that we will first focus on occupation time T・F(K).　（Note that F(K) is rather “percentage converted” occupation time.）

Our work is on-going stage now, but we will show the framework of approaches for deriving the probability distributions, where Feynman-Kac formula and Kac formula described in (1) bring us to obtain Laplace transform of the density functions.

Our first approach uses the first hitting time to the level K and occupation time below K after hitting the level. F(K) can be expressed in terms of Brownian motion (or, via Change of Measure from the Vasicek process). It worked well to find the probability density function of the first hitting time and occupation time for (Geometric) Brownian motion process. But for OU process, it is more complicated. We end up to need the joint distribution of

$\left(W_t, \int_0^t 1_{\{W_s \le 0\}}ds, \int_0^t W_s ds, \int_0^t W_s^2 ds\right)$, which requires a large amount of calculations (not completed yet).

The first attempt to obtain the probability density of first hitting time for OU process was

made in 1998 and corrected two years later in (5) by the same authors. Then it had a small error which was pointed out and corrected by (7), (however, some ambiguity remains to me for the case b not equal to zero).

Another approach is to work directly on Laplace transform of the probability distributions of F(K), without setting the first hitting time. The results from (10) may be applied (my tentative, but not the final view).

We will also briefly explain that similar approaches can be applied to study Ranks R(t).

As for an application to Quantitative Finance, the stochastic corridor which is an Exotic option based on Rank described in (9) may be an interesting candidate, while the other options based on these quantities are also possible, the design of which can be seen in (6) for the case of (Geometric) Brownian Motion.

References.

: (1) Karatzas, I. and Shreve, S. E. (1988). Brownian Motion and Stochastic Calculus. Springer-Verlag. New York.

: (2) Miura, R. (1992). A note on look-back option based on order statistics. Hitotsubashi Journal of Commerce and Management. 27, 15-28.

: (3) Akahori, J. (1995). Some formulae for a new type of path-dependent option. Annals of Applied Probability. 5, 383-388.

: (4) Yor, M. (1995). The Distribution of Brownian Quantiles. Journal of Applied Probability. 32. , 405-416.

: (5) Leblanc, B., Renault, O. and Scailet, O. (2000). A Correction Note on the First Passage time of an Ornstein-Uhlenbeck process to a Boundary. Finance and Stochastics 4, 109-111.

: (6) Fujita, T. and Miura, R. (2002). Edokko Options: A New framework of Barrier Options. Asia Pacific Financial Markets, 9(2), 141-151.

: (7) Going-Jaeschke, A. and Yor, M. (2003). A clarification note about hitting times densities for Ornstein-Uhlenbeck processes. Finance and Stochastics, 7, 413-415.

: (8) Fujita, T. and Miura, R. (2006). The Distribution of Continuous time Rank Processes. Advances in Mathematical Economics. 9, 25-32.

: (9) Miura, R. (2007). Rank Process, Stochastic Corridor and Applications to Finance. Advances in Statistical Modeling and Inference. 529-542.

: (10) Li, B. and Zhou, X. (2013). The Joint Laplace Transforms for Diffusion Occupation times. Advances in Applied Probability. 45, 1049-1067.

SPATIAL REGRESSION DISCONTINUITY DESIGNS[1,2]

TAKUYA ISHIHARA[a], DAISUKE KURISU[b], YASUMASA MATSUDA[a] AND MASAYUKI SAWADA[c]

We propose a mean squared error optimal estimator for regression discontinuity (RD) designs with a vector of running variables. For the estimator, we show the asymptotic normality of a multi-dimensional local polynomial estimator with dimension-specific bandwidths. We illustrate our results for two-dimensional cases. Consider a non-parametric regression model

$$Y_i = m(X_i) + \epsilon_i, E[\epsilon_i|X_i] = 0, i = 1, \ldots, n, Y_i \in \mathbb{R} \text{ and } X_i = (X_{i,1}, X_{i,2})' \in \mathbb{R}^2$$

for a sequence of i.i.d. random vectors $\{(Y_i, X_i)\}_{i=1}^n$. A pair of running variables $X_i = (X_{i1}, X_{i2})'$ is a geographical location on a map. The location X_i determines treatment D_i. Specifically, $D_i = 1\{X_i \in \mathcal{T}\}$ where $\mathcal{T}$ is a subset of the support of X_i.

Let c be a particular point on the boundary of $\mathcal{T}$, our target parameter is $\theta := \lim_{x\to c, x\in\mathcal{T}} m(x) - \lim_{x\to c, x\in\mathcal{T}^C} m(x)$. Under appropriate conditions (Hahn et al., 2001, Keele and Titiunik, 2015, for example), θ is the average treatment effect for the units at the boundary point c.

Local linear estimation such as Calonico et al. (2014) is dominant for RD designs because of its intuitive construction from the identification strategy that compares units around the boundary. However, existing local linear estimators are limited to a uni-variate running variable, and multi-dimensional case is available only in a non-kernel procedure such as Imbens and Wager (2019). Consequently, empirical practices often imposes that $m(\cdot)$ is a function of the Euclidean distance from a point or the boundary. We provide a local linear estimator without such a strong restriction.

Consider the local linear estimator $\hat{\beta}^+ = (\hat{\beta}_0^+, \hat{\beta}_1^+, \hat{\beta}_2^+)'$ solving the following problem

$$\hat{\beta}^+ = \underset{(\beta_0,\ldots,\beta_2)'\in\mathbb{R}^3}{\arg\min} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1(X_{i,1} - c_1) - \beta_2(X_{i,2} - c_2))^2 K_h(X_i - c) 1\{X_i \in \mathcal{T}\}$$

where $K_h(X_i - c) = K\left(\frac{X_{i,1}-c_1}{h_1}, \frac{X_{i,2}-c_2}{h_2}\right)$ and each h_j is a sequence of positive constants (bandwidths) such that $h_j \to 0$ as $n \to \infty$. Similarly, let $\hat{\beta}^-$ be the estimator using $1\{X_i \in \mathcal{T}^c\}$

[1]

[2]Prepared for 3rd TIU workshop on 2022 October 12 through 14.

[a]Graduate School of Economics and Management, Tohoku University

[b]Graduate School of International Social Sciences, Yokohama National University

[c]Institute of Economic Research, Hitotsubashi University

subsample. Our spatial RD estimator is $\hat{\beta}_0^+ - \hat{\beta}_0^-$.

Unlike Masry (1996) who study the asymptotic property of the local polynomial estimator with a common bandwidth $h_1 = h_2$, we show the asymptotic normality of the local polynomial estimator with heterogeneous bandwidth. Furthermore, following Imbens and Kalyanaraman (2012), we propose a procedure to choose the pair of bandwidths $h_1 \neq h_2$ by minimizing the asymptotic mean squared error at the cutoff normalized to $c = 0$ without loss of generality.

$$\{h_1^2(\partial_{11}m^+(0)\mathcal{K}_{11}^+ - \partial_{11}m^-(0)\mathcal{K}_{11}^-) + h_2^2(\partial_{22}m^+(0)\mathcal{K}_{22}^+ - \partial_{22}m^-(0)\mathcal{K}_{22}^-) + h_1h_2(\partial_{12}m^+(0)\mathcal{K}_{12}^+ - \partial_{12}m^-(0)\mathcal{K}_{12}^-)\}^2 + \frac{1}{nh_1h_2}\frac{\sigma^2(0)}{f(0)}e_1'S^{-1}\mathcal{K}S^{-1}e_1$$

where $e_1 = (1,0,0)'$, $\partial_{lk}m^+(0) = \lim_{x\to 0, x\in\mathcal{T}}\partial_{lk}m(x)$, $\partial_{lk}m^-(0) = \lim_{x\to 0, x\in\mathcal{T}^C}\partial_{lk}m(x)$, $S, \mathcal{K}$ and $\mathcal{K}_{11}^+, \mathcal{K}_{11}^-, \mathcal{K}_{12}^+, \mathcal{K}_{12}^-, \mathcal{K}_{22}^+$ and $\mathcal{K}_{22}^-$ are constants determined by the kernel $K(\cdot)$, $\sigma^2(0)$ is the conditional variance of the error term at the cutoff, and $f(0)$ is the density at the cutoff.

However, for a particular shape of m, there exists a pair (h_1, h_2) such that the asymptotic bias is zero, and consequently the bandwidths cannot be determined. For a product kernel $K(x) = K_1(x_1)K_2(x_2)$, we resolve this issue by *rotating* the axis so that K_2 becomes the one-sided triangular kernel $K_2(x) = 2(1-x)1\{0 \leq x \leq 1\}$. In Monte Carlo simulation, we demonstrate that our spatial RD estimator outperforms the conventional RD estimator with the shape restriction. We apply our procedure to Ehrlich and Seidel (2018) data for a spatial RD evaluation of a place-based policy.

REFERENCES

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

Ehrlich, M. v. and T. Seidel (2018): "The Persistent Effects of Place-Based Policy: Evidence from the West-German Zonenrandgebiet," *American Economic Journal: Economic Policy*, 10, 344–374.

Hahn, J., P. Todd, and W. V. d. Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

Imbens, G. and K. Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *The Review of Economic Studies*, 79, 933–959.

Imbens, G. and S. Wager (2019): "Optimized Regression Discontinuity Designs," *The Review of Economics and Statistics*, 101, 264–278.

Keele, L. J. and R. Titiunik (2015): "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 23, 127–155.

Masry, E. (1996): "Multivariate regression estimation local polynomial fitting for time series," *Stochastic Processes and their Applications*, 65, 81–101.

Shrinkage Methods for Treatment Choice

Takuya Ishihara Daisuke Kurisu

This paper studies the problem of determining whether or not to treat individuals based on observed covariates. In this problem, the most common decision rule is the conditional empirical success (CES) rule proposed by Manski (2004), which is a rule assigning individuals to treatments that yield the best experimental outcomes conditional on observed covariates. The CES rule uses only the average treatment effect (ATE) estimate conditional on each covariate value. In contrast, a common method in statistical estimation problem is to shrink unbiased but noisy preliminary estimates toward the average of these estimates and it is well known that shrinkage estimators have smaller mean squared error than unshrunk estimators. In this study, we assume that the dispersion of the conditional ATEs (CATEs) is bounded and propose the shrinkage rule that is a rule assigning individuals to treatments based on shrinkage estimators. We select the shrinkage factor by minimizing an upper bound of the maximum regret. By considering the treatment rule for individuals that uses not only each CATE but also the CATEs of other individuals, it is possible to incorporate information across individuals. This allows the proposed shrinkage rule to perform as well as or better than existing treatment rules in the sense of regret and to be more flexible to the heterogeneity of the individuals. In addition, we compare the shrinkage rule with other rules when the parameter space is correctly specified and misspecified.

The contributions of this paper are threefold. First, our approach is attractive from a computational point of view. Computation of the exact minimax regret rule often becomes challenging in the context of statistical treatment choice. Indeed, when the parameter space is restricted and the number of the possible covariate values is large, it is difficult to obtain the shrinkage rule that minimizes the maximum regret. To overcome this problem, we propose the shrinkage rule that minimizes tractable upper bound of the maximum regret. Because in this approach, each shrinkage factor is obtained by optimizing over a single parameter, the proposed shrinkage rule is easy to compute.

Second, we compare the maximum regrets of the shrinkage, CES, and pooling rules when the parameter space is correctly specified. As an alternative to the CES and shrinkage rules, one could consider using the pooling rule that determines whether or not to treat the individuals based on the average of the CATE estimates. Because the CES and pooling rules are special cases of shrinkage rules, the proposed shrinkage rule is expected to outperform these two rules. However, because the proposed shrinkage rule does not minimize the exact maximum regret, its maximum

regret may be larger than that of the CES and pooling rules. Hence, it is important to compare these maximum regrets. If the dispersion of the standard errors of estimated CATEs is small compared to the dispersion of the CATEs, then the proposed shrinkage rule has a smaller maximum regret than the CES rule. We also show that the maximum regret of the shrinkage rule is always less than twice that of the pooling rule. Furthermore, when the dispersion of CATEs is sufficiently small or large, the shrinkage rule is no worse than the pooling rule.

Third, we evaluate the maximum regret of the shrinkage rule when the parameter space is misspecified. Because the minimax decision rule depends on the parameter space, the choice of the parameter space is important in practice. For example, Armstrong and Kolesar (2018) and Armstrong and Kolesar (2021) consider the minimax estimation and inference problem for the treatment effects and show that it is not possible to choose the parameter space automatically in a data-driven way. Hence, it is important to analyze the decision rule under misspecification of the parameter space. We investigate the performance of the shrinkage rule and show that our results are robust against misspecification of the parameter space. To the best of our knowledge, this is the first study that considers misspecification of the parameter space in the treatment choice problem.

References

- Armstrong, Timothy B., and Michal Kolesár. "Optimal inference in a class of regression models." *Econometrica* 86.2 (2018): 655-683.
- Armstrong, Timothy B., and Michal Kolesár. "Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness." *Econometrica* 89.3 (2021): 1141-1177.
- Manski, Charles F. "Statistical treatment rules for heterogeneous populations." *Econometrica* 72.4 (2004): 1221-1246.

Forecasting Stock Returns with Conditional Quantile-level Dependence

Stanley Iat-Meng Ko* Sung Y. Park†

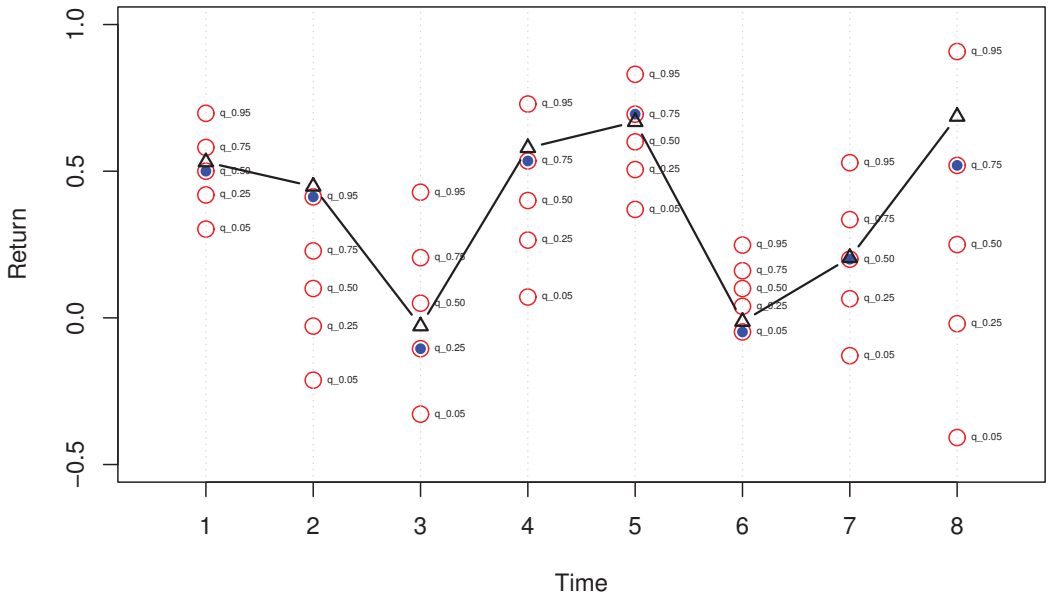
Finance practitioners and researchers have long attempted to forecast the equity premium, which is essential for asset pricing, portfolio construction and so on. Rozeff (1984), Fama and French (1988) and Campbell and Shiller (1988*a*) formally study the predictability of the dividend-price ratio for stock returns. Since then, various predictors have been developed and studied; for example, the dividend-earnings ratio (Campbell and Shiller, 1988*b*; Lamont, 1998), yields on treasury and corporate bonds (Keim and Stambaugh, 1986; Campbell, 1987; Fama and French, 1989), and book-to-market ratio (Kothari and Shanken, 1997; Pontiff and Schall, 1998), among many others. Welch and Goyal (2008) comprehensively summarizes various stock return predictors and extensively examines their forecasting performance. They conclude that most predictors forecast poorly both in-sample and out-of-sample. Moreover, the prevailing mean model (average historical return) simply provides better forecasts; see also Butler, Grullon and Weston (2005). Recent studies have utilized financial economic theory to improve the forecastability of predictors. Using the same set of predictors, Campbell and Thompson (2008) and Pettenuzzo, Timmermann and Valkanov (2014) propose a constrained forecasting method in which the restrictions on mean and variance suggested by investment theory are imposed.

In this paper, we propose a novel approach to forecasting stock excess returns using conditional quantile levels. We define the "optimal forecasting quantile level" as the corresponding conditional quantile value that is the closest to the true y_t. To illustrate the idea, we consider a hypothetical time series and its conditional quantiles at $t = 1, \ldots, 8$ in Figure 1. For illustrative purposes, we only consider five quantiles, at $0.05, 0.25, 0.5, 0.75,$ and 0.95. The red circles denote the quantiles of returns at different times. The realized returns are plotted as triangles at each time. The quantile levels with blue dots are the closest to the true y_t, which we call the "optimal forecasting quantile level", $\tilde{\alpha}_t$. In this example, the optimal forecasting quantile-level sequence or trajectory is thus $0.5 \to 0.95 \to 0.25 \to 0.75 \to 0.75 \to 0.05 \to 0.5 \to 0.75$.

*Graduate School of Economics and Management, Tohoku University, Sendai, Japan. Tel: +81 022-795-6288. Email: stanleyko@tohoku.ac.jp.

†School of Economics, Chung-Ang University, Seoul, Korea. Tel: +82-2-820-5622. Email: sungpark@cau.ac.kr.

Figure 1: Hypothetical optimal forecasting quantile levels.



We discover an interesting phenomenon that the $\tilde{\alpha}_t$ sequence forms a stationary time series with stable autocorrelation. Specifically, there is significant lag-5 autocorrelation in $\tilde{\alpha}_t$ of the monthly equity premium of S&P 500, and significant lag-7 autocorrelation in the quarterly data. However, the autocorrelation of $\tilde{\alpha}_t$ disappears in the annual data. This provides us a way to forecast excess return by exploring the time series property of the optimal forecasting quantile levels. Typically, we first forecast the future optimal forecasting quantile level, based on the past optimal levels; for example, for the monthly return, $\hat{\tilde{\alpha}}_t = \hat{\gamma}_0 + \hat{\gamma}_1 \tilde{\alpha}_{t-5}$. Then we forecast the excess return using quantile regression at $\hat{\tilde{\alpha}}_t$; that is, $\hat{y}_{t|t-1} \equiv \hat{q}_{\hat{\tilde{\alpha}}_t, t+1|t}$. We show that a time varying risk premium adjustment is implied by our method. In the empirical study, we show that the proposed two-step approach outperforms other methods in forecasting the equity premium. Our empirical results demonstrate the superiority of our method for forecasting the monthly and quarterly equity premium. Using the conventional out-of-sample R^2 performance measure, we show extraordinary robust predictive power for our method in different sample periods. Our method also achieves better prediction performance under the alternative absolute loss measure.

3rd Tohoku-ISM-UUlm workshop

Issues in Spatial Processes with Long Range Dependence

Peter M Robinson

London School of Economics

Abstract

We discuss issues of statistical inference for spatial processes with 'long range dependence'. Long range dependence, or strong dependence, in time series is a topic that has been quite extensively studied in recent years. After a number of probabilistic contributions and empirical studies, serious treatment of issues of statistical inference could be said to have begun in the mid-1980s, with activity then increasing through the 1990s and this century. Predominately this literature has focussed on observations that are regularly-spaced over time, and the bulk of the theoretical development has been in terms of asymptotic statistical theory, with the number of observations regarded as diverging, finite-sample theory proving mathematically intractable, even under the precise distributional assumptions that are typically not required in a large-sample treatment.

Parametric, semiparametric and nonparametric models have all featured. The major characteristic feature of a long range dependent covariance stationary time series process is that its autocovariance function decays so slowly with increase in lag length as not to be summable, or, nearly equivalently, its spectral density diverges, typically at zero frequency, while some nonstationary processes, such as ones with a unit root can, a fortiori, be regarded as having even longer memory. By contrast, 'short range dependent' time series typically have autocovariances that are summable and spectral density that is more or less smooth (for example a stationary autoregressive moving average (ARMA) has exponentially decaying autocovariance and analytic spectral density), though for some relevant purposes a short range dependent process is sometimes defined as merely having spectral density that is positive and finite at zero frequency.

Spatial data have long attracted the attention of statisticians, and the configurations of some such data, especially some arising in such fields as meteorology, cosmology and agriculture, can be viewed as generalisations of the typical regularly-spaced time series one mentioned above. In particular, they constitute observations in $M \geq 2$ dimensions, (where $M=1$ in the time series case).

This brings to mind regularly spaced agricultural plots on a field. In fact, `spatial long range dependence' goes back at least to the agriculturally-motivated paper of Smith (1938), which is also a very early reference relative to the literature on time series long range dependence. It is interesting that Smith (1938) thought of a power law decay: this might seem natural to one coming to the subject unschooled in time series modelling which, after World War II, stressed exponential decay, as in ARMA modelling.

Since then, many papers on 'spatial long range dependence' have appeared, but the topic has not been developed as systematically or comprehensively as 'long range dependent time series'. Some distinctive issues arising in the 'spatial' case, all of which have been studied far more under short range dependence than long range dependence, are:

1. Is there isotropy or not? If not, we might model each dimension separately, or with some interaction, and possibly with a different memory parameter for each dimension.

2. Is sampling regularly or irregularly spaced? Whereas in time series regular spacing has been predominately studied, irregular spacing is perhaps more likely to be found with spatial data.

3. Should modelling be unilateral or multilateral? For time series unilateral modelling, reflecting one-sided transition from past to future, is usually natural, but this is not the case with spatial data, where, for example, the dimensions might be latitude and longitude.

4. The edge effect. In estimating lagged quantities there is loss of information at the boundary of the observation region, which has negligible effect when M=1, but increasing, and damaging, effect as M increases unless corrected for.

We discuss the following topics:

1. Inference on location and regression with long range dependent errors.

2. Inference on second-order properties of long range dependent stationary processes.

3. Miscellaneous topics: nonstationary processes, irregular spacing, adaptive estimation, nonparametric regression.

We do not consider 'spatial autoregressive'-type ('SAR') models (which depend on a user-chosen weight matrix of geographic or economic inverse distances); these do not fit into our framework and typically possess a kind of short range dependence.

On estimation of
fractional Browninan fields and sheets

Yoshihiro Yajima
University of Tokyo(Emeritus Professor)
at Tohoku University Oct. 12-14

Abstract

Originally a Gaussian semiparametric estimator is an approximate likelihood estimator in a frequency domain for long memory models of stationary and nonstationary time series(Robinson(1995); Velasco(1999)). We apply it to fractional Brownian fields(FBF) and Brownian sheet(FBS) observed on a regular grid and derive its asymptotic properties. They are Gaussian random fields and used to model many physical processes in space. The estimator is consistent and has the limiting normal distribution as the sample size goes to infinity.

Keywords:spatio-temporal models; fractional Brownian field; fractional Brownian sheet.

1. Introduction FBF is a popular model for intrinsic stationary random fields(ISRFs). Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \boldsymbol{R}^d\}$ be a random field. If $\{X(\boldsymbol{s})\}$ satisfies that for any fixed $\boldsymbol{h}(\in \boldsymbol{R}^d)$, the increment $Z_{\boldsymbol{h}}(\boldsymbol{s}) = X(\boldsymbol{s}+\boldsymbol{h}) - X(\boldsymbol{s})$ is a stationary random field, $\{X(\boldsymbol{s})\}$ is called an ISRF. Then $\{X(\boldsymbol{s})\}$ is characterized by

$$\begin{aligned} E(X(\boldsymbol{s}+\boldsymbol{h}) - X(\boldsymbol{s})) &= 0, \\ \mathrm{Var}(X(\boldsymbol{s}+\boldsymbol{h}) - X(\boldsymbol{s})) &= 2\gamma(\boldsymbol{h}), \end{aligned}$$

where $2\gamma(\boldsymbol{h})$ is the variogram function(see Chilès & Delfiner (2012); Cressie (1993)). Hereafter we also assume that $X(\boldsymbol{0}) = 0$.
For $\boldsymbol{\lambda}$ and $\boldsymbol{h}$, let $(\boldsymbol{\lambda}, \boldsymbol{h})$ be the inner product and $\| \boldsymbol{\lambda} \|$ be the norm.
Then if $2\gamma(\boldsymbol{h})$ is a continuous function on $\boldsymbol{R}^d$ satisfying $\gamma(\boldsymbol{0}) = 0$, it has the spectral representation

$$2\gamma(\boldsymbol{h}) = \int_{\boldsymbol{R}^d} \frac{1 - \cos((\boldsymbol{\lambda}, \boldsymbol{h}))}{(2\pi)^d} G(d\boldsymbol{\lambda}) + Q(\boldsymbol{h}), \tag{1}$$

where $Q(\boldsymbol{h})(\geq 0)$ is a quadratic form and $G(\boldsymbol{\lambda})$ is a positive, symmetric measure such that $\| \boldsymbol{\lambda} \|^2 \, G(\boldsymbol{\lambda})$ is continuous at the origin and

$$\int_{\boldsymbol{R}^d} \frac{\| \boldsymbol{\lambda} \|^2}{1 + \| \boldsymbol{\lambda} \|^2} G(d\boldsymbol{\lambda}) < \infty. \tag{2}$$

(See Chilès& Delfiner (2012); Cressie(1993); Solo (1992); Yaglom (1957).)
Hereafter we assume that $Q(\boldsymbol{h}) \equiv 0$ and $G(\boldsymbol{\lambda})$ is absolutely continuous with density $g(\boldsymbol{\lambda})$. Then (1) and (2) reduce to

$$2\gamma(\boldsymbol{h}) = \int_{\boldsymbol{R}^d} \frac{1 - \cos((\boldsymbol{\lambda}, \boldsymbol{h}))}{(2\pi)^d} g(\boldsymbol{\lambda}) d\boldsymbol{\lambda},$$

and

$$\int_{\boldsymbol{R}^d} \frac{\| \boldsymbol{\lambda} \|^2}{1 + \| \boldsymbol{\lambda} \|^2} g(\boldsymbol{\lambda}) d\boldsymbol{\lambda} < \infty,$$

respectively.
An interesting special class of ISRF's that is often applied to empirical data analysis in space is a fractional Brownian field(FBF)(see Adler (1981); Mandelbrot& Van Ness (1968);Yaglom (1957); Zhu& Stein (2002) and the references therein). FBF has $2\gamma(\boldsymbol{h}) = C \| \boldsymbol{h} \|^{2H}$, which is equivalent to

$$g(\boldsymbol{\lambda}) = \frac{CHK_H}{\| \boldsymbol{\lambda} \|^{d+2H}},$$

where

$$K_H = \pi^{d/2} 2^{2H+d} \Gamma((d+2H)/2) / \Gamma(1-H), \ 0 < H < 1,$$

and C is a scale parameter and H is a smoothness parameter with larger values corresponding to smoother surfaces.
The other popular random field is FBS. FBS is a Gaussian random field and its covariance is defined by

$$\mathrm{Cov}(X(\boldsymbol{s}), X(\boldsymbol{t})) = C\prod_{i=1}^{d}(s_i^{H_i} + t_i^{H_i} - |s_i - t_i|^{H_i}), 0 < H_i < 1.$$

where $\boldsymbol{s} = (s_1, \cdots, s_d)'$ and $\boldsymbol{t} = (t_1, \cdots, t_d)'$.
2.Theoretical results Hereafter for simplicity of calculation and notation we assume that $d = 2$. For FBF, we also denote $\boldsymbol{\lambda}$ by (λ_1, λ_2) and $g(\boldsymbol{\lambda})$ by $g(\lambda_1, \lambda_2)$ respectively.
First we consider FBF.

Assumption 1 *$g(\lambda_1, \lambda_2)$ is expressed by*

$$g(\lambda_1, \lambda_2) = \| \boldsymbol{\lambda} \|^{-2H-2} g_o(\lambda_1, \lambda_2), \quad 0 < H < 1,$$

where $g_o(\lambda_1, \lambda_2)$ is a nonnegative with $g_o(0, 0) > 0$, symmetric , $g_o(\lambda_1, \lambda_2) = g_o(-\lambda_1, -\lambda_2)$, twice continuously differentiable function for $-\infty < \lambda_1, \lambda_2 < \infty$ and is bounded with bounded first and second order partial derivatives.

Theorem 1 *Under Assumption 1 and additional conditions, GSE $\hat{H}_n$ converges to H_0(the true parameter) in probability as $n \to \infty$ and for $p \geq 2$, $m^{1/2}(\hat{H}_n - H_0)$ converges to $N(0, 1)$ in distribution as $n \to \infty$ where m is the number of the discrete Fourier transforms used to construct the estimator and p is the order of the data taper..*

Next we have the following result for FBS.

Theorem 2 *Assume $p \geq 3$. Under some assumptions, GSE $\hat{H}_{in}(i = 1, 2)$converges to H_{i0}(the true parameters) in probability as $n \to \infty$ and $m(\hat{H}_{1n} - H_{10}, \hat{H}_{2n} - H_{20})'$ converges to $N(\mathbf{0}, I_2)$ in distribution as $n \to \infty$ where m^2 is the number of the discrete Fourier transforms used to construct the estimator.*

References

Adler,R.J. (1981). *The Geometry of Random Fields.*Wiley, New York.

Chilès, J.-P. and Delfiner,P. (2012). *Geostatistics :Modeling Spatial Uncertainty.* 2nd ed. Wiley, New York.

Cressie, N.A.C. (1993). *Statistics for Spatial Data.* Revised ed. Wiley, New York.

Mandelbrot, B.B. and Van Ness, J.W. (1968). Fractional Brownian motion, fractal noises and applications. *SIAM. Rev.* **10** 422-437.

Robinson,P.M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630-1661.

Stein,M.L. (1999). *Interpolation of Spatial Data:Some Theory for Kriging.* Springer, New York.

Solo,V. (1992). Intrinsic stationary random functions and the paradox of $1/f$ noise. *SIAM J. Appl. Math.* **52** 270-291.

Velasco,C. (1999). Gaussian semiparametric estimation of non-stationary time series. *J. Time Ser. Anal.* **20** 87-127.

Yaglom,A.M. (1957). Some classes of random fields in n-dimensional space, related to stationary random processes. *Theory Probab. Appl.* **2** 273-320.

Zhu,Z. and Stein,M.L. (2002). Parametric estimation for fractional Brownian surfaces. *Statist. Sinica***12** 863-883.

Joint circular distributions in view of higher order spectra of time series and copula

Masanobu Taniguchi* Waseda University,Tokyo,
Shogo Kato, Institute of Statistical Mathematics,Tokyo,
Hiroaki Ogata, Tokyo Metropolitan University,
Arthur Pewsey, University of Extremadura, Spain

Abstract

Circular data analysis is emerging as an important component of statistics. For this half century, various circular distributions have been proposed, e.g., von Mises distribution, wrapped Cauchy distribution, among other things.
Also, regarding the joint distribution, Wehrly and Johnson(1980) proposed a bivariate circular distribution which is related to a family of Markov processes on the circle. Because the sample space is on a circle, various new statistical methods have been developed. In this talk we provide a new look at circular distributions in view of spectral distributions of time series because the typical circular distributions correspond to spectral densities of time series models. For example, autoregressive AR(1) spectral density corresponds to wrapped Cauchy distribution, and von Mises distribution corresponds to exponential spectral density (Bloomfield(1973)), etc. Furthermore we introduce a class of joint circular distributions from the higher order spectra of time series, which can describe very general joint circular distributions. Hence we can develop the statistical inference for dependent observations on the circle. We present a family of distributions on the circle derived from the ARMA spectral density. It is seen that the proposed family includes some existing circular families as special cases. For these special cases, the normalizing constant and trigonometric moments are shown to have simple and closed form. We develop the asymptotic optimal inference theory based on the local asymptotic normality (LAN) on the circle. Because the observations are permitted to be dependent, the theory opens a new paradigm in the estimation for joint circular distributions. Because we introduced very general joint circular distributions, we can discuss the problem of copula for them.