International Symposium on Theories and Methodologies for Large Complex Data

November 21-23, 2019

Venue:

Conference Room 406, Tsukuba International Congress Center
2-20-3 Takezono, Tsukuba, Ibaraki 305-0032, Japan

Organizers:

Makoto Aoshima	(University of Tsukuba)
Mika Sato-Ilic	(University of Tsukuba)
Kazuyoshi Yata	(University of Tsukuba)
Aki Ishii	(Tokyo University of Science)

Supported by

Grant-in-Aid for Scientific Research (A) 15H01678 (Project Period: 2015-2019)
"Theories and methodologies for large complex data"
(Principal Investigator: Makoto Aoshima)

Grant-in-Aid for Challenging Research (Exploratory) 19K22837 (Project Period: 2019-2021)
"Tackling individualized modeling with ultra-high dimensional data"
(Principal Investigator: Makoto Aoshima)

Program

November 21 (Thursday)

14:00∼14:10 Opening

14:10∼14:50 Aki Ishii[*,a], Kazuyoshi Yata[b] and Makoto Aoshima[b]
[a](Department of Information Sciences, Tokyo University of Science)
[b](Institute of Mathematics, University of Tsukuba)

Tests for high-dimensiomal covariance structures under the SSE model

15:00∼15:40 Takahiro Nishiyama[*,a], Masashi Hyodo[b] and Tatjana Pavlenko[c]
[a](Department of Business Administration, Senshu University)
[b](Department of Mathematical Sciences, Osaka Prefecture University)
[c](Department of Mathematics, KTH Royal Institute of Technology)

On error bounds for high-dimensional asymptotic distribution of L_2-type test statistic

15:55∼16:35 Hiroumi Misaki (Faculty of Engineering, Information and Systems, University of Tsukuba)

Financial risk management with high-frequency data

(∗ Speaker)

16:45∼17:25 Junichi Hirukawa[*,a] and Kou Fujimori[b]

[a](Faculty of Science, Niigata University)

[b](School of Fundamental Science and Engineering, Waseda University)

Weak convergence of the partial sum of $I(d)$ process to a fractional Brownian motion in finite interval representation

November 22 (Friday)

9:20∼10:00 Kengo Kamatani (Graduate School of Engineering Science, Osaka University, and JST CREST)

High-dimensional analysis of the piecewise deterministic Markov process for Bayesian inference

10:10∼10:50 Shogo Kato[*,a] and Peter McCullagh[b]

[a](The Institute of Statistical Mathematics)

[b](Department of Statistics, University of Chicago)

A Cauchy family derived by the Möbius transformations of the sphere

11:00∼17:35 **Special Invited and Keynote Sessions**

18:30∼ Dinner

November 23 (Saturday)

9:20∼10:00 Shota Katayama (Faculty of Economics, Keio University)

Direct estimation of conditional averaging treatment effect in high dimensions

10:10∼10:50 Kei Hirose[a,*] and Hiroki Masuda[b]

[a](Institute of Mathematics for Industry, Kyushu University)

[b](Faculty of Mathematics, Kyushu University)

Statistical modeling for electricity load forecasting

11:00∼11:40 Takuma Bando[a], Tomonari Sei[*,a] and Kazuyoshi Yata[b]

[a](Graduate School of Information Science and Technology, University of Tokyo)

[b](Institute of Mathematics, University of Tsukuba)

Consistency of the objective general index in high dimensional settings

11:40∼ 11:50 Closing

(∗ Speaker)

Special Invited Session

11:00∼11:50 **Data beyond the euclidean space**

Speaker: Jörn Schulz

(Department of Electrical engineering and Computer science, University of Stavanger)

Chair: Shogo Kato (The Institute of Statistical Mathematics)

11:50∼13:15 Lunch

13:15∼14:05 **Change points detection and identification for high dimensional dependent data**

Speaker: Ping-Shou Zhong

(Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago)

Chair: Fumiya Akashi (Graduate School of Economics, University of Tokyo)

14:15∼15:05 **Towards a sparse, scalable, and stably positive definite (inverse) covariance estimator**

Speaker: Joong-Ho (Johann) Won

(Department of Statistics, Seoul National University)

Chair: Shota Katayama (Faculty of Economics, Keio University)

Keynote Session

15:20∼16:20 **A two-stage dimension reduction method and its applications on highly contaminated image sets**

Speaker: I-Ping Tu

(Institute of Statistical Science, Academia Sinica)

Discussion Leader: Yuan-Tsung Chang (Department of Social Information, Mejiro University)

16:35∼17:35 **Sample covariance matrices from "bad populations"**

Speaker: Jeff Yao

(Department of Statistics and Actuarial Science, The University of Hong Kong)

Discussion Leader: Kazuyoshi Yata (Institute of Mathematics, University of Tsukuba)

Tests for high-dimensiomal covariance structures under the SSE model

Aki Ishii[a], Kazuyoshi Yata[b] and Makoto Aoshima[b]

[a] Department of Information Sciences, Tokyo University of Science
[b] Institute of Mathematics, University of Tsukuba

1 Introduction

In this talk, we consider testing the high-dimensional intraclass covariance matrix. We produce a new test statistic for each covariance structure by using the extended cross-data-matrix (ECDM) methodology. We show that the test statistic is an unbiased estimator of its test parameter. We prove that the test statistic has a consistency property and establishes the asymptotic normality. We propose a new test procedure for the high-dimensional intraclass covariance matrix and evaluate its asymptotic size and power theoretically.

Suppose we take samples, $\boldsymbol{x}_j = (x_{1j}, \dots, x_{pj})^T,\ j = 1, \dots, n$, of size $n\ (\geq 4)$, which are independent and identically distributed (i.i.d.) as a $p\ (\geq 2)$-variate distribution. We assume that $\boldsymbol{x}_j$ has an unknown mean vector $\boldsymbol{\mu}$ and unknown (positive-semidefinite) covariance matrix $\boldsymbol{\Sigma}$. Let $\sigma = \text{tr}(\boldsymbol{\Sigma})/p$. Let σ_{ij} be the (i,j) element of $\boldsymbol{\Sigma}$ for $i,j = 1, \dots, p$. We assume that $\sigma_{jj} \in (0,\infty)$ as $p \to \infty$ for all j. For a function, $f(\cdot)$, "$f(p) \in (0,\infty)$ as $p \to \infty$" implies that $\liminf_{p\to\infty} f(p) > 0$ and $\limsup_{p\to\infty} f(p) < \infty$. Then, it holds that $\sigma \in (0,\infty)$ as $p \to \infty$. Let $\rho = \sum_{i\neq j}^{p} \sigma_{ij}/\{\sigma p(p-1)\}$. Note that

$$\frac{\boldsymbol{1}_p^T \boldsymbol{\Sigma} \boldsymbol{1}_p}{p} = \sigma\{1 + \rho(p-1)\} \tag{1.1}$$

and $\rho \in [-(p-1)^{-1}, 1]$, where $\boldsymbol{1}_p = (1, \dots, 1)^T$. We denote the identity matrix of dimension p by $\boldsymbol{I}_p$.

In this paper, we consider testing

$$H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_* \quad \text{vs.} \quad H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_*, \tag{1.2}$$

where $\boldsymbol{\Sigma}_*$ is a candidate (positive-semidefinite) covariance matrix. For $\boldsymbol{\Sigma}_*$ we consider the following covariance structures: (i) identity matrix, (ii) scaled identity matrix, (iii) diagonal matrix, and (iv) intraclass covariance matrix. Let

$$\boldsymbol{\Sigma}_{\text{S}} = \sigma \boldsymbol{I}_p,\ \boldsymbol{\Sigma}_{\text{D}} = \text{diag}(\sigma_{11}, \dots, \sigma_{pp}) \ \text{ and } \ \boldsymbol{\Sigma}_{\text{IC}} = \sigma\{(1-\rho)\boldsymbol{I}_p + \rho \boldsymbol{1}_p \boldsymbol{1}_p^T\}.$$

Let

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_* \quad \text{and} \quad \Delta = \|\boldsymbol{\Sigma}_0\|_F^2 = \text{tr}(\boldsymbol{\Sigma}_0^2),$$

where $\|\cdot\|_F$ is the Frobenius norm. Note that $\Delta = 0$ under H_0 and $\Delta > 0$ under H_1. We regard Δ as a test parameter and construct a test procedure for (1.2) by using an estimator of Δ.

In the current paper, we take a different nonparametric approach and produce a new test statistic for (1.2). We utilize the extended cross-data-matrix (ECDM) method developed by Yata and Aoshima [2] which is an extension of the cross-data-matrix methodology created by Yata and Aoshima [1]. The ECDM method is a nonparametric method to produce an unbiased estimator for a function of $\boldsymbol{\Sigma}$ at a low computational cost even for ultra high-dimensional data.

2 Unbiased estimator of Δ

Let $\boldsymbol{A}_j$ be a $p \times p$ known idempotent matrix with rank $r_j\ (\geq 1)$ for $j = 1, \dots, q$, such that $\sum_{j=1}^{q} r_j = p$ and $\sum_{j=1}^{q} \boldsymbol{A}_j = \boldsymbol{I}_p$, where $r_1 \leq \cdots \leq r_q$ when $q \geq 2$. Note that $\text{tr}(\boldsymbol{A}_j) = r_j$, $\boldsymbol{A}_j^2 = \boldsymbol{A}_j$ and

$\boldsymbol{A}_j\boldsymbol{A}_{j'} = \boldsymbol{O}$ for all j $(\neq j')$. Let κ_j (≥ 0) be an unknown scalar such that $\text{tr}(\boldsymbol{\Sigma}\boldsymbol{A}_j) = r_j\kappa_j$ for all j. Hereafter, we assume that $\boldsymbol{\Sigma}_*$ has the following structure:

$$\boldsymbol{\Sigma}_* = \kappa_1\boldsymbol{A}_1 + \cdots + \kappa_q\boldsymbol{A}_q. \tag{2.1}$$

Note that $\text{tr}(\boldsymbol{\Sigma}_*^2) = \sum_{j=1}^q r_j\kappa_j^2$ and $\Delta = \text{tr}(\boldsymbol{\Sigma}^2) - \text{tr}(\boldsymbol{\Sigma}_*^2)$, so that $\text{tr}(\boldsymbol{\Sigma}^2) \geq \text{tr}(\boldsymbol{\Sigma}_*^2)$. One can summarize as follows:

(I) $\boldsymbol{A}_1 = \boldsymbol{I}_p$, $\kappa_1 = \sigma$, $r_1 = p$ and $q = 1$ when $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_{\text{S}}$;

(II) $\boldsymbol{A}_j = \text{diag}(0,\dots,0,1,0,\dots,0)$ whose j-th diagonal element is 1, $\kappa_j = \sigma_{jj}$, $r_j = 1$ for all j and $q = p$ when $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_{\text{D}}$;

(III) $\boldsymbol{A}_1 = \boldsymbol{1}_p\boldsymbol{1}_p^T/p$, $\boldsymbol{A}_2 = \boldsymbol{I}_p - \boldsymbol{1}_p\boldsymbol{1}_p^T/p$, $\kappa_1 = \sigma\{1 + (p-1)\rho\}$, $\kappa_2 = \sigma(1-\rho)$, $r_1 = 1$, $r_2 = p-1$ and $q = 2$ when $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_{\text{IC}}$.

We note that $\text{tr}(\boldsymbol{\Sigma}\boldsymbol{A}_j)^2/r_j = r_j\kappa_j^2$ for all j.

We first give an unbiased estimator of Δ by using the ECDM method. Let $n_{(1)} = \lceil n/2 \rceil$ and $n_{(2)} = n - n_{(1)}$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. Let

$$\boldsymbol{V}_{n(1)(k)} = \begin{cases} \{\lfloor k/2 \rfloor - n_{(1)} + 1, \dots, \lfloor k/2 \rfloor\} & \text{if } \lfloor k/2 \rfloor \geq n_{(1)}, \\ \{1,\dots,\lfloor k/2 \rfloor\} \cup \{\lfloor k/2 \rfloor + n_{(2)} + 1, \dots, n\} & \text{otherwise;} \end{cases}$$
$$\boldsymbol{V}_{n(2)(k)} = \begin{cases} \{\lfloor k/2 \rfloor + 1, \dots, \lfloor k/2 \rfloor + n_{(2)}\} & \text{if } \lfloor k/2 \rfloor \leq n_{(1)}, \\ \{1,\dots,\lfloor k/2 \rfloor - n_{(1)}\} \cup \{\lfloor k/2 \rfloor + 1, \dots, n\} & \text{otherwise} \end{cases}$$

for $k = 3,\dots,2n-1$, where $\lfloor x \rfloor$ denotes the largest integer $\leq x$. Let $\#\boldsymbol{S}$ denote the number of elements in a set $\boldsymbol{S}$. Note that $\#\boldsymbol{V}_{n(l)(k)} = n_{(l)}$, $l = 1,2$, $\boldsymbol{V}_{n(1)(k)} \cap \boldsymbol{V}_{n(2)(k)} = \emptyset$ and $\boldsymbol{V}_{n(1)(k)} \cup \boldsymbol{V}_{n(2)(k)} = \{1,\dots,n\}$ for $k = 3,\dots,2n-1$. Also, note that $i \in \boldsymbol{V}_{n(1)(i+j)}$ and $j \in \boldsymbol{V}_{n(2)(i+j)}$ for $i < j$ $(\leq n)$. Let

$$\overline{\boldsymbol{x}}_{(1)(k)} = n_{(1)}^{-1} \sum_{j \in \boldsymbol{V}_{n(1)(k)}} \boldsymbol{x}_j \quad \text{and} \quad \overline{\boldsymbol{x}}_{(2)(k)} = n_{(2)}^{-1} \sum_{j \in \boldsymbol{V}_{n(2)(k)}} \boldsymbol{x}_j$$

for $k = 3,\dots,2n-1$. Let $u_{n(l)} = n_{(l)}/(n_{(l)} - 1)$ for $l = 1,2$,

$$\boldsymbol{y}_{ij(1)} = \boldsymbol{x}_i - \overline{\boldsymbol{x}}_{(1)(i+j)} \quad \text{and} \quad \boldsymbol{y}_{ij(2)} = \boldsymbol{x}_j - \overline{\boldsymbol{x}}_{(2)(i+j)}$$

for all $i < j$. We note that $u_{n(l)}E(\boldsymbol{y}_{ij(l)}\boldsymbol{y}_{ij(l)}^T) = \boldsymbol{\Sigma}$ for $l = 1,2$, and $\boldsymbol{y}_{ij(1)}$ and $\boldsymbol{y}_{ij(2)}$ are independent for all $i < j$. For example, Yata and Aoshima [2] gave an estimator of $\text{tr}(\boldsymbol{\Sigma}^2)$ as

$$W_n = \frac{2u_{n(1)}u_{n(2)}}{n(n-1)} \sum_{i<j}^{n} (\boldsymbol{y}_{ij(1)}^T\boldsymbol{y}_{ij(2)})^2 \tag{2.2}$$

by the ECDM method. Then, it holds that $E(W_n) = \text{tr}(\boldsymbol{\Sigma}^2)$. We also give an unbiased estimator of $\text{tr}(\boldsymbol{\Sigma}_*^2)$ by using the ECDM method.

References

[1] Yata, K. and Aoshima, M. (2010). Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. *Journal of Multivariate Analysis, 101*, 2060–2077.

[2] Yata, K. and Aoshima, M. (2013). Correlation tests for high-dimensional data using extended cross-data-matrix methodology. *Journal of Multivariate Analysis, 117*, 313–331.

On error bounds for high-dimensional asymptotic distribution of L_2-type test statistic

Takahiro Nishiyama[a], Masashi Hyodo[b] and Tatjana Pavlenko[c]

[a] Department of Business Administration, Senshu University
[b] Department of Mathematical Sciences, Osaka Prefecture University
[c] Department of Mathematics, KTH Royal Institute of Technology

This talk was concerned with the canonical testing problem in modern statistical inference, namely the two-sample test for equality of means of independent multivariate populations with very large dimensions. Precisely, let $\boldsymbol{x}_{gk} = (x_{g1k}, \ldots, x_{gpk})^\top$, $k \in \{1, \ldots, n_g\}$, be n_g iid random vectors with $\boldsymbol{x}_{gk} \quad \mathcal{N}_p(\ _g, \Sigma_g)$, $g \in \{1, 2\}$, where $\ _g \in \mathbb{R}^p$, $\Sigma_g \in \mathbb{R}^{p\times p}_{>0}$, represent the usual parameters.

We are interested in testing the hypothesis $\mathcal{H} : \| \ _1 \quad \ _2\| = 0$, vs. $\mathcal{A} : \| \ _1 \quad \ _2\| > 0$ where $\| \quad \|$ denotes L_2-norm. The best-known test procedure which accommodates high-dimensional data and allows for $\Sigma_1 \neq \Sigma_2$ is the L_2-type test statistics introduced by Chen and Qin (2010) (hereafter called for Ch-Q test):

$$T_n = \|\overline{\boldsymbol{x}}_1 \quad \overline{\boldsymbol{x}}_2\|^2 \quad \sum_{g=1}^{2} \mathrm{tr}(S_g)/n_g,$$

where $n = n_1 + n_2$, $\overline{\boldsymbol{x}}_g$ and S_g are the sample mean and sample covariance matrix of gth population. To the unbiasedness property of T_n, Chen and Qin (2010) quantified its variance

$$\sigma_n^2 = \mathrm{var}(T_n) = \sum_{g=1}^{2} \frac{2\mathrm{tr}(\Sigma_g^2)}{n_g(n_g \quad 1)} + \frac{4\mathrm{tr}(\Sigma_1\Sigma_2)}{n_1 n_2},$$

under $\mathcal{H}$, and show that the distribution of T_n admits a normal limit after appropriate rescaling. We first presented a theoretical analysis of this problem which requires understanding of the rate of convergence of T_n to its normal limit, and then introduced two new approximations which are more accurate in the asymptotic regime where both n and p tend to infinity.

To provide the intuition behind the approximations which we propose, let $\psi_{\widetilde{T}_n}(t)$ denote the characteristic functions of $\widetilde{T}_n = T_n/\sigma_n$. Also, let $\lambda_r(\Lambda)$ be the r-th largest eigenvalue of the matrix $\Lambda = (n/n_1)\Sigma_1 + (n/n_2)\Sigma_2$ and let $\quad = \lambda_1(\Lambda)^2/\mathrm{tr}(\Lambda^2)$.

Then we obtained the characteristic function of $\widetilde{T}_n$ as

$$\psi_{\widetilde{T}_n}(t) \approx e^{\quad t^2/2} + a(it)^3 e^{\quad t^2/2} \tag{1}$$

where $a = 4b/(3\sigma_n^3)$ and

$$b = \sum_{g=1}^{2} \frac{(n_g \quad 2)\mathrm{tr}(\Sigma_g^3)}{n_g^2(n_g \quad 1)^2} + \frac{3\mathrm{tr}(\Sigma_1^2\Sigma_2)}{n_1^2 n_2} + \frac{3\mathrm{tr}(\Sigma_1\Sigma_2^2)}{n_1 n_2^2}.$$

By this result, the normal approximation of $\widetilde{T}_n$ is immediately achieved, that is $F_{\widetilde{T}_n}(t) = \Phi(t) + o(1)$, as $\quad = 0$. On the other hand, inverting the right side of (1) term by term provided the approximating distribution of $\widetilde{T}_n$ of the form $F_2(x) = \Phi(x) + a(1 \quad x^2)\phi(x)$, where $\phi(x)$ denotes the density of $\Phi(x)$.

Another avenue of research delivers a χ^2-approximation of $\widetilde{T}_n$; this look on the problem is motivated by the work of Buckley and Eagleson (1988) and Zhang (2005). Let $G_d(\ )$ denote the cumulative distribution function of χ^2-distribution with d degrees of freedom. Then we could approximate $F_{\widetilde{T}_n}(x)$ by $F_3(x) = G_d(\sqrt{2d}x + d)$.

To establish the rate of convergence, we obtained the explicit bounds for the Kolmogorov distance between the distribution of $\widetilde{T}_n$ and its approximations $\Phi(x)$, $F_2(x)$, and $F_3(x)$.

Theorem 1. *The distribution of $\widetilde{T}_n$ satisfies the following properties under $\mathcal{H}$:*

(i) For any $n_1, n_2, p, \Sigma_1, \Sigma_2$ such that $\quad < 1/8$,

$$\sup_{x\in\mathbb{R}} |F_{\widetilde{T}_n}(x) \quad F_2(x)| \quad \frac{3}{2\pi\omega}\left\{2+\frac{8!^{1/4}}{8(1 \quad 8 \quad)^2}\right\}+\frac{8 \quad (2+\omega)}{9\pi\omega^2}.$$

(ii) For any $n_1, n_2, p, \Sigma_1, \Sigma_2$ such that $\quad < 1/8$,

$$\sup_{x\in\mathbb{R}} |F_{\widetilde{T}_n}(x) \quad F_3(x)| \quad \frac{3}{2\pi\omega}\left\{2+\frac{8!^{1/4}}{8(1 \quad 8 \quad)^2}\right\}+\frac{\{10+3(1 \quad 8 \quad) \quad ^2\}}{2\pi}.$$

(iii) For any $n_1, n_2, p, \Sigma_1, \Sigma_2$ such that $\quad < 1/6$,

$$\sup_{x\in\mathbb{R}} |F_{\widetilde{T}_n}(x) \quad \Phi(x)| \quad \frac{2 \quad ^{1/2}}{2\pi\omega}\left\{3\sqrt{\pi}+\frac{6!^{1/4}\sqrt{2}}{3(1 \quad 6 \quad)^{3/2}}\right\},$$

where ω is omega constant which is a mathematical constant defined by $\omega \exp(\omega) = 1$, and $\omega \approx 0.56714$.

Based on Theorem 1, we established convergence rate of the proposed approximations. Also, we proposed the Ch-Q test with the proposed approximations rests on the adjusted α-quantiles, $q_2(\alpha) = z_\alpha + 4b(z_\alpha^2 \quad 1)/(3\sigma_n^3)$ and $q_3(\alpha) = (\chi_d^2(\alpha) \quad d)/(2d)^{1/2}$. Besides, we evaluated empirical quantiles of $\widetilde{T}_n$ to assess the accuracy of the proposed approximations $q_2(\alpha)$ and $q_3(\alpha)$, and the sizes of the proposed tests for some selected parameters.

References

[1] Buckley, M.J., Eagleson, G.K., 1988. An approximation to the distribution of quadratic forms in normal random variables. *Austral. J. Statist.*, **30**, 150–159.

[2] Chen, S.X., Qin, Y.L., 2010. A two-sample test for high dimensional data with applications to gene-set testing. *Ann. Statist.*, **38**, 808–835.

[3] Zhang, J.T., 2005. Approximate and asymptotic distributions of chi-squared-type mixtures with applications. *J. Amer. Statist. Assoc.*, **100**, 273–285.

Financial Risk Management with High-Frequency Data

Hiroumi Misaki [1]

1 Introduction

Methods for estimating covariance and correlation between multiple asset prices have been investigated intensively in the field of financial econometrics. In recent decades, daily covariance and correlations have been estimated directly using high-frequency financial data or *tick-by-tick* data. The estimation object is *integrated covariance*, which is a natural measure of the covariation of multivariate high-frequency asset prices. Kunitomo and Sato [1, 2, 3] proposed a statistical method called the separating information maximum likelihood (SIML) estimator.

The methods used to estimate the integrated volatility and covariance (subsequently, correlation) from high-frequency data are called realized measures. In this presentation, we first compare the accuracy of realised measures using a number of computer simulations in univariate case. Second, we introduce a brief result when we combine high-frequency covariance estimation with optimal portfolio selection methods.

2 Robustness of the SIML estimation

We compare the accuracy of realised measures using a number of computer simulations. We consider a simple realised volatility (RV), a 5-minute RV, a subsampled 5-minute RV, a two-scale estimator (TS), a realised kernel (RK), a pre-averaging estimator (PA) and a separating information maximum likelihood estimator (SIML).

We use several non-linear transformation models to obtain the form of the market microstructure noise. We perform 132 cases of simulation in total.

Our findings are as follows. RV and RV5 are dominated by RV5ss in all cases. If this simplest class of realised measures is used, subsampling is recommended. RV5ss is heavily biased when the market microstructure noise or/and round-off is large. TS, RK and PA are reasonable in most cases, but we have observed some drawbacks in the case of a large round-off with a small noise and a small adjustment with a large noise case. TS is also problematic in the SSAR model with large noise.

SIML is not biased irrationally in any case; therefore we can conclude that SIML is sufficiently robust to the form of the market microstructure noise. We have also found that SIML is the only realised measure to maintain the consistency in all simulations. Then, we can conclude that SIML is a reasonable choice in practice when we do not know the exact form of the market microstructure noise, particularly for assets that are actively traded.

[1] Faculty of Engineering, Information and Systems, University of Tsukuba, Tennodai 1-1-1, Tsukuba City, Ibaraki 305-8577, JAPAN, hmisaki@risk.tsukuba.ac.jp

3 Portfolio selection

Meucci [4] derived the *principal portfolios* using principal component analysis. The *diversification index* is defined as the dispersion of the *diversification distribution*, based on its entropy. We maximize the *diversification index* to obtain the optimal weight $\mathbf{w}^*$ to construct the risk diversified portfolio.

We compare the following six methods to construct the risk diversified portfolios:

(i) Naive. An equally weighted portfolio, i.e., $w_i = 1/N$.

(ii) OC-125. An analogy of the conventional method of the intraday case; $\tilde{\mathbf{\Sigma}}_t$ are obtained by 125-day moving average of outer product of open to close returns.

(iii) SIML-1. We use the previous day's SIML estimate for asset allocation in the current day: $\tilde{\mathbf{\Sigma}}_t = \hat{\mathbf{\Sigma}}_{t-1}$.

(iv) SIML-5. We use the averages of previous estimates: $\tilde{\mathbf{\Sigma}}_t = (\hat{\mathbf{\Sigma}}_{t-1} + \cdots + \hat{\mathbf{\Sigma}}_{t-5})/5$ to moderate the estimation errors using a single day.

(v) SIML-22. One-month version of SIML-5: $\tilde{\mathbf{\Sigma}}_t = (\hat{\mathbf{\Sigma}}_{t-1} + \cdots + \hat{\mathbf{\Sigma}}_{t-22})/22$.

(vi) SIML-0. We use the estimates of the covariance matrix of the same day to carry out portfolio optimization: $\tilde{\mathbf{\Sigma}}_t = \hat{\mathbf{\Sigma}}_t$.

We selected 10 assets from the stock market. We find that the risk diversified portfolio is feasible when combining with high-frequency covariance estimation. However, the performances of the portfolios with SIML estimation are not always better than the naive portfolio. Therefore, a prudent choice of prediction method of covariance matrices is recommended.

References

[1] Kunitomo, N., Sato S.: Separating information maximum likelihood estimation of realized volatility and covariance with micro-market noise. Discussion paper CIRJE-F-581, Graduate School of Economics, University of Tokyo (2008)

[2] Kunitomo, N., Sato S.: The SIML estimation of the integrated volatility of Nikkei-225 futures and hedging coefficients with micro-market noise. *Math. Comput. Simulat* 8, 1272-1289 (2011)

[3] Kunitomo, N., Sato S.: Separating information maximum likelihood estimation of the integrated volatility and covariance with micro-market noise. *N. Am. J. Econ. Financ.* 26, 282-309 (2013)

[4] Meucci, A.: Managing diversification. *Risk*, 22(5), 74-7 (2009)

Weak convergence of the partial sum of $I(d)$ process to a fractional Brownian motion in finite interval representation

Junichi Hirukawa and Kou Fujimori
Niigata University and Waseda University

ABSTRACT

An integral transformation which changes a fractional Brownian motion to a process with independent increments has been given. A representation of a fractional Brownian motion through a standard Brownian motion on a finite interval has also been given. On the other hand, it is known that the partial sum of the discrete time fractionally integrated process ($I(d)$ process) weakly converges to a fractional Brownian motion in infinite interval representation. In this talk we derive the weak convergence of the partial sum of $I(d)$ process to a fractional Brownian motion in finite interval representation.

1 Introduction

Stochastic analysis for FBM has been developed by Decreusefond and Üstünel (1997) using Malliavin calculus. Norros et al. (1999) showed that many basic results can be obtained more directly with rather elementary arguments and computations. Norros et al. (1999) considered a normalized fractional Brownian motion (FBM) $(Z_t)_{g\geq 0}$ with self-similarity parameter $H \in (0,1)$. Mandelbrot and Van Ness (1968) defiend the process more constructively as the integral

$$Z_t - Z_s = c_H \left(\int_s^t (t-u)^{H-1/2}\, dW_u + \int_{-\infty}^t \left\{ (t-u)^{H-1/2} - (s-u)^{H-1/2} \right\} dW_u \right),$$

where W_t is the standard Brownian motion. The normalization $E\left(Z_1^2\right) = 1$ is achieved with the choice

$$c_H = \left(\frac{2H\Gamma\left(\frac{3}{2} - H\right)}{\Gamma\left(H + \frac{1}{2}\right)\Gamma(2-2H)} \right)^{1/2},$$

where $\Gamma(\cdot)$ denotes the Gamma function.

1.1 The fundamental martingale M

Norros et al. (1999) considered the following process. Let $w(t,s)$ be the function

$$w(t,s) = \begin{cases} c_1 s^{1/2-H}(t-s)^{1/2-H}, & \text{for } s \in (0,t), \\ 0, & \text{for } s \notin (0,t), \end{cases}$$

where

$$c_1 = \left\{ 2HB\left(\frac{1}{2} - H, H + \frac{1}{2} \right) \right\}^{-1}$$

and B is the beta funtion

$$B(u,v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}.$$

Then, the centered Gaussian process

$$M_t = \int_0^t w(t,s)\, dZ_s$$

has independent increments and variance function

$$E\left(M_t^2\right) = c_2^2 t^{2-2H},$$

where

$$c_2 = \frac{c_H}{2H(2-2H)^{1/2}}.$$

In particular, M is a martingale.

2 Weak convergence of $I(d)$ process

Now, we obtain the following functional central limit result for $I(d)$ process

$$\begin{aligned}\frac{1}{\sigma n^{d+1/2}}\widetilde{Z}_{[nt]} &= \frac{1}{\sigma n^{d+1/2}}\sum_{s=1}^{[nt]} v_{s-1}^{1/2} DW_s + \frac{1}{\sigma n^{d+1/2}}\sum_{s=1}^{[nt]-1}\left\{\sum_{u=s}^{[nt]-1}\theta_{u,u+1-s} v_{s-1}^{1/2}\right\} DW_s \\ &\Rightarrow \frac{1}{\Gamma(d)}\int_0^t s^{-d}\left\{\int_s^t (u-s)^{d-1} u^d du\right\} dW(s) := \int_0^t dZ(s) = Z(t).\end{aligned}$$

References

Mandelbrot B. B. and van Ness J. W. (1968). Fractional Brownian motions, fractional noises and applications *SIAM Review,* **10**, 422–437.

Norros, I. and Valkeila, E. and Virtamo, J. (1999). An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions *Bernoulli* **4**, 571–587.

Pipiras, V. and Taqqu, Murad S. (2001). Are classes of deterministic integrands for fractional Brownian motion on an interval complete? *Bernoulli* **7**, 873–897.

Tanaka, K. (2013). Distributions of the maximum likelihood and minimum contrast estimators associated with the fractional Ornstein-Uhlenbeck process *Stat. Inference Stoch. Process.* **16**, 173–192.

High-dimensional analysis of the piecewise deterministic Markov process for Bayesian inference

Kengo Kamatani (Osaka University, JST CREST)

In this talk, we review the recent results of piecewise deterministic Markov processes for Bayesian inference in a high-dimension context. Suppose we wish to sample from a probability distribution

$$\Pi(\mathrm{d}x) = \exp(-\mathcal{H}(x))\mathrm{d}x \ (x \in \mathbb{R}^d)$$

where $\mathcal{H} : \mathbb{R}^d \to \mathbb{R}$ is a continuously differentiable function. For the Bayesian context, this probability distribution is the posterior distribution of interest. If we have an i.i.d. sample from Π, we can approximate Π-integral of any function $f(x)$ by the law of large numbers

$$\frac{1}{M}\sum_{m=1}^{M} f(X_m) \longrightarrow \int_{\mathbb{R}^d} f(x)\Pi(\mathrm{d}x). \tag{0.1}$$

In most of the cases, direct i.i.d. sampling is impossible or computationally very expensive. For these cases, the Markov chain Monte Carlo method is useful which originated with the classic paper by Metropolis et al. (1953) almost 70 years ago. The Markov chain Monte Carlo method is designed to construct an ergodic Markov kernel P which is Π-invariant. If a Markov chain $X_1, X_2, \ldots$ is generaetd from the Markov kernel P then the law of large numbers (0.1) is satisfied. The Markov chain Monte Carlo is now a gold standard for Bayesian inference.

Recently, its continuous process version, the Markov **process** Monte Carlo method is of substantial interest for Monte Carlo analysis. Known Markov process Monte Carlo methods rely on an auxiliary variable trick which uses an auxiliary variable v with a probability density ν on Ξ and considers the joint probability distribution $\mu := \Pi(\mathrm{d}x) \otimes \nu(\mathrm{d}v)$ as an extended target distribution on $\mathcal{Z} = \mathbb{R}^d \times \Xi$. The original target distribution is a marginal distribution of the extended target distribution. Since Brownian motion does not have an absolutely continuous path, we can not simulate processes driven by Brownian motion exactly. For our Monte Carlo analysis, exact sampling is necessary. Therefore, the Markov processes of interest should not have a Brownian part. Known processes consist of a deterministic part and a pure jump part. These processes are known as the **piecewise deterministic Markov processes**.

Here we follow Azaïs et al. (2014) for the expression of the piecewise deterministic Markov processes. The processes are constructed by characteristics $(\phi, \lambda_k, Q_k : l = 1, \ldots, K)$. The flow $\phi : \mathcal{Z} \times \mathbb{R} \to \mathcal{Z}$ is a one-parameter group of homeomorphism, that is, ϕ is continuous, $\phi(\cdot, t)$ is a homeomorphism for each $t \in \mathbb{R}$ and $\phi(\phi(\cdot, s), t) = \phi(\cdot, s + t)$. For each $k = 1, \ldots, K$, the jump rate $\lambda_k : \mathcal{Z} \to \mathbb{R}_+$ determines the jump time of pure jump processes, and Q_k is a Markov kernel on $\mathcal{Z}$. Let $\Lambda_k(z, t) = \int_0^t \lambda_k(\phi(z, s))\mathrm{d}s$.

The Markov process is defined by the following way. Suppose $z(0) = (x(0), t(0)) \in \mathcal{Z}$. Let $T_1, \ldots, T_K$ be independent processes with $\mathbb{P}(T_k \geqslant t) = \exp(-\Lambda_k(z, t))$. Let $T_* = \min_{k=1,\ldots,K} T_k$. If $T_k = T_*$, then Z is generated from $Q_k(\phi(z, T_*), \cdot)$ and set

$$X(t) = \begin{cases} \phi(z(0), t) & \text{for } t < T_* \\ Z & \text{for } t = T_*. \end{cases}$$

After T_*, the process evolves in the same way with starting value Z. There are several choices of characteristics. We introduce three piecewise deterministic Markov processes.

Two popular piecewise deterministic Markov processes use the same flow ϕ defined by $x'(t) = v(t)$ and $v'(t) = 0$. The **Zig-Zag sampler** proposed by Bierkens et al. (2019) uses d Markov kernels $Q_1, \ldots, Q_d$ with d jump rates $\lambda_1, \ldots, \lambda_d$. For each $i = 1, \ldots, d$, the Markov kernel is a deterministic kernel Q_i defined by a map $(x, v) \mapsto (x, F_i(v))$ where F_i is an operator which flips the sign of the i-th coordinate of x. The jump rate is defined by $\lambda_i((x, v)) = \max\{0, \partial_i \mathcal{H}(x) v_i\}$.

The **bouncy particle sampler** proposed by Peters and de With (2012), Bouchard-Côté et al. (2018) uses two Markov kernels Q_{bounce} and Q_{ref} with corresponding jump rates λ_{bounce} and λ_{ref}. The kernel Q_{bounce} is a deterministic kernel defined by a map $(x, v) \mapsto (x, \kappa(x, v))$ where

$$\kappa(x, v) = v - 2\frac{\langle \nabla\mathcal{H}(x), v\rangle}{\|\nabla\mathcal{H}(x)\|^2}\nabla\mathcal{H}(x)$$

and $\lambda_{\text{bounce}}(x, v) = \max\{0, \langle \nabla\mathcal{H}(x), v\rangle\}$. The jump rate λ_{ref} is a positive constant, and Q_{ref} is a μ-invariant Markov kernel. For our analysis, for simplicity, we assume $Q_{\text{ref}}((x, v), \mathrm{d}(y, w)) = \nu(\mathrm{d}w)$.

We review some recent results on asymptotic properties of the above deterministic Markov processes such as Bierkens et al. (2018), Andrieu et al. (2018), Deligiannidis et al. (2018).

References

Christophe Andrieu, Alain Durmus, Nikolas Nüsken, and Julien Roussel. Hypocoercivity of Piecewise Deterministic Markov Process-Monte Carlo. *arXiv e-prints*, art. arXiv:1808.08592, Aug 2018.

Romain Azaïs, Jean-Baptiste Bardet, Alexandre Génadot, Nathalie Krell, and Pierre-André Zitt. Piecewise deterministic Markov process—recent results. In *Journées MAS 2012*, volume 44 of *ESAIM Proc.*, pages 276–290. EDP Sci., Les Ulis, 2014. doi: 10.1051/proc/201444017.

Joris Bierkens, Kengo Kamatani, and Gareth O. Roberts. High-dimensional scaling limits of piecewise deterministic sampling algorithms. *arXiv e-prints*, art. arXiv:1807.11358, Jul 2018.

Joris Bierkens, Paul Fearnhead, and Gareth Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.*, 47(3):1288–1320, 2019. ISSN 0090-5364. doi: 10.1214/18-AOS1715.

Alexandre Bouchard-Côté, Sebastian J. Vollmer, and Arnaud Doucet. The bouncy particle sampler: A nonreversible rejection-free markov chain monte carlo method. *Journal of the American Statistical Association*, 113(522):855–867, 2018. doi: 10.1080/01621459.2017.1294075. URL `https://doi.org/10.1080/01621459.2017.1294075`.

George Deligiannidis, Daniel Paulin, Alexandre Bouchard-Côté, and Arnaud Doucet. Randomized Hamiltonian Monte Carlo as Scaling Limit of the Bouncy Particle Sampler and Dimension-Free Convergence Rates. *arXiv e-prints*, art. arXiv:1808.04299, Aug 2018.

N. Metropolis, W. A. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

Elias AJF Peters and G. de With. Rejection-free monte carlo sampling for general potentials. *Physical Review E*, 85(2):026703, 2012.

A Cauchy family derived by the Möbius transformations of the sphere

Shogo Kato[*,a] and Peter McCullagh[b]

[a] *The Institute of Statistical Mathematics, Japan*
[b] *Department of Statistics, University of Chicago, USA*

1 A Cauchy family on the sphere

This paper discusses a family of distributions on the unit sphere $S^d \subset \mathbb{R}^{d+1}$ with probability density function

$$f(y;\mu,\rho) = \frac{\Gamma\{(d+1)/2\}}{2\pi^{(d+1)/2}} \left(\frac{1-\rho^2}{1+\rho^2-2\rho\mu^T y} \right)^d, \quad y \in S^d, \tag{1}$$

with respect to surface area, where $\mu \in S^d$ is the location parameter, $\rho \in [0,1)$ is the concentration parameter, and $S^d = \{x \in \mathbb{R}^{d+1} \,;\, \|x\| = 1\}$ denotes the unit sphere in $\mathbb{R}^{d+1}$. The circular case ($d = 1$) is well-known as the wrapped Cauchy or circular Cauchy family; see, e.g., Kent & Tyler (1988) and McCullagh (1996). In this paper, the distribution (1) is called the Cauchy distribution on the sphere or the spherical Cauchy distribution.

McCullagh (1996) showed that the wrapped Cauchy family is closed under conformal maps preserving the unit circle which are called the Möbius transformations on the unit circle, and that there are similar induced transformations on the parameter space. Related results about the Cauchy family on the real line and on the Euclidean space have been given by McCullagh (1992) and Letac (1986), respectively. To our knowledge, however, there has been no literature about the association between the Möbius transformations and the spherical Cauchy family (1). Since there have been various statistical applications of the wrapped Cauchy family and/or the Möbius transformations in directional statistics (McCullagh, 1996; Downs & Mardia, 2002; Downs, 2003; Jones, 2004; Kato, 2010; Kato & Jones, 2010; Kato & Pewsey, 2015; Uesu *el al.*, 2015), it is potentially useful to consider the Cauchy family on the sphere and its relationship with the Möbius transformations.

2 Some properties of a Cauchy family on the sphere

This paper presents some properties of the Cauchy family on the sphere, especially, those related to the Möbius transformations. The Möbius transformation is defined by

$$\mathcal{M}_{R,\psi}(x) = R \left\{ \frac{1-\|\psi\|^2}{\|\tilde{x}+\psi\|^2}(\tilde{x}+\psi) + \quad \right\}, \quad x \in \mathbb{R}^{d+1} \setminus \{0, -\psi/\|\psi\|^2\}. \tag{2}$$

where $\tilde{x} = x/\|x\|^2$, $\psi \in \mathbb{R}^{d+1} \backslash S^d$, and R is a $(d+1)\times(d+1)$ rotation matrix. Also, we define $\mathcal{M}_{R,\psi}(0) = R\psi$, $\mathcal{M}_{R,\psi}(-\psi/\|\psi\|^2) = \infty$ and $\mathcal{M}_{R,\psi}(\infty) = R\psi/\|\psi\|^2$. The Möbius transformation (2) is a bijective conformal map which maps $\overline{\mathbb{R}}^{d+1}$ onto itself, where $\overline{\mathbb{R}}^{d+1}$ denotes the

[*] *Address for correspondence*: Shogo Kato, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan. `E-mail: skato@ism.ac.jp`

$(d+1)$-dimensional compactified Euclidean space $\mathbb{R}^{d+1} \cup \{\infty\}$. For any $\psi \in \overline{\mathbb{R}}^{d+1} \setminus S^d$, the transformation (2) maps the unit sphere S^d onto itself.

The spherical Cauchy family is closed under the Möbius transformations (2) on the sphere, and the transformed parameter is given by the Möbius transformation (2) on $\overline{\mathbb{R}}^{d+1}$. The statistical benefits of this property include: (i) an efficient algorithm for random variate generation; (ii) a simple pivotal statistic for parametric inference; (iii) straightforward calculation of probabilities of a region; (iv) closed form expression for the maximum likelihood estimator for $n \leq 3$; and (v) straightforward calculation of the Fisher information matrix. A method of moments estimator can be expressed in simple form. A simple algorithm for maximum likelihood estimation is available. The likelihood for the spherical Cauchy is equivalent to that for the t-family with a certain degree of freedom which is related to the spherical Cauchy via stereographic projection. An asymptotically efficient estimator is presented which our simulation study suggests outperforms the method of moments estimator and the maximum likelihood estimator in certain settings. Comparing the densities of the spherical Cauchy and von Mises–Fisher, the spherical Cauchy density takes greater values around the mode and antimode and smaller values in the other area of the sphere. (See, e.g., Section 9.3.2 of Mardia & Jupp, 1999, for the definition and properties of the von Mises–Fisher family.) The advantages of the spherical Cauchy over the von Mises–Fisher in terms of properties include the closure under the Möbius transformations and the related properties, while the von Mises–Fisher compares favourably with the spherical Cauchy in terms of its membership in the exponential family, straightforward maximum likelihood estimation and well-developed theory of hypothesis testing.

The preprint version (Kato & McCullagh, 2018) of this paper includes details of the properties of the spherical Cauchy family discussed above.

Acknowledgment

The work of the first author was supported by JSPS KAKENHI Grant Number 17K05379.

References

Downs, T. D. (2003). Spherical regression. *Biometrika*, **90**, 655–668.

Downs, T. D. & Mardia, K. V. (2002). Circular regression. *Biometrika*, **89**, 683–697.

Jones, M. C. (2004). The Möbius distribution on the disc. *Ann. Inst. Statist. Math.*, **56**, 733–742.

Kato, S. (2010). A Markov process for circular data. *J. R. Statist. Soc.* B **72**, 655–672.

Kato, S. & Jones, M. C. (2010). A family of distributions on the circle with lines to, and applications arising from, Möbius transformation. *J. Am. Statist. Assoc.*, **105**, 249–262.

Kato, S. & McCullagh, P. (2018). Möbius transformation and a Cauchy family on the sphere. *arXiv preprint arXiv:1510.07679v2 [math.ST].*

Kato, S. & Pewsey, A. (2015). A Möbius transformation-induced distribution on the torus. *Biometrika*, **102**, 359–370.

Kent, J. T. & Tyler, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Statist.*, **15**, 247–254.

Letac, G. (1986). Seul le groupe des similitudes-inversions préserve le type de la loi de Cauchy-conforme de $\mathbb{R}^n$ pour $n > 1$. *J. Funct. Anal.*, **68**, 43–54.

Mardia, K. V. & Jupp, P. E. (1999). *Directional Statistics.* Chichester: Wiley.

McCullagh, P. (1992). Conditional inference and Cauchy models. *Biometrika*, **79**, 247–259.

McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation. *Ann. Statist.*, **24**, 787–808.

Uesu, K., Shimizu, K. & SenGupta, A. (2015). A possibly asymmetric multivariate generalization of the Möbius distribution for directional data. *J. Multi. Anal.*, **134**, 146–162.

Data Beyond the Euclidean Space

Jörn Schulz

Department of Electrical Engineering and Computer Science,
University of Stavanger, Stavanger, Norway

Complex data such as non-Euclidean or a mixture of Euclidean and non-Euclidean data has gained growing attention recently. However, only few methods are available to do sensitive statistical inferences on these types of data and only little is known about their asymptotic properties. In the following, we assume that the non-Euclidean data lives on a smooth manifold and in particular we will focus on directional data, i.e. data on the hypersphere $\mathbb{S}^d = \{\mathbf{x} \in \mathbb{R}^{d+1} : \mathbf{x}^T\mathbf{x} = 1\}$ and data on polyspheres $(\mathbb{S}^2)^d$. Examples of these data types are i.) shape representations including directions such as skeletal representations that live on $\mathbb{S}^{d_1} \times (\mathbb{S}^2)^{d_2}$ (Hong et al. (2016); Pizer et al. (2013); Schulz et al. (2016)), ii.) dihedral angles of protein structures on $(\mathbb{S}^1 \times \mathbb{S}^1)^d$ (Eltzner et al. (2017)) or iii.) to analyze temporal sequences of molecules on $\mathbb{S}^d$ (Dryden et al. (2019)). Especially, in examples i.) and ii.) we have usually a high dimension low sample size setting, i.e. $d \gg n$ where n is the sample size and d is the dimension.

A crucial step in the analysis in all these applications is principal nested spheres (PNS) (Jung et al. (2012)), a method for decomposition and dimension reduction of directional data on $\mathbb{S}^d$. In opposite to principal component analysis, PNS is a backward dimension reduction method. In each step, a submanifold of successively lower dimension, containing the largest total variance, is fitted to the data. A submanifold can be either a small-sphere or a great sphere, i.e. a sphere with radius $r < \pi/2$ or $r = \pi/2$. The choice of a small or a great sphere is a critical question in the PNS procedure. The fitting of a small sphere to the data might result in an overfitting, e.g. if the data is concentrated around a point at $\mathbb{S}^d$. We will discuss a new testing procedure that outperforms alternative testing methods during a simulation study and the analysis of skeletal 3D models of hippocampi. The proposed method is based on a measure of multivariate kurtosis for directional data. Given a suitable decomposition of the data, statistical inference by hypothesis testing (Schulz et al. (2016)), classification (Hong et al. (2016)) or clustering (Dryden et al. (2019)) might be performed.

In addition, we will briefly review and discuss some recent works on asymptotic results within this framework.

References

Dryden, I. L., Kim, K.-R., Laughton, C. A., and Le, H. (2019), "Principal nested shape space analysis of molecular dynamics data," *arXiv preprint arXiv:1903.09445*.

Eltzner, B., Huckemann, S., and Mardia, K. V. (2017), "Torus Principal Component Analysis with an Application to RNA Structures," *Annals of Applied Statistics*, ISSN 1932-6157 (In Press).

Hong, J., Vicory, J., Schulz, J., Styner, M., Marron, J. S., and Pizer, S. M. (2016), "Non-Euclidean classification of medically imaged objects via s-reps," *Medical image analysis*, 18, 37–45.

Jung, S., Dryden, I. L., and Marron, J. S. (2012), "Analysis of Principal Nested Spheres," *Biometrika*, 99, 551–568.

Pizer, S. M., Jung, S., Goswami, D., Zhao, X., Chaudhuri, R., Damon, J. N., Huckemann, S., and Marron, J. S. (2013), "Nested Sphere Statistics of Skeletal Models," in *Innovations for Shape Analysis: Models and Algorithms*, eds. Breus, M., Bruckstein, A., and Maragos, P., New York: Springer, pp. 93–115.

Schulz, J., Pizer, S. M., Marron, J. S., and Godtliebsen, F. (2016), "Non-linear Hypothesis Testing of Geometric Object Properties of Shapes Applied to Hippocampi," *Journal of Mathematical Imaging and Vision*, 54, 15–34, issue 1.

Ping-Shou Zhong

University of Illinois at Chicago

Title: Change point detection and identification for high dimensional dependent data

Abstract: High-dimensional functional data appear in practice when a dense number of repeated measurements are taken on a large number of variables for a relatively small number of experimental units. The spatial-temporal dependence and high-dimensional nature of the data structure make statistical analysis and computation a challenge. This talk will introduce computationally efficient procedures to detect and identify change points among covariance matrices from high-dimensional functional data. The change point detection procedure is presented in the form of a hypothesis test, and the asymptotic distributions of the proposed test statistics are established under an asymptotic framework with “large p, large T and small n”, where p is data dimension, T is the number of repeated measurements and n is the sample size. We also propose change-point estimators for both single and multiple change points. These estimators are proven to be consistent under a mild set of conditions. The rate of convergence of the estimator depends on the data dimension, sample size, number of repeated measurements, and signal-to-noise ratio. Computation efficiency is carefully studied to address the challenges due to the large number of repeated measurements and high-dimensionality. Simulation results demonstrate that the size of the detection procedure is well controlled at the nominal level, and the locations of multiple change points can accurately be identified. We apply the proposed approach to find event boundaries in a continuous movie by identifying change points among functional connectivity using functional MRI data.

Towards a sparse, scalable, and stably positive definite (inverse) covariance estimator

Joong-Ho (Johann) Won
Department of Statistics, Seoul National University

Abstract: High-dimensional covariance estimation and graphical model selection is a contemporary topic in statistics and machine learning, and has widespread applications. The problem is notoriously difficult in high dimensions as the traditional estimate is not even positive definite, let alone sufficiently stable. An important line of research is to shrink the spectrum to yield stable well-conditioned estimators. A separate line of research has considered sparse estimation using nonsmooth regularization methods and provides interpretable models with fewer parameters. Though an estimator which is both stable and sparse is often desirable in numerous downstream applications, obtaining such estimators is inherently challenging in modern high-dimensional regimes due to the very different nature of the two approaches. In this talk we propose a unifying and scalable framework which addresses this problem. Our general methodology takes an arbitrary covariance loss functions (such as the ones which have been proposed in the literature) and yields estimates that are both spectrally regularized and sparse. The framework leads to an enriched class of estimators which are computationally tractable and enjoy good asymptotic properties. In addition, when the covariance loss function is orthogonally invariant, we further demonstrate that a solution path algorithm can be derived, involving a series of ordinary differential equations. The path algorithm is attractive because it provides the entire family of estimates for all possible values of the regularization parameter, at the same computational cost of a single estimate with a fixed parameter. An important finding is

that an iterative path algorithm can be devised even when the loss function is not orthogonally invariant, utilizing modern operator splitting techniques. We illustrate the efficacy of our approach on both real and simulated data.

This is a collaboration with Sang-Yun Oh (UC Santa Barbara) and Bala Rajaratnam (UC Davis).

A Two-Stage Dimension Reduction Method and its Applications on Highly Contaminated Image Sets

I-Ping Tu
Institute of Statistical Science, Academia Sinica, Taiwan

Abstract

Principal component analysis (PCA) is arguably the most popular dimension reduction method for vector type data. When applied on image data, PCA demands the images to be portrayed as vectors. The resulting computation is heavy because it would solve an eigenvalue problem of a covariance matrix whose size equals the square of the pixel number. To mitigate the computation burden, multi-linear PCA that uses column and row basis with a Kronecker product to compose the matrix structure was proposed, for which the success was demonstrated on face image sets. However, when we apply MPCA on the particle images of the single particle cryo-electron microscopy (cryo-EM) experiments, the results are not satisfying. Here, we propose a dimension reduction method called Two Stage Dimensional Reduction (2SDR) where we first apply MPCA to extract its projection scores, and then apply PCA on these scores to further reduce the dimension. Tests using single particle cryo-EM benchmark experimental data sets demonstrate that 2SDR reduce huge computation costs compared to PCA, and show 2SDR can reconstruct better quality images than MPCA. Further application of 2SDR on a cryo-EM micrograph data set significantly reduces the noise to clearly reveal the individual particles. Remarkably, the de-noised particles boxed out from the micrograph allow subsequent structural analysis to reach a high-quality 3D density map. This is a joint work with Szu-Chi Chung, Po-Yao Niu, Su-Yun Huang and Wei-Hau Chang.

Sample covariance matrices from "bad populations"

Jeff Yao

Department of Statistics and Actuarial Science,
The University of Hong Kong

Recent spectral analysis of large covariance matrices is largely based on the celebrated Marcenko-Pastur law and subsequent applications of the theory to high-dimensional inference involve central limit theorems for the corresponding eigenvalue statistics. However it has recently appeared that there are some important multivariate populations with strongly dependent coordinates for which the exiting theories do not apply. High-dimensional mixtures are one of such "bad populations". In this talk, I will describe this phenomenon and then present some alternative results for the case of high-dimensional mixtures.

Direct estimation of conditional averaging treatment effect in high dimensions

Shota Katayama

Keio University, Japan

The estimation of conditional average treatment effect (CATE) is a general and fundamental problem in observational studies. Such estimation problem is essential for policy evaluation, personalized medicine, offline or online marketing and advertising. Usually, to identify CATE, one requires the strong ignorability condition which says that outcomes and treatment assignment is independent conditional on covariates. In other words, only the covariates we collect affect both of outcomes and treatment assignment. If we fail to collect such a covariate, the strong ignorability does not hold. Clearly, a large number of covariates tends to meet the strong ignorability, although it is uncheckable condition from observations. With advances of information technology and database system, it would be plausible to consider the high dimensional covariates.

In this talk, we consider the estimation of CATE in high dimensions. Following the Neyman–Rubin's potential outcome framework (Rubin, 1974, Neyman et al., 1990), assume that there is a potential outcomes $(Y_i(0), Y_i(1))$ for each sample $i \in \{1, 2, \ldots, n\}$. Let $T_i \in \{0, 1\}$ be the assignment indicator. Then, $Y_i(0) \in \mathbb{R}$ is the potential outcome when the sample i is assigned to the control ($T_i = 0$) and $Y_i(1) \in \mathbb{R}$ is the potential outcome when it is assigned to the treatment ($T_i = 1$). Assume that we have n independent and identically distributed examples $\{(\boldsymbol{X}_i, T_i, Y_i(T_i))\}_{i=1}^n$ where $\boldsymbol{X}_i \in \mathbb{R}^p$ is the covariates with possibly high dimensions, that is, $p \gg n$. Our goal is to estimate the conditional average treatment effect (CATE) given by

$$\tau^*(\boldsymbol{x}) = \mathbb{E}\big\{Y_i(1) - Y_i(0) \,|\, \boldsymbol{X}_i = \boldsymbol{x}\big\}.$$

To identify the CATE, we assume the following strong ignorability condition.

Assumption 1. $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \,|\, \boldsymbol{X}_i$

Moreover, we assume the linearity for the potential outcomes.

Assumption 2. $\mathbb{E}\big\{Y_i(0) \,|\, \boldsymbol{X}_i = \boldsymbol{x}_i\big\} = \boldsymbol{x}_i^T \boldsymbol{\beta}_0^*$ and $\mathbb{E}\big\{Y_i(1) \,|\, \boldsymbol{X}_i = \boldsymbol{x}_i\big\} = \boldsymbol{x}_i^T \boldsymbol{\beta}_1^*$.

From Assumption 2, we have $\tau^*(\boldsymbol{x}) = \boldsymbol{x}^T(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_0^*)$ and we can estimate $\boldsymbol{\beta}_1^*$ from the treated examples and can estimate $\boldsymbol{\beta}_0^*$ from the control examples under Assumption 1. Since the covariates $\boldsymbol{X}_i$ is high dimension, a natural approach would be applying the Lasso proposed by Tibshirani (1996) for each treated and control examples, i.e.,

$$\hat{\boldsymbol{\beta}}_t = \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{2} \sum_{T_i = t} (Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta})^2 + \lambda_t \|\boldsymbol{\beta}\|_1, \quad t = 0, 1,$$

where $Y_i = Y_i(T_i)$. Thus, we obtain the estimator of CATE as $\hat{\tau}(\boldsymbol{x}) = \boldsymbol{x}^T(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0)$. However, such the procedure estimate $\boldsymbol{\beta}_0^*$ and $\boldsymbol{\beta}_1^*$ separately. The treated (control) outcomes are predicted by treated (control) covariates. Hence, if $\boldsymbol{x}$ is coming from the distribution of $\boldsymbol{X} \,|T = 1$, then $\boldsymbol{x}^T \hat{\boldsymbol{\beta}}_1$ would be accurate but $\boldsymbol{x}^T \hat{\boldsymbol{\beta}}_0$ be not. Moreover, the non-zero elements of $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_0$ usually do not imply zero elements of $\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0$ even when the corresponding elements of $\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_0^*$ are zero.

Our goal is to construct a direct estimation procedure for $\boldsymbol{\theta}^* = \boldsymbol{\beta}_1^* - \boldsymbol{\beta}_0^*$ via the well-known consequence of the strong ignorability condition, given by

$$\tau^*(\boldsymbol{x}) = \mathbb{E}\left[Y_i \left\{ \frac{T_i}{e(\boldsymbol{x})} - \frac{1 - T_i}{1 - e(\boldsymbol{x})} \right\} \middle| \boldsymbol{X}_i = \boldsymbol{x} \right],$$

where $e(\boldsymbol{x}) = \mathbb{P}(T_i = 1 \,|\boldsymbol{X}_i = \boldsymbol{x})$ is the propensity score function at $\boldsymbol{x}$. Thus, $\boldsymbol{\theta}^*$ can be estimated by regressing the appropriately weighted outcomes on the covariates. The propensity score is unknown in most cases. An approach to estimate it in high dimensions may be generalized linear regression with sparse regularization (see, e.g., Fan and Li (2001) and Van de Geer (2008)), but it may lead to an biased estimator for $\boldsymbol{\theta}^*$ when the propensity score function is misspecified.

In this talk, inspired by Athey et al. (2018), a two-step estimation procedure of $\boldsymbol{\theta}^*$ is proposed. The first step obtains weightings for outcomes without specifying the propensity score and then Lasso is applied to the weighted outcomes. The weights are computed by the alternating direction method of multipliers (ADMM) with smoothing technique by Nesterov (2005).

Statistical modeling for electricity load forecasting

Kei Hirose [1,3] and Hiroki Masuda [2]

[1] Institute of Mathematics for Industry, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

[2] Faculty of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

[3] RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan

E-mail: hirose@imi.kyushu-u.ac.jp, hiroki@math.kyushu-u.ac.jp

Short and medium-term load forecasting with high accuracy is essential to decision making in a trade on electricity markets. Among various electricity markets, the day-ahead (or spot) market is popular in many electricity exchanges, including Japan Electric Power Exchange (JEPX) (`http://www.jepx.org/english/index.html`). In the day-ahead market, the contracts for the delivery of electricity on the following day are made, and the transaction is typically carried out in 30 minutes interval; the suppliers must forecast the loads in 30 minutes interval. In this paper, we consider the problem of forecasting electricity consumption, which can be used for the day-ahead market.

The electricity demand is mainly determined by past electricity consumption and external effect, such as weather information. To incorporate the weather information into the regression model, we may use the weather forecast in 30 minutes interval. In general, however, the weather forecast information are obtained per not 30 minutes but daily; for example, only maximum temperature or average humidity is available. Therefore, ordinary linear regression may not perform well.

In this paper, we introduce statistical modeling which elaborately captures the nonlinear structure of the weather factor with limited weather forecast information. We use the varying coefficient model with basis expansion, in which the coefficients are assumed to be different depending on time intervals. The effect of weather variables is expressed as a nonlinear function using basis expansions to capture the yearly seasonal effect of weather information. For interpretation of the estimated model, we eliminate the effect of weather from the past consumption data. Our regression model turns out to be a linear regression model, so that we can apply the least-squares estimation.

In least-squares estimation, however, we found that the regression coefficients concerning the effect of past electricity consumption often become negative, which makes the interpretation of the estimated model difficult. To handle this issue, we employ the non-negative least squares (NNLS; [1], for example) estimation, where we impose a constraint that the regression coefficients are nonnegative. Furthermore, we employ the post-selection inference to construct the prediction interval after the model selection based on [2] and [3].

References

[1] D. Chen and R. J. Plemmons. Nonnegativity constraints in numerical analysis. In *The birth of numerical analysis*, pages 109–139. World Scientific, 2010.

[2] J. D. Lee and J. E. Taylor. Exact Post Model Selection Inference for Marginal Screening. *arXiv.1402.5596*, 2014.

[3] J. D. Lee, D. L. Sun, Y. Sun, and J. E. Taylor. Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3):907–927, June 2016.

Consistency of the objective general index in high dimensional settings

Takuma Bando*, Tomonari Sei* and Kazuyoshi Yata†

1 Introduction

Rankings are often determined by multivariate data. For example, the world university ranking provided by [3] is based on five attributes of universities: teaching, research, citations, industry income and international outlook. For a happiness index of prefectures in Japan [2], 65 attributes are used to make a ranking of 47 prefectures. In heptathlon of athletics, the scores of seven events are unified into an overall score. These rankings are, after some transformations, based on a weighted sum of variables.

We focus on the weights. In [1], an objective weight is proposed via diagonal scaling of the sample covariance matrix. The resultant index called the objective general index (OGI) has positive correlation with all the variables and is invariant with respect to scale transformation of the data.

In some applications like the happiness ranking mentioned above, the number of variables is often large and comparable with the sample size. In other words, we have to deal with high-dimensional data for ranking. If we use the objective general index for such a data, a reliable estimator will be required. The aim of this paper is to study consistency of the weight determined from a random sample.

2 Problem setting and main results

The objective general index is one of possible general indices for multivariate data. We consider a weighted sum $\boldsymbol{w}^\top \boldsymbol{x}$ of an observation $\boldsymbol{x} \in \mathbb{R}^p$ as a

*The University of Tokyo
†University of Tsukuba

general index, where $\boldsymbol{w} \in \mathbb{R}^p$ is a weight vector. Each variate x_i is assumed to have a meaning that "larger is better" without loss of generality. Then it is natural to suppose that every coordinate of $\boldsymbol{w}$ is positive.

Let $n > p$ and consider a random sample $\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(n)} \in \mathbb{R}^p$ according to the multivariate normal distribution $N(\boldsymbol{0}, \boldsymbol{\Sigma})$ with covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$. Denote the sample covariance matrix by $\boldsymbol{S} = n^{-1}\sum_{t=1}^{n} \boldsymbol{x}_{(t)}\boldsymbol{x}_{(t)}^\top$.

Definition 1 ([1])**.** The *objective weight* $\boldsymbol{w}$ is defined by a solution of

$$\boldsymbol{\Sigma}\boldsymbol{w} = \frac{\boldsymbol{1}}{\boldsymbol{w}}, \quad \boldsymbol{w} \in \mathbb{R}^p_{>0}. \tag{1}$$

Similarly, *the sample objective weight* $\hat{\boldsymbol{w}}$ is defined by

$$\boldsymbol{S}\hat{\boldsymbol{w}} = \frac{\boldsymbol{1}}{\hat{\boldsymbol{w}}}, \quad \hat{\boldsymbol{w}} \in \mathbb{R}^p_{>0}. \tag{2}$$

The weighted sum $\boldsymbol{w}^\top \boldsymbol{x}$ of an observation $\boldsymbol{x} \in \mathbb{R}^p$ using the objective weight $\boldsymbol{w}$ is called *the objective general index* (OGI).

We consider a high-dimensional setting in that the dimension p grows with the sample size n. Denote the entries of $\boldsymbol{\Sigma}$ as $(\sigma_{ij})_{i,j=1}^p$.

Theorem 1. *Suppose that $\boldsymbol{\Sigma}\boldsymbol{1} = \boldsymbol{1}$. Then there exists a constant $C > 0$ such that*

$$\mathrm{P}(\|\hat{\boldsymbol{w}} - \boldsymbol{1}\| \geq \varepsilon) \leq 4p \exp\left(-\frac{nC\varepsilon^2}{(\max_i \sigma_{ii})p^2}\right) \tag{3}$$

for any $\varepsilon > 0$ and any $n \geq n_0$ with some $n_0 = n_0(\varepsilon)$. In particular, if $\max_i \sigma_{ii} = \mathrm{O}(1)$ and $(p^2 \log p)/n = \mathrm{o}(1)$ as $n \to \infty$, then $\hat{\boldsymbol{w}}$ is strongly consistent in the sense that $\|\hat{\boldsymbol{w}} - \boldsymbol{1}\|_2$ converges to 0 almost surely.

References

[1] Sei, T. (2016). An objective general index for multivariate ordered data, *J. Multivariate Anal.*, **147**, 247–264.

[2] Terashima, J., Japan Research Institute / Nihon Unisys Ltd. (eds.) (2016) *Happiness ranking of all the 47 prefectures in Japan, 2016* (in Japanese), Toyo Keizai.

[3] Times Higher Education, World University Rankings. https://www.timeshighereducation.com/