科研費シンポジウム

「構造化された大規模複雑データ解析とその統計・機械学習モデルの理論と方法論の展開」
（Theoretical and methodological development of structured large-scale complex data analysis and its statistical and machine learning models）

日時：2025年10月10日（金）～10月12日（日）
(Date: Friday, 10, October – Sunday, 12, October, 2025)

場所：
2025年10月10日（金）STATION Ai M3F 大会議室１・２
2025年10月11日（土），12日（日）南山大学 G棟 25教室
Place:
October 10, 2025 (Friday) STATION Ai, M3F Large Conference Rooms 1, 2
October 11-12, 2025 (Saturday-Sunday) Nanzan University, Building G Room 25

科学研究費・基盤研究（A）（課題番号：25H01107）
「大規模複雑データの理論と方法論の深化と展開」
研究代表者：青嶋 誠（筑波大学）
Makoto Aoshima (University of Tsukuba)
開催責任者：蛭川 潤一，塩濱 敬之（南山大学）
Junichi Hirukawa, Takayuki Shiohama (Nanzan University)

Program

Friday, 10, October

Registration & Opening
13:00 - 13:10 **Junichi Hirukawa** (Nanzan University)

Session I - Friday Afternoon Session I (English) 13:10 - 14:30
Chair **Hannah L. H. Lai** (National University of Singapore)
1. 13:10 - 13:50
Liu Guo (Waseda University)
A New Lasso Refitting Strategy
2. 13:50 - 14:30
Shi Chen (Waseda University)
Yan Liu (Waseda University)
Portmanteau tests for copula time series

Coffee Break 14:30 - 14:50

Session II - Friday Afternoon Session II 14:50 - 16:50
Chair **Yuichi Goto** (Kyushu University)
3. 14:50 - 15:30 **吉田 耀晟 (Yosei Yoshida)**, 喜多 美琴 (Mikoto Kita), 劉 言 (Yan Liu)
早稲田大学 (Waseda University)
Multivariate linear process bootstrap for testing covariance structures in high-dimensional time series
4. 15:30 - 16:10 **巖名 佑務**, 石井 晶, 矢田 和善, 青嶋 誠
筑波大学数理物質科学研究群, 東京理科大学創域理工学部, 筑波大学数理物質系, 筑波大学数理物質系
単一強スパイク固有値モデルにおける高次元相関行列の検定
5. 16:10 - 16:50 **柳本 武美 (Takemi Yanagimoto)**, 小椋 透 (Tohru Ogura)
統計数理研究所 (Institute of Statistical Mathematics), 三重大学 (Mie University)
バイアス低減法による条件付き尤度法の改良 (Improving inferential procedures through the conditional likelihood method by those through the bias reduction method)

Saturday, 11, October

Session III - Saturday Morning Session 10:00 - 10:40
Chair **Takayuki Shiohama** (Nanzan University)
6. 10:00 - 10:40 **藤森 洸 (Kou Fujimori)**, 白石 博 (Hiroshi Shiraishi), 蛭川 潤一 (Junichi Hirukawa), Konstantinos Fokianos
信州大学 (Shinshu University), 慶應義塾大学 (Keio University), 南山大学 (Nanzan University), University of Cyprus
Sparse estimators for multivariate integer-valued autoregressive models with applications to inference for Hawkes processes

Lunch 10:40-13:20

Session IV - Saturday Afternoon Session I (English) 13:20-15:20
Chair **Lei Zhou** (National University of Singapore)
7. 13:20-14:00 **Yuichi Goto**, Abdelhakim Aknouche, Christian Francq
Kyushu University, Qassim University, ENSAE-CREST
Flexible Modeling of $\mathbb{Z}$-valued Time Series
8. 14:00-14:40 **Muneya Matsui**
Nanzan University
Self-normalized partial sums of heavy-tailed time series
9. 14:40-15:20 **Masanobu Taniguchi**, Anna Clara Monti, Yujie Xue
Waseda University, University of Sannio (Italy), ISM (Japan)
Higher-order investigation of general time series divergences

Coffee Break 15:20 - 15:40

Session V - Saturday Afternoon Session II (Guest Session) 15:40 - 17:40

Chair **Junichi Hirukawa** (Nanzan University)

10. 15:40 - 16:40 **Hannah L. H. Lai**, Ying Chen, Maria Grith

National University of Singapore, National University of Singapore, Erasmus University (Rotterdam)

Neural Tangent Kernel in Implied Volatility Forecasting: A Nonlinear Functional Autoregression Approach

11. 16:40 - 17:40 **Lei Zhou**, Ying Chen

National University of Singapore

Debt Structure and Recovery Rates

Sunday, 12, October

Session VI - Sunday Morning Session 10:00 - 12:40
Chair **Kou Fujimori** (Shinshu University)
12. 10:00 - 10:40 阿部 俊弘
法政大学
Cauchy 型の EM アルゴリズムとその加速法について
13. 10:40 - 11:20 井本 智明
静岡県立大学
シンプルなトーラス上分布の隠れマルコフモデルへの応用
14. 11:20 - 12:00 宮田 庸一，塩濱 敬之，阿部 俊弘
高崎経済大学，南山大学，法政大学
円柱上の統計モデルを出力分布に持つ隠れマルコフモデルについて (A Hidden Markov Model with an Emission Distribution on the Cylinder)
15. 12:00 - 12:40 清水 邦夫
統計数理研究所 大学統計教員育成センター
（ハイパー）シリンダー上の分布構築に向けて

Closing 12:40 - 12:45 **Junichi Hirukawa** (Nanzan University)

A New Lasso Refitting Strategy

Liu Guo[1]

[1]Department of Pure and Applied Mathematics, Waseda University

1 Introduction

The least absolute shrinkage and selection operator (Lasso) is a popular method for high-dimensional statistics. However, it has limitations such as bias and prediction error. To address such disadvantages, many alternatives and refitting strategies have been proposed and studied. This work introduces a novel Lasso–Ridge method. We show that our estimator achieves better prediction performance than the standard Lasso, even when its tuning parameter is set at the optimal rate $\sqrt{\log p/n}$. Moreover, the proposed method retains key advantages of the Lasso, including prediction consistency and reliable variable selection under standard conditions. Through extensive simulations, we further demonstrate that our estimator consistently outperforms the Lasso in both prediction and estimation accuracy, highlighting its potential as a powerful tool for high-dimensional linear regression. Details and references are not included here due to space limitations, but they will be presented and discussed during the conference talk.

2 Methodology

We consider a standard linear regression model given by:

$$y = X\beta^0 + \epsilon, \;\; \epsilon \sim N(0, \sigma^2 I_n),$$

where $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n\times p}$ is the design matrix, β^0 denotes the unknown coefficient vector, and ϵ is the noise vector. A new Lasso-Ridge estimator could be defined as:

$$\hat{\beta} = \operatorname*{arg\,min}_{\beta\in\mathbb{R}^p, \beta_{-E}=0} \left\{ \frac{1}{2n}\|y - X\beta\|_2^2 + \frac{\lambda}{2}\|\beta - \bar{\beta}\|_2^2 \right\},$$

where $\bar{\beta}$ is the Lasso estimator, defined as follows:

$$\bar{\beta} = \operatorname*{arg\,min}_{\beta\in\mathbb{R}^p} \left\{ \frac{1}{2n}\|y - X\beta\|_2^2 + \lambda_L\|\beta\|_1 \right\},$$

and E is the equicorrelation set, defined by:

$$E \coloneqq \left\{ i \in \{1, 2, \dots, p\} : |X_i^T(y - X\bar{\beta})/n| = \lambda_L \right\}.$$

3 Theoretical Results

We summarize the main theoretical properties as follows:

1. A theoretical guarantee of improved prediction under mild assumptions on the design, noise, and Lasso tuning parameter.

2. Retention of key strengths of the Lasso, including variable selection and prediction consistency.

4 Numerical Studies

We adopt a simulation framework similar to that used in the Relaxed Lasso experiments to compare the performance of our estimator with that of the Lasso under diverse conditions. The response variable follows the linear model:

$$y = X\beta + \varepsilon,$$

where the design matrix $X \in \mathbb{R}^{n\times p}$ consists of rows independently drawn from a multivariate normal distribution with zero mean and covariance matrix Σ, given by

$$\Sigma_{ij} = \rho^{|i-j|}, \quad \text{with } \rho \in \{0, 0.5\}.$$

The noise vector $\varepsilon \in \mathbb{R}^n$ follows a normal distribution $\mathcal{N}(0, \sigma^2 I_n)$, with $\sigma \in \{0.5, 1\}$ representing different noise level. The response vector y is then generated according to the linear model above. We consider sample sizes $n \in \{50, 100, 200\}$ and numbers of predictors $p \in \{100, 200, 400\}$, and the number of relevant (nonzero) coefficients is $s \in \{5, 10, 30\}$. For $k \leq s$, we set $\beta_k = 1$, and for $k > s$, we set $\beta_k = 0$. In addition to the synthetic simulations, we have also conducted experiments on semi-real and real datasets to further demonstrate the practical performance of the proposed refinement, and the results will be presented and discussed during the conference talk. In summary, the simulation results demonstrate that the Lasso–Ridge estimator generally outperforms the Lasso across a variety of scenarios.

5 Conclusion and Further Studies

In this work, we proposed a Lasso–Ridge estimator. We provided theoretical guarantees under minimal assumptions, and empirical evidence demonstrated consistent gains over the standard Lasso across various simulation regimes.

Several directions remain open for future research. First, extending the theoretical analysis to more general noise structures and non-Gaussian designs would broaden the application of the method. Second, the integration of adaptive or data-driven tuning strategies within the refinement framework could further improve robustness in practice. Finally, extending the method to broader model classes, such as generalized linear models or structured sparsity settings, represents a promising direction for future work.

Portmanteau tests for copula time series

Shi Chen

(Waseda Univ.)*

Yan Liu

(Waseda Univ.)*

1. Summary

This study introduces two portmanteau tests for copula time series models. Copula provides a framework to separate marginal properties and correlation structure. We define a copula-based Markov process and propose portmanteau tests as diagnostic methods. The first test uses residuals, and the asymptotic distribution of its sample correlation is a weighted chi-square distribution; the second applies the probability integral transform (PIT) to recover i.i.d. innovations. Simulation and empirical results show that PIT-based methods are generally more stable and more powerful.

2. Methodology

2.1 Copula Time Series and Generation

Consider a Markov process $\{X_t.t \in Z\}$. Denote $\boldsymbol{F}$ as the joint distribution of $X_t, X_{t-1}, \dots, X_{t-k}, \forall t$. A copula time series of order k is defined by

$$\boldsymbol{F}(x_0, \dots, x_k) = C(F(x_0), \dots, F(x_k)),$$

where C is a $(k+1)$-dimensional copula and F is the marginal CDF. The process is strictly stationary.

Generation Algorithm. Given initial values $\{X_t\}_{t=1}^{k}$:

1. Transform $U_t = F(X_t)$ for $t \leq k$.
2. For $t = k+1, \dots, n$:
   (a) Draw $V_t \sim U(0,1)$.
   (b) Compute the conditional distribution
   $$H_k(x) = \frac{C^{[0,1,\dots,1]}(x, U_{t-1}, \dots, U_{t-k})}{C^{[0,1,\dots,1]}(1, U_{t-1}, \dots, U_{t-k})}.$$
   (c) Solve $U_t = H_k^{-1}(V_t)$, then $X_t = F^{-1}(U_t)$.

2.2 Portmanteau Tests of copula time series

(1) Residual-based Test. Define residuals $\xi_t = X_t - \hat{X}_t$ where $\hat{X}_t$ is the conditional predictor from H_k^{-1}. Compute sample autocorrelations $\hat{R}_l$, and the Ljung–Box statistic:

$$Q_m = n(n+2)\sum_{l=1}^{m}\frac{\hat{R}_l^2}{n-l}.$$

Under the null hypothesis, we conclude that the asymptotic distribution is

$$Q_m \xrightarrow{d} \sum_{l=1}^{m} c_l\chi_1^2(l), \qquad c_l = \frac{E[\xi_t^2\xi_{t-l}^2]}{E[\xi_t^2]^2},$$

where c_l could be calculated analytically or estimated by sample estimation. We could estimate the distribution of Q_m by a Gamma distribution.

(2) Probability Integral Transform-based Test. With the given conditional CDF $\hat{H}_k$. Compute innovations $u_t = \hat{H}_k(U_t)$. If the model is correctly specified, u_t are i.i.d. $U(0,1)$. Then apply the Ljung–Box test to u_t for p value. The asymptotic distribution should be a standard chi-square distribution.

3. Questions and Answers

(i) **Generally, we do not use the non-standard chi-square distribution. Can we use some techniques to make its asymptotic distribution become a standard chi-square distribution?**

Since all asymptotic variances c_l are available, each $\hat{R}_l$ can be standardized. Then the Ljung–Box statistic will follow a standard χ_m^2.

(ii) **How to estimate the copula model?**

We could estimate parameters by maximizing the likelihood based on PIT:

$$\ell(\theta) = \sum_{t=k+1}^{n} \log h_k(U_t \mid U_{t-1}, \ldots, U_{t-k}; \theta),$$

where h_k is the conditional copula density. This corresponds to maximizing the probability that the recovered innovations are i.i.d. $U(0,1)$.

(iii) **How does the performance compare with existing algorithms?**

There are no existing portmanteau tests for copula time series yet, but we could compare their performance with others as goodness-of-fit tests.

References

Akashi, F., Taniguchi, M., Monti, A. C. and Amano, T. (2021). *Diagnostic Methods in Time Series.* Springer.

Berg, D. and Bakken, H. (2005). Copula Goodness-of-fit Tests: A Comparative Study. Technical report, Norwegian Computing Center and University of Oslo.

Sun, L.-H., Huang, X.-W., Alqawba, M. S., Kim, J.-M. and Emura, T. (2020). *Copula-based Markov Models for Time Series: Parametric Inference and Process Control.* Springer Nature.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* Springer Science & Business Media.

Multivariate linear process bootstrap for sphericity testing in high-dimensional time series

Yoshida, Yosei (Waseda Univ.)
Kita, Mikoto (Waseda Univ.)
Liu, Yan (Waseda Univ.)

1. Introduction

We use two statistics U and V, following the terminology of Ledoit and Wolf (2002) and Nagao (1973). They are defined respectively as

$$U = \frac{1}{p} \operatorname{tr} \left[\left(\frac{\boldsymbol{S}}{(1/p) \operatorname{tr} \boldsymbol{S}} - \boldsymbol{I}_p \right)^2 \right] \quad \text{and} \quad V = \frac{1}{p} \operatorname{tr} \left[(\boldsymbol{S} - \boldsymbol{I}_p)^2 \right],$$

where $\boldsymbol{S}$ denotes the sample covariance matrix, p is the data dimension, and $\boldsymbol{I}_p$ is the p-dimensional identity matrix. These statistics are utilized to assess structural properties of the covariance matrix.

Recent work, including Liu et al. (2018) and Yoshida and Liu (2024), has successfully extended the asymptotic theory of statistics U and V to high-dimensional time series with temporal dependence. These studies established the asymptotic normality of the statistics under the regime $p/N \to c \in (0, \infty)$ as both the sample size N and dimension p diverge. However, their derivations of the bias and asymptotic variance require specific structural assumptions on the data-generating process, such as Gaussianity and linear process forms.

Motivated by this limitation, the present study introduces a bootstrap-based refinement using the Multivariate Linear Process Bootstrap (MLPB) under the null hypothesis. Unlike classical asymptotic theory, which often fails to capture the finite-sample behavior of eigenvalue-based statistics in high-dimensional settings (Ledoit and Wolf (2002), Bai et al. (2018), Zheng et al. (2015)), the MLPB provides a flexible resampling scheme that preserves the second-order dependence structure of the data. As a result, we improved theoretical justification and make better finite-sample performance for sphericity testing in high-dimensional, temporally dependent time series.

2. Model

We describe the high-dimensional time series model and present the null and alternative hypotheses for testing sphericity. We also clarify the asymptotic regime and modeling assumptions adopted throughout this study.

Let $\{\boldsymbol{X}(t) = (X_1(t), X_2(t), \ldots, X_p(t))^\top : t \in \mathbb{Z}\}$ be the p-dimensional Gaussian stationary processes with $p \times p$ covariance function $\boldsymbol{R}(h) := (R_{ij}(h))$.

We focus on the problem of testing the sphericity hypothesis:

$$H_0 : \boldsymbol{R}(0) = \sigma^2 I_p \quad \text{v.s.} \quad A : \boldsymbol{R}(0) \neq \sigma^2 I_p, \quad \text{with } \sigma^2 > 0,$$

when the sample size n and the dimension p diverge to infinity at the same asymptotic order, where $\boldsymbol{I}_p \in \mathbb{R}^{p \times p}$ denotes the identity matrix.

Suppose the observation stretch $\boldsymbol{X}(1), \boldsymbol{X}(2), \ldots, \boldsymbol{X}(N)$ is obtained from the process $\{\boldsymbol{X}(t)\}$. Let us define the sample covariance matrix as

$$\boldsymbol{S} = \frac{1}{N-1} \sum_{t=1}^{N} (\boldsymbol{X}(t) - \bar{\boldsymbol{X}})(\boldsymbol{X}(t) - \bar{\boldsymbol{X}})^\top.$$

3. Sphericity MLPB Procedure

We refer to the bootstrap algorithm that uses $\widehat{\mathbf{\Gamma}}^{\epsilon,\mathcal{H}_0}_{\kappa,\ell}$ in place of $\widehat{\mathbf{\Gamma}}^{\epsilon}_{\kappa,\ell}$ as the *Sphericity MLPB*. The procedure is summarized as follows:

1. Let $\boldsymbol{Y}(t) = \boldsymbol{X}(t) - \overline{\boldsymbol{X}}$ and $\boldsymbol{Y} := \mathrm{vec}([\boldsymbol{Y}(1) : \ldots : \boldsymbol{Y}(N)])$.
2. Compute $\boldsymbol{W} = (\widehat{\mathbf{\Gamma}}^{\epsilon,\mathcal{H}_0}_{\kappa,\ell})^{-1/2}\boldsymbol{Y}$, where $(\widehat{\mathbf{\Gamma}}^{\epsilon,\mathcal{H}_0}_{\kappa,\ell})^{-1/2}$ denotes the lower left triangular matrix $\boldsymbol{L}^{\mathcal{H}_0}$ of the Cholesky decomposition $\widehat{\mathbf{\Gamma}}^{\epsilon,\mathcal{H}_0}_{\kappa,\ell} = \boldsymbol{L}^{\mathcal{H}_0}\boldsymbol{L}^{\mathcal{H}_0\top}$.
3. Let $\boldsymbol{Z}$ be the standardized $\boldsymbol{W}$, that is, $\boldsymbol{Z}_i = (W_i - \overline{W})/\widehat{\sigma}_W, i = 1, \ldots, pN$, where $\overline{W} = \frac{1}{pN}\sum_{i=1}^{pN} W_i$ and $\widehat{\sigma}^2_W = \frac{1}{pN}\sum_{i=1}^{pN}(W_i - \overline{W})^2$.
4. Generate bootstrap sample $\boldsymbol{Z}^*$ by i.i.d. resampling from $\{Z_1, \ldots, Z_{pN}\}$.
5. Reconstruct the bootstrap sample as $\boldsymbol{Y}^* = (\widehat{\mathbf{\Gamma}}^{\epsilon,\mathcal{H}_0}_{\kappa,\ell})^{1/2}\boldsymbol{Z}^* = \mathrm{vec}([\boldsymbol{Y}^*(1) : \ldots : \boldsymbol{Y}^*(N)])$.

4. Theory

Theorem 0.1. *Suppose some Assumptions. Under the null hypothesis $H_0 : \boldsymbol{R}(0) = \sigma^2\boldsymbol{I}_p$, we obtain*

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left(\sqrt{np}\ \mathscr{A}^{-1/2}(U - \Delta_u) \le u\right) - \mathbb{P}^*\left(\widehat{\sigma}^{-1}_{u^*}(U^* - \widehat{\Delta}_{u^*}) \le u\right)\right| = o_P(1),$$

and

$$\sup_{v\in\mathbb{R}}\left|\mathbb{P}\left(\sqrt{np}\ \mathscr{B}^{-1/2}(V - \Delta_v) \le v\right) - \mathbb{P}^*\left(\widehat{\sigma}^{-1}_{v^*}(V^* - \widehat{\Delta}_{v^*}) \le u\right)\right| = o_P(1),$$

where $\mathbb{P}^*$ denotes the conditional probability, and $\widehat{\Delta}_{u^*}$ and $\widehat{\sigma}^2_{u^*}$, as well as $\widehat{\Delta}_{v^*}$ and $\widehat{\sigma}^2_{v^*}$, are the bootstrap estimators of the bias and asymptotic variance of U and V, respectively.

During the Q&A session following the presentation, several insightful questions were raised. One was whether the proposed method could be extended to non-Gaussian processes. I received advice that the linear bootstrap might not perform well in such cases, and that a hybrid bootstrap approach could be a potential alternative. Another question concerned why the results deteriorate when $p = 32$ and $N = 32$. Finally, there was a question about why the distributions were not plotted in the simulation results. I explained that the focus of this study was on the behavior of the test statistics, but visualizing the distributions would be a valuable ingredients for future work.

References

Zhidong Bai, Kwok Pui Choi, and Yasunori Fujikoshi. Limiting behavior of eigenvalues in high-dimensional manova via rmt. *The Annals of Statistics*, 46(6A):2985–3013, 2018.

Carsten Jentsch and Dimitris N. Politis. Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *Annals of Statistics*, 43(3):1117–1140, 2015. doi: 10.1214/14-AOS1301.

Olivier Ledoit and Michael Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4):1081–1102, 2002. doi: 10.1214/aos/1031689018.

Yan Liu, Yurie Tamura, and Masanobu Taniguchi. Asymptotic theory of test statistic for sphericity of high-dimensional time series. *Journal of Time Series Analysis*, 39(3):402–416, 2018. doi: 10.1111/jtsa.12288.

Hisao Nagao. On some test criteria for covariance matrix. *The Annals of Statistics*, pages 700–709, 1973.

Yosei Yoshida and Yan Liu. A note on testing for homoscedasticity in high-dimensional time series. *Research in Statistics*, 2 (1):2394557, 2024. doi: 10.1080/27684520.2024.2394557.

Shurong Zheng, Zhidong Bai, and Jianfeng Yao. Substitution principle for clt of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. 2015.

単一強スパイク固有値モデルにおける高次元相関行列の検定

筑波大学・数理物質科学研究群　　巖名　佑務
筑波大学・数理物質系　　矢田　和善
東京理科大学・創域理工学部　　石井　晶
筑波大学・数理物質系　　青嶋　誠

1　はじめに

Aoshima and Yata [2] は，弱スパイク固有値 (Non-Strongly Spiked Eigenvalue) モデルと強スパイク固有値 (Strongly Spiked Eigenvalue) モデルの 2 つの固有値モデルを考案した．本講演では，強スパイク固有値モデルのもとで拡張クロスデータ行列法 (ECDM) を用いた新たな検定統計量を与え，その漸近分布を用いた高次元相関行列の検定手法を提案した．高次元データにおける相関行列の検定は，Aoshima and Yata [1]，Yata and Aoshima [5, 6] によって，弱スパイク固有値モデルのもとで ECDM を用いた検定統計量が与えられた．母集団分布に p 次元の分布を考え，n 個のデータ $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ を無作為に抽出する．ただし，$\boldsymbol{x}_j = (\boldsymbol{x}_{1j}^\top, \boldsymbol{x}_{2j}^\top)^\top, j = 1, ..., n$ とし，$\boldsymbol{x}_{ij} \in \mathbb{R}^{p_i}$，$p_1 \in \{1, ..., p-1\}$，$p_2 = p - p_1$ とする．ここで，$\boldsymbol{x}_j$ は未知の平均ベクトル $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$，共分散行列

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_* \\ \boldsymbol{\Sigma}_*^\top & \boldsymbol{\Sigma}_2 \end{pmatrix} (\geq \boldsymbol{O})$$

をもつとする．各 i, j で $\mathrm{E}(\boldsymbol{x}_{ij}) = \boldsymbol{\mu}_i, \mathrm{Var}(\boldsymbol{x}_{ij}) = \boldsymbol{\Sigma}_i, \mathrm{Cov}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = \mathrm{E}(\boldsymbol{x}_{1j}\boldsymbol{x}_{2j}^\top) - \boldsymbol{\mu}_1\boldsymbol{\mu}_2^\top = \boldsymbol{\Sigma}_*$ とする．相関行列を $\boldsymbol{P} = \mathrm{diag}(\sigma_{11}, \sigma_{12}, \dots, \sigma_{1p_1})^{-1/2}\boldsymbol{\Sigma}_*\mathrm{diag}(\sigma_{21}, \sigma_{22}, \dots, \sigma_{2p_2})^{-1/2}$ とおく．ただし，$\sigma_{i1}, \sigma_{12}, \dots, \sigma_{ip_i}$ は $\boldsymbol{\Sigma}_i$ の対角成分である．このとき，次の検定を考える．

$$H_0 : \boldsymbol{P} = \boldsymbol{O} \ \ vs. \ \ H_1 : \boldsymbol{P} \neq \boldsymbol{O}. \tag{1}$$

2　単一強スパイク固有値モデルにおける相関行列の検定

次を仮定する．

(A-i)　p_1は固定;　$\dfrac{\lambda_{21}}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_2^2)}} \to 1, \quad p_2 \to \infty$. ただし，$\lambda_{21}$ は $\boldsymbol{\Sigma}_2$ の最大固有値.

固有値モデル (A-i) は強スパイク固有値モデルの 1 つであり，Ishii et al.[3] で定義され，単一強スパイク固有値モデルと呼ばれる．$\Delta = \mathrm{tr}(\boldsymbol{\Sigma}_*\boldsymbol{\Sigma}_*^\top)\ (= \|\boldsymbol{\Sigma}_*\|_F^2)$ とおく．ただし，$\|\cdot\|_F$ はフロベニウスノルムを表す．ここで，ECDM を用いた Δ の不偏推定量について考える．$n_{(1)} = \lceil n/2 \rceil$，$n_{(2)} = n - n_{(1)}$ とおく．ここで，$\lceil x \rceil$ は x 以上の最小の整数を表す．2 つの集合 $\boldsymbol{V}_{n(1)(k)}$，$\boldsymbol{V}_{n(2)(k)}$，$k = 3, \dots, 2n-1$ を次のように定義する．

$$\boldsymbol{V}_{n(1)(k)} = \begin{cases} \{\lfloor k/2 \rfloor - n_{(1)} + 1, ..., \lfloor k/2 \rfloor\}, & \lfloor k/2 \rfloor \geq n_{(1)} \text{のとき}, \\ \{1, \dots, \lfloor k/2 \rfloor\} \cup \{\lfloor k/2 \rfloor + n_{(2)} + 1, ..., n\}, & \text{それ以外}. \end{cases}$$

$$\boldsymbol{V}_{n(2)(k)} = \begin{cases} \{\lfloor k/2 \rfloor + 1, ..., \lfloor k/2 \rfloor + n_{(2)}\}, & \lfloor k/2 \rfloor \leq n_{(1)} \text{のとき}, \\ \{1, ..., \lfloor k/2 \rfloor - n_{(1)}\} \cup \{\lfloor k/2 \rfloor + 1, ..., n\}, & \text{それ以外}. \end{cases}$$

ただし，$\lfloor x \rfloor$ は x 以下の最大の整数を表す．そのとき，$k=3,...,2n-1$ について，$\#\boldsymbol{V}_{n(\ell)(k)}=n_{(\ell)}$，$\ell=1,2,\quad \boldsymbol{V}_{n(1)(k)}\cap\boldsymbol{V}_{n(2)(k)}=\emptyset,\quad \boldsymbol{V}_{n(1)(k)}\cup\boldsymbol{V}_{n(2)(k)}=\{1,...,n\}$ となること，及び，$i<j\ (\le n)$ について，$i\in\boldsymbol{V}_{n(1)(i+j)},\quad j\in\boldsymbol{V}_{n(2)(i+j)}$ となることに注意する．ここで，$\#\boldsymbol{S}$ は集合 $\boldsymbol{S}$ の要素の個数を表す．$\ell\ (=1,2)$，$k\ (=3,...,2n-1)$ について，2 分割した集合の平均を $\overline{\boldsymbol{x}}_{\ell(1)(k)}=n_{(1)}^{-1}\sum_{j\in\boldsymbol{V}_{n(1)(k)}}\boldsymbol{x}_{\ell j},\quad \overline{\boldsymbol{x}}_{\ell(2)(k)}=n_{(2)}^{-1}\sum_{j\in\boldsymbol{V}_{n(2)(k)}}\boldsymbol{x}_{\ell j}$ とし，ある $i,j\ (1\le i<j\le n)$ について，Δ の 1 つの不偏推定量として

$$\widehat{\Delta}_{ij}=u_n(\boldsymbol{x}_{1i}-\overline{\boldsymbol{x}}_{1(1)(i+j)})^\top(\boldsymbol{x}_{1j}-\overline{\boldsymbol{x}}_{1(2)(i+j)})(\boldsymbol{x}_{2i}-\overline{\boldsymbol{x}}_{2(1)(i+j)})^\top(\boldsymbol{x}_{2j}-\overline{\boldsymbol{x}}_{2(2)(i+j)})$$

を計算する．ただし，$u_n=n_{(1)}n_{(2)}/\{(n_{(1)}-1)(n_{(2)}-1)\}$ である．全ての組み合わせで平均を取り，

$$\widehat{T}_n=\frac{2u_n}{n(n-1)}\sum_{i<j}^{n}\widehat{\Delta}_{ij}$$

を定義する．このとき，$\mathrm{E}(\widehat{T}_n)=\Delta$ となることに注意する．

定理 1 ([4])**.** *(A-i)* といくつかの正則条件を仮定する．$m=\min\{p,n\}\to\infty$ のとき次が成り立つ．

$$\frac{n(\widehat{T}_n-\Delta)}{\mathrm{tr}(\boldsymbol{S}_1)\lambda_{NR(21)}}+1\Rightarrow\sum_{s=1}^{p_1}\frac{\lambda_{1s}\chi_{1s}^2}{\mathrm{tr}(\boldsymbol{\Sigma}_1)}.$$

ただし，$\boldsymbol{S}_1$ は $\boldsymbol{x}_{1j}$ の標本分散共分散行列，$\lambda_{NR(21)}$ はノイズ掃き出し法による λ_{21} の推定量，λ_{1s} は $\boldsymbol{\Sigma}_1$ の s 番目の固有値，$\chi_{1s}^2, s=1,\dots,p_1$ は互いに独立な自由度 1 のカイ二乗分布に従う確率変数を表す．

参考文献

[1] M. Aoshima, K. Yata, Two-stage procedures for high-dimensional data. Sequential Anal. (*Editor's special invited paper*) 30 (2011) 356-399.

[2] M. Aoshima, K. Yata, Two-sample tests for high-dimension, strongly spiked eigenvalue models. Statist. Sinica 28 (2018) 43-62.

[3] A. Ishii, K. Yata, M. Aoshima, Hypothesis tests for high-dimensional covariance structures. Ann. Inst. Statist. Math. 73 (2021) 599-622.

[4] Y. Iwana, A. Ishii, K. Yata, M. Aoshima, Correlation tests for high-dimensional data under the strongly spiked eigenvalue model. Revised in Jpn. J. Stat. Data Sci. (2025).

[5] K. Yata, M. Aoshima, Correlation tests for high-dimensional data using extended cross-data-matrix methodology. J. Multivariate Anal. 117 (2013) 313-331.

[6] K. Yata, M. Aoshima, High-dimensional infrence on covariance structures via the extended cross-data-matrix methodology. J. Multivariate Anal. 151 (2016) 151-166.

バイアス低減法による条件付き尤度法の改良

柳 本 武 美: 統計数理研究所
小 椋 透 : 三重大学

1. 序

R^1 上の大きさ n の標本ベクトル $\boldsymbol{x}$ が従う分布の族を $\{p(\boldsymbol{x}|\theta)|\theta \in \Theta \subset R^p\}$ とする。ベイズ推定では、事前密度 $\pi(\theta)$ を仮定して推論を行う。推測法は事後密度 $\pi(\theta|\boldsymbol{x})$ に基づいて構成される。母数 θ の推定は事後平均 $\hat{\theta} = \mathrm{E}_{\mathrm{post}}[\theta]$ で行うことが多い。また、事後モードも用いられる。

この発表では、条件付き尤度法を改めて評価して、その推測方式がバイアス低減事前関数を仮定した下でのベイズ推論方式で代替する可能性を議論した。

2. 条件付き尤度法とバイアス低減法

2.1. 条件付き尤度法

母数の次元 $p = 2$ として、$\theta = (\xi, \psi)$ と書く。条件付き尤度法では、統計量 t が存在して、標本密度 $p(\boldsymbol{x}|\xi, \psi)$ が t の周辺密度と t を与えたときの条件付き密度が $p(\boldsymbol{x}|\xi, \psi) = pc(\boldsymbol{x}|t;\psi) \cdot pm(t|\xi, \psi)$ と分解出来ることを前提とする。ここで条件付き密度は ψ のみに依存する。この前提は強い制約条件である。興味のある因子 ψ は $\hat{\psi}_C = \mathrm{Argmax}_\psi\, pc(\boldsymbol{x}|t;\psi)$ で推定する。因子 ξ は邪魔者因子として推定しない。Profile 尤度推定方程式 $\partial \log p(\boldsymbol{x}|\xi_{ML}(\psi), \psi)/\partial\psi = 0$ は不偏ではないが、条件付き尤度推定方程式 $\mathrm{d} \log pc(\boldsymbol{x}|t;\psi)/\mathrm{d}\psi = 0$ は弱い正則条件の下で不偏になる。

2.2. バイアス低減法

ベイズ推定における、無情報事前関数の導出規準は様々提案されてきた。Reference prior, Jeffreys prior がよく知られている。Firth (1993) は事後モード $\hat{\theta}_F$ の漸近バイアスが $\mathrm{E}\{\hat{\theta}_F - \theta; p(\boldsymbol{x}|\theta)\} = o(1/n)$ となる事前 (罰金) 関数を提案している。最近、Miyata and Yanagimoto (2024) は事後平均の漸近バイアスを低減させる事前関数が満たすべき条件式を与えた。その条件は制約的であるが、条件付き推論で必要な分解条件のような非常に強い制約ではない。

3. 条件付き尤度法の限界

条件付き尤度法は、通常の (無条件) 尤度法の改良として考案された。大きな欠陥は制約の条件が強いことにある。また、因子 ξ が邪魔者であるとの仮定は適用範囲を狭める。データを記述するモデルの母数の因子はそれぞれに意味があることが通常だからである。

確かに、条件付き尤度推測法は無条件尤度推測法を改善する (Yanagimoto and Anraku, 1989)。しかし、芳しくない後者の性能を少しだけ改善しているだけの可能性がある。実際、MLE は通常小標本の下ではその性能は改良の余地がある。また、母数の次元が大きくなると更に悪くなる。実

際ある強い制約条件の下ではベイズ推定量が最善で MLE が最悪である条件を与えることが出来る (Yanagimoto and Ohnishi, 2011)。これらの事実は、単純な事前関数であっても条件付き推測法を改善させることが期待される。バイアス低減事前関数と条件付き尤度推定方程式は、バイアスの低減の目標を共有している。

4. 二つの二項分布の場合

条件付き推測が適用される代表的な場合である、二つの二項分布 $x \sim \mathrm{Bi}(n,p)$ $y \sim \mathrm{Bi}(m,q)$ の場合を議論する。Poisson 分布へも拡張できる。

4.1. オッズ比の推測

母数 $(\xi,\psi) = (np+mq, \log p(1-q)/q(1-p))$ を考える。通常は因子 ψ に関心があるとされ、因子 ξ は邪魔者であると扱われる。因子 ψ は統計量 $t = x+y$ を与えた下での条件付き MLE で推定する。検定問題 $\mathrm{H}_0 : \psi = 0$, $\mathrm{H}_1 : \psi > 0$ は Fisher の厳密検定が標準である。しかし、Fisher の厳密検定はその検出力が低いことから批判の対象である。

ロジット変換 $\gamma_1 = \log p/(1-p)$, $\gamma_2 = \log q/(1-q)$ を施すと、 母数 (γ_1,γ_2) は指数分布族と見たときの自然母数となる。自然母数の事後平均には二つの最適性を満たす (Corcuera and Giummole, 1999)。対数オッズ比 ψ とは $\psi = \gamma_1 - \gamma_2$ で結ばれる。この母数のバイアス低減 prior は $\pi_{BR}(\gamma_1,\gamma_2) \propto \exp(\gamma_1+\gamma_2)/(1+\exp\gamma_1)^2(1+\exp\gamma_2)^2$ で与えられる。

検定問題 $\mathrm{H}_0 : \psi = 0$, $\mathrm{H}_1 : \psi > 0$ は信用水準 $\mathrm{cl}(x,y) = \Pr\{A; \pi(\xi,\psi|x,y)\}$, 但し $A = \{(\xi,\psi)|\psi > 0\}$, に基づいて構成できる。仮説検定の水準を遵守するためには、検定統計量を $\mathrm{T}(x,y) = \Pr\{C(x,y)\}$, 但し $C(x,y) = \{(x',y')|\mathrm{cl}(x',y') \geq \mathrm{cl}(x,y)\}$, として構成する。このベイズ検定の拡張はすべての場合に適用できる。この検定方式について、Ogura and Yanagimoto (2016) は条件付き検定法である McNemar 検定を例にして論じている。

4.2. リスク比の推測

別の母数として 対数変換を施して $\eta_1 = \log p$, $\eta_2 = \log q$ の場合を議論する。この場合、対数リスクの差を $\psi = \eta_1 - \eta_2$ とおいても、条件付き推測に必要な分解式は成り立たない。尤度法での母数推定では MLE が用いられる。このために、リスクの推測は扱い難いと見られてきた。

一方、バイアス低減 prior は $\pi_{BR}(\eta_1,\eta_2) \propto \exp(\eta_1+\eta_2)/(1-\exp\eta_1)(1-\exp\eta_2)$ となる。推論方式は対数オッズ比の場合と同様に議論できる。

文献：1) Corcuera and Giummole (1999). Scand. J. Stat., 26, 265-279. 2) Firth, D. (1993). Biometrika 80, 27-38. 3) Miyata, T. and Yanagimoto, T. (2024). arXiv:2409.19673 4) Ogura, T. and Yanagimoto, T. (2016). Stat. in Med., 35, 2455-2466. 5) Yanagimoto, T. and Anraku, K. (1989). Ann. Inst. Stat. Math., 41, 269-278. 6) Yanagimoto, T. and Ohnishi, T. (2011). J. Stat. Plann. Inf., 41, 1990-2000.

Sparse estimators for multivariate integer-valued autoregressive models with applications to inference for Hawkes processes

Kou Fujimori [1], Hiroshi Shiraishi[2], Junichi Hirukawa[3], and Konstantinos Fokianos[4]

[1]Department of Economics, Shinshu University, kfujimori@shinshu-u.ac.jp

[2]Department of Mathematics, Keio University, shiraishi@math.keio.ac.jp

[3]Department of Data Science, Nanzan University, hirukawa@nanzan-u.ac.jp

[4]Department of Mathematics and Statistics, University of Cyprus, fokianos@ucy.ac.cy

September 14, 2025

1 Introduction

Univariate stationary Hawkes processes are known to be approximated by integer-valued autoregressive (INAR) models with infinite order; see e.g., [2]. For multivariate cases, the similar approximation are widely used. To estimate the self-excitement structures for multivariate Hawkes processes, the conditional least squares estimation for INAR models which approximates Hawkes processes are widely used; see e.g., [1], [3]. However, a rigorous proof of the discrete approximation has not yet been established in the multidimensional case. Furthermore, sparse estimators for high-dimensional settings have not been sufficiently investigated in estimating the coefficient matrix of an integer-valued time series approximating a multidimensional Hawkes process. In this talk, we discussed the discrete approximation of multidimensional Hawkes processes via integer-valued autoregressive models, as well as the sparse estimators for high-dimensional INAR models.

2 Main results

Let $\boldsymbol{N}$ be a d-dimensional stationary Hawkes process with the following intensity function:

$$\boldsymbol{\lambda}(t|\boldsymbol{N}) := \boldsymbol{\eta} + \int_{\mathbb{R}} \boldsymbol{H}(t-s)\boldsymbol{N}(ds), \quad t \in \mathbb{R},$$

where $\boldsymbol{\eta} \in \mathbb{R}^d$, $\boldsymbol{H}(t) \in \mathbb{R}^{d\times d}, t \in \mathbb{R}$ have non-negative components. For the Hawkes process $\boldsymbol{N}$ and given $\Delta > 0$, we define $\boldsymbol{A}_k^{(\Delta)} = \boldsymbol{H}(k\Delta), k \geq 1$, $\boldsymbol{\eta}^{(\Delta)} = \Delta\boldsymbol{\eta}$ and consider the following INAR (∞) model.

$$\boldsymbol{X}_t^{(\Delta)} = \sum_{k=1}^{\infty} \boldsymbol{A}_k^{(\Delta)} \circ \boldsymbol{X}_{t-j} + \boldsymbol{\epsilon}_t^{(\Delta)}, \quad t \in \mathbb{Z}, \tag{1}$$

where $\circ$ is the multivariate Poisson reproduction operator, $\{\boldsymbol{\epsilon}_t^{(\Delta)}\}$ is a d-dimensional i.i.d. random sequence such that for every $i = 1, \ldots, d$, $\epsilon_{t,i}^{(\Delta)} \sim Poisson(\Delta\eta_i)$. For multivariate reproduction operator, see e.g., [4]. Then, we define the following process $\boldsymbol{N}^{(\Delta)}$

$$\boldsymbol{N}^{(\Delta)}(A) = \sum_{t:t\Delta\in A} \boldsymbol{X}_t^{(\Delta)}, \quad A \in \mathcal{B},$$

where $\mathcal{B} \subset \mathbb{R}$ is the family of Borel sets. Then, under appropriate conditions, it holds that

$$\boldsymbol{N}^{(\Delta)} \to^w \boldsymbol{N}, \quad \Delta \to 0,$$

where $\to^w$ denoted the weak convergence.

When the non-negative valued function $\boldsymbol{H}(\cdot)$ is integrable, for every $\epsilon_H > 0$, there exists $\tau_H > 0$ such that $\|\boldsymbol{H}(t)\|_\infty < \epsilon_H, t > \tau_H$. Therefore, if $\Delta > 0$ is sufficiently small, the INAR (∞) defined in (1) can be approximated by the following INAR with finite order.

$$\boldsymbol{X}_t^{(\Delta)} = \sum_{k=1}^{p_\Delta} \boldsymbol{A}_k^{(\Delta)} \boldsymbol{X}_{t-k}^{(\Delta)} + \boldsymbol{\epsilon}_t^{(\Delta)}, \quad p_\Delta = \lceil \tau_H/\Delta \rceil.$$

This talk addressed sparse estimation of the coefficient matrix in finite-order INAR models and its application to the inference of Hawkes processes.

References

[1] P. Embrechts and M. Kirchner. Hawkes graphs. *Theory of Probability & Its Applications*, 62(1):132–156, 2018.

[2] M. Kirchner. Hawkes and INAR(∞) processes. *Stochastic Processes and their Applications*, 126(8):2494–2525, 2016.

[3] M. Kirchner. An estimation procedure for the Hawkes process. *Quantitative Finance*, 17(4):571–595, 2017.

[4] Alain Latour. The multivariate GINAR(p) process. *Advances in Applied Probability*, 29(1):228–248, 1997.

Flexible modeling of $\mathbb{Z}$-valued time series

Abdelhakim Aknouche (Qassim University)
Yuichi Goto (Kyushu University)[*1]
Christian Francq (ENSAE-CREST)

1. 導入

整数値を取る時系列データは, さまざまな分野において観測される. 例えば, サッカーにおける得失点数や顧客数や取引件数の増減などが挙げられる. また, 非負整数値時系列データが非定常の場合, 階差を取ることで定常データに変換することが考えられる. このとき, 差分を取った結果として, 正負両方向に値を持つ整数値の時系列データが生成される.

このような整数値時系列データをモデル化するために, 非負整数値時系列モデルを基にした拡張が試みられてきた. 代表的な非負整数値時系列モデルには, INGARCH (integer-valued generalized autoregressive conditional heteroscedasticity) モデルや INAR (integer-valued autoregressive) モデルがある. これらのモデルはそれぞれの構造を活かす形で整数値モデルへ拡張されてきた.

INGARCH モデルに基づく方法としては, 従来のポアソン分布に限定せず, スケラム分布をはじめとする整数値分布を条件付き分布として用いる手法が考案されている. また, INGARCH 過程で生成される非負整数に対し, $-1, 1$ の値を取るベルヌーイ分布を掛け合わせることにより, 符号を持つ整数値データを生成する枠組みも提案されている. 一方で, INAR モデルに基づく方法としては, 符号付き間引き演算子を導入し, 従来の非負整数モデルを整数全体へ拡張する試みがなされてきた. ただし, INAR に基づく方法は尤度が複雑になり, 推定や解析が困難になるという問題を抱えている.

従って, 解析的に取り扱いやすい INGARCH モデルの拡張を本研究の出発点とした. 既存の INGARCH モデルに基づく拡張手法は, 限られたクラスにしか適用できないという制約がある. 本講演では, 柔軟な整数値時系列モデルを提案し, その統計的性質を紹介した.

2. 提案モデル

そこで我々は, 次の混合差分型 (mixed difference) INGARCH モデルを提案した.

$$
\begin{aligned}
&Z_t := B_t X_t - (1 - B_t) Y_t,\\
&X_t | Z_{t-1}, Z_{t-2}, \ldots \sim \mathrm{Pois}(\lambda_{t,X}), \quad \lambda_{t,X} := \omega_X + \alpha_X |Z_{t-1}| + \beta_X \lambda_{t-1,X},\\
&Y_t | Z_{t-1}, Z_{t-2}, \ldots \sim \mathrm{ShiftedPois}(\lambda_{t,Y}), \quad \lambda_{t,Y} := \omega_Y + \alpha_Y |Z_{t-1}| + \beta_Y \lambda_{t-1,Y},\\
&B_t | Z_{t-1}, Z_{t-2}, \ldots \sim \mathrm{Ber}(\lambda_{t,B}), \quad \lambda_{t,B} := \omega_B + \alpha_B B_{t-1} + \beta_B \lambda_{t-1,B}.
\end{aligned}
$$

ここで, $\mathrm{Pois}(\lambda)$ は平均 $\lambda > 0$ のポアソン分布, $\mathrm{ShiftPois}(\lambda)$ はサポート $\{1, 2, \ldots\}$ である平均 λ の shifted ポアソン分布, $\mathrm{Ber}(\pi)$ は平均 $\pi \in (0, 1)$ のベルヌーイ分布である. また, Z_t の符号から B_t の値が計算できることに注意する. パラメータは, $\omega_1 > 0$, $\alpha_X, \alpha_Y \geq$

本研究は JSPS 科研費 JP23K16851, 日本数学会在外研究奨励フェローの助成を受けたものである.
[*1] e-mail: `yuichi.goto@math.kyushu-u.ac.jp`

$0,\ \beta_X,\ \beta_Y \geq 0,\ \omega_B > 0,\ \alpha_B > 0,\ \beta_B \geq 0,\ \omega_B + \alpha_B + \beta_B < 1$ を満たす．さらに, ω_2 に関しては, $\omega_2 > 1$ または $0 < 1 - \sum_{j=1}^{p} \beta_{2j} < \omega_2$ を満たす.

このモデルは, B_t が Z_t の符号を制御しており, X_t は非負整数値の構造を, Y_t は負整数値の構造をモデリングしている．上のモデルでは Poisson 分布や Shifted Poisson 分布を用いているが, 例えば Negative Binomial (NB) 分布や Shifted NB 分布など, さまざまな分布の組み合わせを考えることが可能である．本モデルの特色は, 二峰性分布のモデリングが可能である点や, 過去との相関を考慮した符号の遷移確率を導入できる点にある.

2.1. 提案モデルの性質

講演では，提案モデルに従う過程 Z_t の定常エルゴード性および β-mixing 性が成り立つための十分条件を与えた．まず，π_t に対して

$$\pi_t \leq \pi_1^+ \text{ a.s.}, \qquad 1 - \pi_t \leq \pi_0^+ \text{ a.s.}$$

を満たす定数 $\pi_1^+ \in (0,1)$ および $\pi_0^+ \in (1 - \pi_1^+, 1)$ が存在すると仮定する．さらに，X_t と Y_t は *stochastic-equal-mean order property* を満たすものとする．次の行列

$$\boldsymbol{A} = \begin{pmatrix} \alpha_X \pi_1^+ + \beta_X & \alpha_X \pi_0^+ \\ \alpha_Y \pi_1^+ & \alpha_Y \pi_0^+ + \beta_Y \end{pmatrix}$$

のスペクトル半径が 1 未満であれば，提案モデルに従う定常エルゴードな過程 Z_t が存在する．また，β-mixing 係数について，ある定数 $K > 0$ と $\varrho \in (0,1)$ が存在して

$$\beta_Y(h) \leq K \varrho^h, \qquad h \geq 0$$

が成り立つ.

2.2. 未知母数推定法と漸近的性質

次に, 未知パラメータの推定方法として，次で定義される mixed Poisson 擬似最尤法を提案した．$s = B, X, Y$ に対し $\boldsymbol{\psi}_s = (\omega_s, \alpha_s, \beta_s)^\top$ とおく．$\boldsymbol{\theta} = \left(\boldsymbol{\psi}_B^\top, \boldsymbol{\psi}_X^\top, \boldsymbol{\psi}_Y^\top\right)^\top$ に対して,

$$\widehat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \widetilde{\ell}_n(\boldsymbol{\theta}), \qquad \widetilde{L}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^{n} \widetilde{\ell}_t(\boldsymbol{\theta}).$$

ここで，対数擬似尤度関数は次式で与えられる:

$$\begin{aligned} \widetilde{\ell}_t(\boldsymbol{\theta}) = &\left\{ \log\left(\widetilde{\pi}_t(\boldsymbol{\psi}_B)\right) - \widetilde{\lambda}_{1t}(\boldsymbol{\psi}_X) + Z_t \log\left(\widetilde{\lambda}_{1t}(\boldsymbol{\psi}_X)\right) \right\} \mathbb{1}_{\{Z_t \geq 0\}} \\ &+ \left\{ \log\left(1 - \widetilde{\pi}_t(\boldsymbol{\psi}_B)\right) - \widetilde{\lambda}_{2t}(\boldsymbol{\psi}_Y) - \left(Z_t + 1\right) \log\left(\widetilde{\lambda}_{2t}(\boldsymbol{\psi}_Y) - 1\right) \right\} \mathbb{1}_{\{Z_t < 0\}}. \end{aligned}$$

また，固定された初期値 $\widetilde{\lambda}_{0,X}, \ldots, \widetilde{\lambda}_{1-p,X}, \widetilde{\lambda}_{0,Y}, \ldots, \widetilde{\lambda}_{1-p,Y}, \widetilde{\lambda}_{0,B}, \widetilde{Z}_0, \ldots, \widetilde{Z}_{1-q}, \widetilde{B}_0$ に対し，データから計算可能な近似量を次で定義する．$s = X, Y$ に対して,

$$\widetilde{\lambda}_{t,s}(\boldsymbol{\psi}_s) = \omega_s + \alpha_s |Y_{t-1}| + \beta_s \widetilde{\lambda}_{t-1,s}(\boldsymbol{\psi}_s), \quad \widetilde{\lambda}_{t,B}(\boldsymbol{\psi}_B) = \omega_B + \alpha_B B_{t-1} + \beta_B \widetilde{\lambda}_{t-1,B}(\boldsymbol{\psi}_B).$$

提案推定量に対して，正則性条件の下で推定値 $\widehat{\boldsymbol{\theta}}_n$ が真のパラメータ値 $\boldsymbol{\theta}_0$ に概収束することを示した．さらに，$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$ が漸近的に正規分布に従うことを明らかにした.

Self-normalized partial sums of heavy-tailed time series

Muneya Matsui

Abstract

We study the joint limit behavior of sums, maxima and ℓ^p-type moduli for samples taken from an $\mathbb{R}^d$-valued regularly varying stationary sequence with infinite variance. As a consequence, we can determine the distributional limits for ratios of sums and maxima, studentized sums, and other self-normalized quantities in terms of hybrid characteristic-distribution functions and Laplace transforms. These transforms enable one to calculate moments of the limits and to characterize the differences between the iid and stationary cases in terms of indices which describe effects of extremal clustering on functionals acting on the dependent sequence.

Targeting processes. The stationary sequence $(\mathbf{X}_t)$ is *regularly varying with (tail) index* $\alpha > 0$ if there exist a Pareto(α)-distributed random variable Y, i.e., $\mathbb{P}(Y > y) = y^{-\alpha}$, $y > 1$, and an $\mathbb{R}^d$-valued sequence $(\mathbf{\Theta}_t)$ which is independent of Y such that for every $h \geq 0$,

$$\mathbb{P}\big(x^{-1}(\mathbf{X}_{-h},\ldots,\mathbf{X}_h) \in \cdot \,\big|\, |\mathbf{X}_0| > x\big) \stackrel{d}{\to} \mathbb{P}\big(Y\,(\mathbf{\Theta}_{-h},\ldots,\mathbf{\Theta}_h) \in \cdot\big)\,, \qquad x \to \infty\,.$$

We prove the joint limit behavior of sums, maxima and ℓ^p-type moduli for samples taken from this sequence.

Out tool: hybrid characteristic-distribution function. Following Chow and Teugels [1], we will exploit the idea of *hybrid characteristic-distribution function.* We call a combination of a characteristic function (for sums) and a distribution function (for maxima) *hybrid characteristic-distribution function*:

$$\Psi_n(\mathbf{u},x) \;:=\; \mathbb{E}\big[\exp\big(i\,a_n^{-1}\mathbf{u}^\top\mathbf{S}_n\big)\,\mathbf{1}\big(a_n^{-1}M_n^{|\mathbf{X}|} \leq x\big)\big]\,, \qquad \mathbf{u}\in\mathbb{R}^d\,, \qquad x>0\,.$$

The hybrid characteristic-distribution function of a pair $(\mathbf{Y},U)$ of a random vector $\mathbf{Y}$ and a non-negative random variable U determines the distribution of $(\mathbf{Y},U)$. Moreover, $(\mathbf{Y}_n,U_n) \stackrel{d}{\to} (\mathbf{Y},U)$, $n\to\infty$, if and only if the hybrid characteristic-distribution functions of $(\mathbf{Y}_n,U_n)$ converge pointwise to the corresponding hybrid characteristic-distribution function of $(\mathbf{Y},U)$. We also use this function to describe the joint limit distribution.

Conditions: mixing condition, anti-clustering condition We will use the following mixing condition and anti-clustering condition. **Mixing condition**: for some integer sequences $r_n\to\infty$, $k_n=[n/r_n]\to\infty$, for every $x>0$, $\mathbf{u}\in\mathbb{R}^d$,

$$\Psi_n(\mathbf{u},x) \;=\; \Big(\mathbb{E}\big[\exp\big(i\,a_n^{-1}\mathbf{u}^\top\mathbf{S}_{r_n}\big)\,\mathbf{1}\big(a_n^{-1}M_{r_n}^{|\mathbf{X}|}\leq x\big)\big]\Big)^{k_n} + o(1)\,, \qquad n\to\infty\,. \tag{0.1}$$

Condition (0.1) ensures the asymptotic independence of k_n maxima and sums over disjoint blocks of length r_n. It follows from strong mixing properties of $(\mathbf{X}_n)$ or by coupling arguments.

An $\mathbb{R}^d$-valued stationary regularly varying sequence $(\mathbf{X}_t)$ satisfies the **anti-clustering condition** if for some $r_n\to\infty$ such that $k_n=[n/r_n]\to\infty$,

$$\lim_{l\to\infty}\limsup_{n\to\infty} n\sum_{j=l}^{r_n}\mathbb{E}\big[(|a_n^{-1}\mathbf{X}_j|\wedge x)\,(|a_n^{-1}\mathbf{X}_0|\wedge x)\big]=0\,, \qquad x=1\,. \tag{0.2}$$

Further definitions necessary to describe our result. Under (0.1) and (0.2) we have the property that $\|\mathbf{\Theta}\|_\alpha^\alpha = \sum_{t\in\mathbb{Z}} |\mathbf{\Theta}_t|^\alpha < \infty$ a.s. Then one can define the *spectral cluster process* $\mathbf{Q} = \mathbf{\Theta}/\|\mathbf{\Theta}\|_\alpha$. Defining $\|\cdot\|_p$, $p \in [\alpha,\infty]$, such that $\|\mathbf{x}\|_p^p = \sum_{t\in\mathbb{Z}} |\mathbf{x}_t|^p$ for $\mathbf{x} \in \ell^\alpha(\mathbb{R}^d)$, we introduce another spectral cluster process $\mathbf{Q}^{(p)}$ via the change of measure

$$\mathbb{P}(\mathbf{Q}^{(p)} \in \cdot) \;=\; (\mathbb{E}[\|\mathbf{Q}\|_p^\alpha])^{-1}\mathbb{E}\Big[\|\mathbf{Q}\|_p^\alpha \,\mathbf{1}\Big(\frac{\mathbf{Q}}{\|\mathbf{Q}\|_p} \in \cdot\Big)\Big]\,. \tag{0.3}$$

By construction $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s. and, beside $\mathbf{Q} = \mathbf{Q}^{(\alpha)}$, we also use the notation $\widetilde{\mathbf{Q}} = \mathbf{Q}^{(\infty)}$ and $\widehat{\mathbf{Q}} = \mathbf{Q}^{(2)}$ which are well-defined for $\alpha \in (0,2)$. The extremal index of $(|\mathbf{X}_t|)$ can be expressed as

$$\theta_{|\mathbf{X}|} = \mathbb{P}(Y \max_{t\in\mathbb{Z}} |\mathbf{Q}_t| > 1) = \mathbb{E}\big[\|\mathbf{Q}\|_\infty^\alpha\big] = \mathbb{E}\big[\|\widetilde{\mathbf{Q}}\|_\alpha^{-\alpha}\big]\,. \tag{0.4}$$

We also use the *spectral cluster process* to describe the joint limit distribution.

Main result. We have the following joint convergence of normalized maxima and sums.

Theorem 0.1. *Consider an $\mathbb{R}^d$-valued regularly varying stationary process $(\mathbf{X}_t)$ with index $\alpha \in (0,2)\backslash\{1\}$. If $\alpha \in (1,2)$ we also assume that $\mathbb{E}[\mathbf{X}] = 0$. Choose the normalizing constants (a_n) such that $n\,\mathbb{P}(|\mathbf{X}| > a_n) \to 1$ as $n \to \infty$. Assume the mixing condition* (0.1) *and the anti-clustering condition* (0.2) *for the same integer sequences $r_n \to \infty$, $k_n = [n/r_n] \to \infty$ as $n \to \infty$. Then*

$$a_n^{-1}(M_n^{|\mathbf{X}|}, \mathbf{S}_n) \xrightarrow{d} (\eta_\alpha, \boldsymbol{\xi}_\alpha)\,, \qquad n \to \infty\,,$$

where η_α is Fréchet-distributed with a positive extremal index $\theta_{|\mathbf{X}|}$ and $\boldsymbol{\xi}_\alpha$ is α-stable with characteristic function

$$\varphi_{\boldsymbol{\xi}_\alpha}(\mathbf{u}) = \exp\big(-c_\alpha\,\sigma^\alpha(\mathbf{u})\big(1 - i\,\beta(\mathbf{u})\tan(\alpha\,\pi/2)\big)\big)\,, \qquad \mathbf{u} \in \mathbb{R}^d\,, \tag{0.5}$$

constant $c_\alpha = \Gamma(2-\alpha)\,\cos(\alpha\,\pi/2)\,/(1-\alpha)$, and parameter functions, for $\widetilde{\mathbf{u}} = (\mathbf{u}/|\mathbf{u}|)\,\mathbf{1}(\mathbf{u} \neq \mathbf{0})$,

$$\beta(\mathbf{u}) = \frac{\mathbb{E}\big[(\widetilde{\mathbf{u}}^\top \sum_{t\in\mathbb{Z}} \mathbf{Q}_t)_+^\alpha - (\widetilde{\mathbf{u}}^\top \sum_{t\in\mathbb{Z}} \mathbf{Q}_t)_-^\alpha\big]}{\mathbb{E}\big[|\widetilde{\mathbf{u}}^\top \sum_{t\in\mathbb{Z}} \mathbf{Q}_t|^\alpha\big]}\,, \quad \sigma^\alpha(\mathbf{u}) = |\mathbf{u}|^\alpha\,\mathbb{E}\Big[\Big|\widetilde{\mathbf{u}}^\top \sum_{t\in\mathbb{Z}} \mathbf{Q}_t\Big|^\alpha\Big]\,. \tag{0.6}$$

Moreover, the joint distribution of $(\eta_\alpha, \boldsymbol{\xi}_\alpha)$ is characterized by the hybrid characteristic-distribution function: for $\mathbf{u} \in \mathbb{R}^d$ and $x > 0$,

$$\begin{aligned} &\mathbb{E}\big[\mathrm{e}^{\,i\,\mathbf{u}^\top \boldsymbol{\xi}_\alpha}\mathbf{1}(\eta_\alpha \le x)\big] \\ &= \varphi_{\boldsymbol{\xi}_\alpha}(\mathbf{u})\,\exp\Big(-\int_0^\infty \mathbb{E}\Big[\mathrm{e}^{\,i\,y\,\mathbf{u}^\top \sum_{t=-\infty}^\infty \mathbf{Q}_t}\,\mathbf{1}\Big(y\max_{t\in\mathbb{Z}} |\mathbf{Q}_t| > x\Big)\Big]\,d(-y^{-\alpha})\Big) \\ &= \varphi_{\boldsymbol{\xi}_\alpha}(\mathbf{u})\,\Phi_\alpha^{\theta_{|\mathbf{X}|}}(x)\,\exp\Big(-\theta_{|\mathbf{X}|}\int_x^\infty \mathbb{E}\big[\mathrm{e}^{\,i\,y\,\mathbf{u}^\top \sum_{t=-\infty}^\infty \widetilde{\mathbf{Q}}_t} - 1\big]\,d(-y^{-\alpha})\Big)\,, \end{aligned} \tag{0.7}$$

where $\mathbf{Q} = \mathbf{\Theta}/\|\mathbf{\Theta}\|_\alpha$ and $\widetilde{\mathbf{Q}} = \mathbf{Q}^{(\infty)}$ is defined in (0.3) *and the extremal index $\theta_{|\mathbf{X}|}$ is given in* (0.4).

Based on Theorem 0.1 we determine the distributional limits for ratios of sums and maxima, studentized sums, and other self-normalized quantities in terms of hybrid characteristic-distribution functions and Laplace transforms.

References

[1] Chow, T. L. and Teugels, J. L. (1978) The sum and the maximum of iid random variables. In: *Proceedings of the 2nd Prague Symposium on Asymptotic Statistics*, pp. 81–92.

[2] Matsui, M., Mikosch, T. and Wintenberger, O. (2025) Self-normalized partial sums of heavy-tailed time series. *Stoch. Proc. Appl.* **190**, 104729.

[3] Mikosch, T. and Wintenberger, O. (2024) *Extreme Value Theory for Time Series. The Heavy-Tailed Case.* Springer. New York.

Higher-order investigation of general time series divergences

Anna Clara Monti[(*)], Masanobu Taniguchi[(**)] and Yujie Xue[(***)]

[(*)] *DEMM Department, University of Sannio, Benevento, Italy.*
[(**)] *Faculty of Science and Engineering, Waseda University, Tokyo, Japan.*
[(***)] *The Institute of Statistical Mathematics, Tokyo, Japan.*

1 Report

Suppose that $g = g(\lambda)$ is the spectral density of a stationary process and let $f_\theta = f_\theta(\lambda)$, $\theta \in \Theta$, be a fitted spectral model for $g(\lambda)$. A minimum contrast estimator $\hat{\theta}_n$ of θ is defined by $\hat{\theta}_n = \arg\min_{\theta\in\Theta} D(f_\theta, \hat{g}_n)$ where $D(f_\theta, g)$ is a divergence between f_θ and g, and $\hat{g}_n$ is a nonparametric spectral density estimator based on n observations. [6] showed that $\hat{\theta}_n$ is asymptotically efficient if $g(\lambda) = f_\theta(\lambda)$. Because there are infinitely many candidates of $D(f_\theta, g)$, infinitely many efficient estimators $\hat{\theta}_n$ can be constructed. In view of this, [7] discussed the second-order asymptotic efficiency of $\hat{\theta}_n$, based on the second-order Edgeworth expansion. These authors show that the bias-adjusted version of $\hat{\theta}_n$ is not second-order asymptotically efficient. This is a sharp contrast with regular parametric estimation, where it is known that if an estimator is first-order asymptotically efficient, then it is automatically second-order asymptotically efficient, after a suitable bias adjustment (e.g. [1]).

In this paper we investigate the first- and second-order robustness of $D(f_\theta, g)$ by use of the first-order influence function $IF_1(D)$ and the second-order influence function $IF_2(D)$, where $D = D(f_\theta, g)$. The first-order case is investigated proving that all $IF_1(D) = 0$, i.e. all the divergences are first-order robust. In the analysis of second-order robustness we consider (a) D_{K_1} which is a Whittle divergence, (b) D_{K_2} a log-squares divergence, (c) D_{K_3} an α-power divergence, (d) D_{K_4} an α-entropy divergence and (e) the Hellinger divergence introduced by [8]. The paper shows that (c) and (d) with $\alpha = 1/3$ have better second-order robustness properties than (a), (b) and (e), which is an unexpected result. [4] also discuss higher-order infinitesimal robustness for "regular" statistical estimators. However the current paper deals with "semiparametric estimation", hence methods and results are different.

Also [7] and [8] showed that the D_K's lead to the first-order efficiency, and that (a), (c) with $\alpha = 1/3$ and (e) lead to the second-order efficiency, but the other D_K's do not.

References

[1] Ghosh J. K. (1994) *Higher Order Asymptotics.* Institute of Mathematical Statistics. California.

[2] Hampel F.R. (1974). The Influence Curve and Its Role in Robust Estimation. *J. Amer. Statist. Assoc.* **69**, 383–393.

[3] Hannan E. J. (2008) *Multiple Time Series.* J. Wiley.

[4] La Vecchia, D., Ronchetti, E., and Trojani, F. (2012). Higher-Order Infinitesimal Robustness. *J. Amer. Statist. Assoc.* **107**, 1546–1557.

[5] R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`.

[6] Taniguchi M. (1987). Minimum contrast estimations for spectral densities of stationary processes. *J. Roy. Statist. Soc.* B **49**, 315–325.

[7] Taniguchi, M., van Garderen, K. J., and Puri, M. L. (2003). Higher Order Asymptotic Theory for Minimum Contrast Estimators of Spectral Parameters of Stationary Processes. *Econometric Theory*, **19**, 984–1007.

[8] Taniguchi M and Xue Y. (2023). Hellinger Distance Estimation for non-regular spectra. Theory Probab. Appl., **69**, 565–570.

NEURAL TANGENT KERNEL IN IMPLIED VOLATILITY FORECASTING: A NONLINEAR FUNCTIONAL AUTOREGRESSION APPROACH

HANNAH L. H. LAI

We denote by $\mathcal{H} = L^2(\mathcal{I})$ the Hilbert space consisting of all square-integrable surfaces defined on a compact set $\mathcal{I} \subset \mathbb{R}^q$ and equipped with the inner product $\langle f, g\rangle_{\mathcal{H}} = \int_{\mathcal{I}} f(u)g(u)\, du$, for any $f, g \in L^2(\mathcal{I})$. Define the squared L^2 norm of a function by $\|f\|_{\mathcal{H}} = \langle f, f\rangle_{\mathcal{H}}$. Let $\{Y_i\}_{i=1}^n$ be a series of n random surfaces that take values on $\mathcal{H}_Y = L^2(\mathcal{I}_Y)$. Associated with each Y_i, there is a regressor surface $X_i \in \mathcal{H}_X = L^2(\mathcal{I}_X)$. We consider functions with finite second moment, i.e., $\mathbb{E}[||Y_i||^2_{\mathcal{H}_Y}] < \infty$ and $\mathbb{E}[\|X_i\|^2_{\mathcal{H}_X}] < \infty$. For simplicity, we assume that Y_i and X_i are centered functions, i.e., $\mu_X(v) = \mathbb{E}[X_i(v)] = 0, \forall v \in \mathcal{I}_X$ and $\mu_Y(u) = \mathbb{E}[Y_i(u)] = 0, \forall u \in \mathcal{I}_Y$. Let P_X and P_Y denote the distributions of X and Y, and $P_{Y|X} : \mathcal{H}_X \times \mathcal{H}_Y \to \mathbb{R}$ the conditional distribution of Y given X. If $L_2(P_X)$ represents the class of all measurable functions of X with $\mathbb{E}[f^2(X)] < \infty$ under P_X, then $L_2(P_Y)$ is similarly defined for Y. Our goal is to capture the potential nonlinear dependence between Y_i and X_i through a function $g : \mathcal{H}_X \to \mathcal{H}_Y$

$$Y_i = g(X_i) + \epsilon_i, \tag{0.1}$$

where ϵ_i is a noise function with $\mathbb{E}[\epsilon_i(u)] = 0, \forall u \in \mathcal{I}_Y$ and $\mathbb{E}[\|\epsilon_i\|^2_{\mathcal{H}_Y}] < \infty$. In our study, X_i is a vector of lagged surfaces $Y_{i-1}, Y_{i-2}, \ldots$ or their linear combination. Hence, the model (0.1) is a nonlinear functional autoregression model (NFAR).

We project Y_i onto a set of orthonormal basis functions $\varphi = (\varphi_1, \varphi_2 \ldots)^T$ with $\varphi_j \in \mathcal{H}_Y$

$$Y_i = \sum_{j=1}^{\infty} y_{ij}\varphi_j, \quad \text{with } y_{ij} = \langle Y_i, \varphi_j\rangle_{\mathcal{H}_Y}, \tag{0.2}$$

with $\boldsymbol{y}_i = (y_{i1}, y_{i2}, ...)^T \in \mathcal{H}_{\boldsymbol{y}} \subseteq \mathbb{R}^\infty$ the projection coefficients of Y_i onto the basis functions φ, satisfying $\mathbb{E}[y_{ij}y_{rv}] = 0$ for $j \neq v$, $j, v \in \mathbb{N}_+$ and any $i, r \in \{1, \ldots, n\}$. Similarly, we project X_i onto a sequence of orthogonal basis functions $\psi = (\psi_{1,}, \psi_2, ...)^T$ with $\psi_j \in \mathcal{H}_X$

$$X_i = \sum_{j=1}^{\infty} x_{ij}\psi_j, \quad \text{with } x_{ij} = \langle X_i, \psi_j\rangle_{\mathcal{H}_X}, \tag{0.3}$$

with $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ...)^T \in \mathcal{H}_{\boldsymbol{x}} \subseteq \mathbb{R}^\infty$ the projection coefficients of X_i onto the basis functions ψ, satisfying $\mathbb{E}[x_{ij}x_{rv}] = 0$ for $j \neq v$, $j, v \in \mathbb{N}_+$ and any $i, r \in \{1, \ldots, n\}$. Transitioning from functions to vectors, we define $f : \mathcal{H}_x \to \mathcal{H}_y$

$$\boldsymbol{y}_i = f(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i, \tag{0.4}$$

where $\boldsymbol{\epsilon}_i$ is a noise vector with $\mathbb{E}[\boldsymbol{\epsilon}_{ij}] = 0$ and $\mathbb{E}[\|\boldsymbol{\epsilon}_i\|^2] < \infty$. Although vectors offer a more compact representation of functions, they still exist within an infinite-dimensional framework unless additional restrictions are assumed to hold. This inherent complexity makes the empirical estimation of Equation (0.4) challenging when working with finite sample sizes. To address this issue, we employ classical sieve methods leading to finite-dimensional vector spaces.

To elucidate the nonlinear relation between X_i and Y_i in Equation (0.1), we introduce another Hilbert space of functions generated by a positive-definite kernel $K : \mathcal{H}_X \times \mathcal{H}_X \to \mathbb{R}$ defined on the inner product of $\mathcal{H}_X$ through a function $\rho : \mathbb{R}^3 \to \mathbb{R}^+$, such that

$$K(X_i, X_j) = \rho(\langle X_i, X_i\rangle_{\mathcal{H}_X}, \langle X_i, X_j\rangle_{\mathcal{H}_X}, \langle X_j, X_j\rangle_{\mathcal{H}_X}), \tag{0.5}$$

for any $X_i, X_j \in \mathcal{H}_X$. The function-on-function regression problem in Equation (0.1) can be reformulated as a functional kernel regression, in which the task is to find $B_0 \in \mathcal{B}(\mathcal{H}_Y, \mathfrak{M}_X)$ such that

$$B_0 = \underset{B \in \mathcal{B}(\mathcal{H}_Y, \mathfrak{M}_X)}{\arg\min} \mathbb{E}[\|Y_i - B^* K(., X_i)\|^2_{\mathcal{H}_Y}]. \tag{0.6}$$

We define a new kernel $k : \mathcal{H}_{\boldsymbol{x}} \times \mathcal{H}_{\boldsymbol{x}} \to \mathbb{R}$ such that for any $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{H}_{\boldsymbol{x}}$

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \rho(\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle, \langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle, \langle \boldsymbol{x}_j, \boldsymbol{x}_j\rangle). \tag{0.7}$$

Lemma 0.1 (Isomorphism between Reproducing Kernel Hilbert Spaces). *Under Equations* (0.3) *and* (0.7)*, it holds that*

$$\begin{aligned} k(\boldsymbol{x}_i, \boldsymbol{x}_j) &= \langle k(., \boldsymbol{x}_i), k(., \boldsymbol{x}_j)\rangle \\ &= \langle K(., X_i), K(., X_j)\rangle_{\mathfrak{M}_X} = K(X_i, X_j). \end{aligned} \tag{0.8}$$

Then the RKHS $\mathfrak{M}_X$ nested on $\mathcal{H}_X$ is isometrically isomorphic to the RKHS $\mathfrak{M}_{\boldsymbol{x}}$ nested on $\mathcal{H}_{\boldsymbol{x}}$.

Theorem 0.2 (Vector-to-vector regression). *Given the decomposition of Y_i in Equation* (0.2) *and X_i in Equations* (0.3)*, under some technical Assumptions and Lemma 0.1, for a positive definite kernel k defined by Equation* (0.7)*, if there is a covariance matrix $\Sigma_{\boldsymbol{xx}}$ of $k(., \boldsymbol{x})$ that is diagonal, then the function-to-function regression model in Equation* (0.6) *may be represented equivalently by*

$$\beta_0 = \underset{\beta \in \mathcal{B}(\mathcal{H}_{\boldsymbol{y}}, \mathfrak{M}_{\boldsymbol{x}})}{\arg\min} \mathbb{E}[\|\boldsymbol{y}_i - \beta^* k(., \boldsymbol{x}_i)\|^2], \tag{0.9}$$

with solution $\beta_0 = \Sigma^\dagger_{\boldsymbol{xx}}\Sigma_{\boldsymbol{xy}}$. This leads to

$$\begin{aligned} \mathbb{E}[\boldsymbol{y}_i | \boldsymbol{x}_i] &= \beta_0^* k(., \boldsymbol{x}_i) \\ &= \Sigma_{\boldsymbol{yx}}\Sigma^\dagger_{\boldsymbol{xx}} k(., \boldsymbol{x}_i) \\ &= \mathbb{E}[\{(\Sigma^\dagger_{\boldsymbol{xx}} k(., \boldsymbol{x}_i))(\boldsymbol{x})\}\boldsymbol{y}]. \end{aligned} \tag{0.10}$$

In our work, we utilize the Neural Tangent Kernel (NTK), a flexible kernel class that uses neural networks to capture complex nonlinear dependencies in data effectively. Our empirical analysis includes over 6 million European calls and put options from the S&P 500 Index, covering January 2009 to December 2021. The results confirm the superior forecasting accuracy of the fNTK across different time horizons. When applied to short delta-neutral straddle trading, the fNTK achieves a Sharpe ratio ranging from 1.30 to 1.83 on a weekly to monthly basis, translating to 90% to 675% relative improvement in portfolio returns compared to forecasts based on functional Random Walk model.

Email address: `hlhlai@u.nus.edu`

Debt Structure and Recovery Rates

Lei Zhou * Ying Chen †

September 17, 2025

Recovery rates are a critical determinant in credit risk modeling, bond pricing, and risk management, quantifying the amount creditors reclaim in the event of default. Empirical evidence reveals substantial heterogeneity in recovery rates; for example, Moody's (2023) reports an average ultimate recovery rate of 40% for senior unsecured bonds, yet with significant dispersion where 25% recovered less than 10% and 20% recovered more than 70%. This variation challenges the conventional practice of assuming constant recovery rates in credit pricing models. While factors like debt seniority and collateralization are recognized determinants, the role of the overall debt structure remains underexplored. This paper defines debt structure complexity as the hierarchical composition of a firm's obligations, including seniority layers, debt instruments, and covenant constraints, and examines its profound influence on recovery rates.

Our theoretical framework extends the structural credit risk model by Leland and Toft (1996). It incorporates finite-maturity debt, refinancing opportunities, endogenous default, and a layered capital structure composed of senior and junior bonds.

The firm's asset value, V_t, is modeled as a lognormal diffusion process under the risk-neutral probability measure:

$$\frac{dV_t}{V_t} = (r-q)dt + \sigma dz \quad (1)$$

where r is the risk-free interest rate, q is the payout ratio, σ is the firm's asset value volatility, and dz is a standard Brownian motion. The capital structure comprises equity (E), a senior bond (SB) with face value F_{SB} and coupon C_{SB}, and a junior bond (JB) with face value F_{JB} and coupon C_{JB}. The total levered firm value V_t^L is the sum of these components:

$$V_t^L = E(V_t,t) + B_{SB}(V_t,t) + B_{JB}(V_t,t) \quad (2)$$

where $B_{SB}(V,t)$ and $B_{JB}(V,t)$ are the market values of the senior and junior bonds, respectively. The model incorporates market frictions such as flotation costs for new debt issuance (γ), tax shield benefits from coupon payments (at rate α), and bankruptcy costs (ω) representing a proportion of the firm's asset value forfeited during liquidation.

Equity holders optimally choose the timing of default, τ_D, to maximize their value, which occurs when their value becomes non-positive. This is formally defined by the smooth-pasting

*Asian Institute of Digital Finance, National University of Singapore, 21 Lower Kent Ridge Rd, Singapore. E-mail: leizhou@nus.edu.sg.

†Department of Mathematics & Risk Management Institute & Center for Quantitative Finance & Asian Institute of Digital Finance, National University of Singapore, 21 Lower Kent Ridge Rd, Singapore. E-mail: matcheny@nus.edu.sg.

and value-matching conditions at the default boundary $V_B(t)$:

$$E(V_B(t),t)=0,\quad \frac{\partial E}{\partial V}(V_B(t),t)=0 \quad (3)$$

In bankruptcy, the firm's assets are liquidated. The remaining asset value, RV_{τ_D}, after deducting bankruptcy costs, is:

$$RV_{\tau_D}=(1-\omega)V_{\tau_D} \quad (4)$$

This value is distributed based on strict priority. Senior bondholders are repaid first, up to their face value:

$$SBRV_{\tau_D}=\min(F_{SB},RV_{\tau_D}) \quad (5)$$

Junior bondholders receive any residual value only if the senior claim is fully satisfied:

$$JBRV_{\tau_D}=\min(F_{JB},\max(0,RV_{\tau_D}-F_{SB})) \quad (6)$$

The recovery rates for senior and junior bonds are defined as the ratio of recovered value to face value:

$$RR_{SB}=\frac{SBRV_{\tau_D}}{F_{SB}},\quad RR_{JB}=\frac{JBRV_{\tau_D}}{F_{JB}} \quad (7)$$

This hierarchical structure implies that if $RV_{\tau_D}<F_{SB}$, junior bondholders receive nothing, illustrating the subordination of their claims. The optimal decision-making behavior of equity holders, considering default and refinancing, is governed by the Hamilton–Jacobi–Bellman (HJB) equation:

$$\max\left\{\frac{\partial E}{\partial t}+rV\frac{\partial E}{\partial V}+\frac{1}{2}\sigma^2V^2\frac{\partial^2 E}{\partial V^2}-rE+(1-\alpha)((r-q)V-C_{total}),E\right\}=0 \quad (8)$$

where $C_{total}=C_{SB}+C_{JB}$. The solution to this equation determines the endogenous default boundary and thus the recovery rates under various capital structures.

Simulations demonstrate that as the ratio of junior bonds increases (proportion of senior bonds decreases), the recovery rate of senior bonds increases in a stepwise upward pattern. This is attributed to a larger portion of the fixed liquidation value being available relative to the reduced outstanding senior debt. The recovery rate of junior bonds also exhibits a fluctuating but generally increasing trend, reflecting the expanding pool of junior bondholders eligible for residual liquidation value. While senior bond recovery rates consistently remain higher, these specific trends highlight significant nonlinearities.

This paper conclusively demonstrates that debt structure complexity significantly influences bond recovery rates, moving beyond conventional seniority-based explanations. Our combined theoretical and empirical approach establishes that recovery rates are dynamic and depend not only on the nominal seniority but also on the overall debt composition, leading to nonlinear recovery dynamics for both senior and junior bonds.

Cauchy 型の EM アルゴリズムとその加速法について

Toshihiro Abe[1] and Masahiro Kuroda[2]

[1]Hosei University, Faculty of Economics
[2]Okayama University of Science, Faculty of Management

The t distribution, which includes the heavy-tailed Cauchy distribution, is a useful model for data analysis including outliers. The likelihood equations of the t distribution are often solved by numerical methods. The Newton-Raphson method is a simple and fast approach, but it may not converge due to its dependence on initial values. Liu & Rubin (1995) used stochastic representations of the normal and gamma distributions to estimate the parameters of the multivariate t distribution, and compared the EM, ECM, and ECME. This study focuses on a simple expression of the EM algorithm, A simple expression of the EM algorithm is focused on in this study, while other options are available in literature.

At first, we consider Cauchy and its related distributions, with probability density functions (pdf's) given by:

$$f_C(x) = \frac{1}{\pi\sigma\left(1+\left(\frac{x-\mu}{\sigma}\right)^2\right)} = \frac{\sigma}{\pi\left(\sigma^2+(x-\mu)^2\right)}, \quad x \in (-\infty,\infty), \tag{1}$$

$$f_{logC}(x) = \frac{1}{x\pi\sigma\left(1+\left(\frac{\log x-\mu}{\sigma}\right)^2\right)} = \frac{\sigma}{x\pi\left(\sigma^2+(\log x-\mu)^2\right)}, \quad x \in (0,\infty), \tag{2}$$

where $\mu \in \mathbb{R}$, $\sigma \in (0,\infty)$. The log-Cauchy distribution is obtained by taking the exponential of a Cauchy random variable. It has infinite moments and is suitable for modeling extreme events or outliers. For example, it can be used to model the time from infection to virus onset. Our talk consists as follows. Among various these skew-t distributions, this study gives the simple EM algorithm for the Cauchy and its related distributions. For that purpose, a review of the skew-t distributions is given. Then, the main idea is constructed to borrow the strength of the exponential distribution structure and thus obtain explicit expressions of the EM algorithm for the distribution in a form of Cauchy-type. Based on the idea and the inclusion of a latent variable into the distribution function, explicit updated formulae for the multivariate skew-t distribution are derived. The formulae for the finite mixture and regression models are also derived in the section. Some comments on the initial values and stopping rules for illustrative examples are given. The first example of the EM algorithm for the multivariate skew-t distribution is shown for the CHF 1,000 bill data and the second one is given for real estate data that are skewed and have a heavy tail. The vector epsilon algorithm applied to our EM algorithm is given using the result in this talk. In our numerical experiments, all computations were executed using *Mathematica*, a symbolic mathematical computing package. Finally, conclusions are presented.

References

Liu, C., & Rubin, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. Statistica Sinica, **5**, 19–39.

シンプルなトーラス上分布の隠れマルコフモデルへの応用

Imoto Tomoaki
University of Shiozuoka

Abstract

A probability distribution on the torus, or toroidal distribution, is a bivariate distribution of two circular random variables. The applications of toroidal distribution can be seen in many fields. For example, analysis of the wind directions at two points, the directions of the consecutive movements of some animal, comparison of genome structures between paired bacteria, and the pair of dihedral angles between consecutive amino acids in a protein, and so on.

In this paper, we focus on the parameter estimation of the distribution by Abe et. al. [1] with the pdf

$$f(\theta_1, \theta_2) = 2G\left(\lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right) f_1(\theta_1 - \mu_1) f_2(\theta_2 - \mu_2), \tag{1}$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are the symmetric circular pdfs at 0 and $G(\cdot)$ is the cdf whose pdf is symmetric at 0. The marginal pdfs of the distribution (1) are easily shown to be $f_1(\theta_1 - \mu_1)$ and $f_2(\theta_2 - \mu_2)$, and the parameter λ controls the relation between two circular variables. Therefore, this construction is one of the methods of specifying marginals, but it does not need any complicated functions like cdf, and an additional normalizing constant. A more interesting property is that the Fisher information matrix is expressed by

$$\begin{pmatrix} \iota_{\lambda\lambda} & 0 & 0 \\ 0 & \iota_{\mu_1\mu_1} & \iota_{\mu_1\mu_2} \\ 0 & \iota_{\mu_1\mu_2} & \iota_{\mu_2\mu_2} \end{pmatrix}, \text{ where } \iota_{\alpha\beta} = \mathrm{E}\left[\frac{\partial^2 \log f}{\partial\alpha\partial\beta}\right],$$

and $\iota_{\mu_1\mu_2} = 0$ when $\lambda = 0$. This property will be useful when the test of independence between two variables is considered, since the usual asymptotic normality is satisfied for the maximum likelihood estimate (MLE) of λ.

Consider the trivariate distribution with the pdf

$$f(w, \theta_1, \theta_2) = 2\phi\left(w - \lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right) f_1(\theta_1 - \mu_1) f_2(\theta_2 - \mu_2)$$

for $(w, \theta_1, \theta_2) \in \mathbb{R}^+ \times [-\pi, pi) \times [-\pi, pi)$, where $\phi(\cdot)$ is the standard normal pdf. The marginal pdf of (θ_1, θ_2) is

$$\begin{aligned} f(\theta_1, \theta_2) &= \int_0^\infty 2\phi\left(w - \lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right) f_1(\theta_1 - \mu_1) f_2(\theta_2 - \mu_2) dw \\ &= 2\int_{-\lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)}^\infty \phi(w) dw f_1(\theta_1 - \mu_1) f_2(\theta_2 - \mu_2) \\ &= 2\{1 - \Phi\left(-\lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right)\} f_1(\theta_1 - \mu_1) f_2(\theta_2 - \mu_2) \\ &= 2\Phi\left(\lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right) f_1(\theta_1 - \mu_1) f_2(\theta_2 - \mu_2), \end{aligned}$$

corresponding to (1) with $G(x) = \Phi(x)$. Therefore, the conditional distribution of w given (θ_1, θ_2) is the truncated normal distribution with the pdf

$$f(w \mid \theta_1, \theta_2) = \frac{\phi\left(w - \lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right)}{\Phi\left(\lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right)}, \quad w > 0$$

with the expectation

$$\begin{aligned} \mathrm{E}[w \mid \theta_1, \theta_2] &= \lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2) + \frac{\phi\left(-\lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right)}{1 - \Phi\left(-\lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right)} \\ &= \lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2) + \frac{\phi\left(\lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right)}{\Phi\left(\lambda \sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2)\right)}. \end{aligned}$$

Using this property and considering w as a latent variable, we can construct the complete log-likelihood function for the observations $(\theta_{11}, \theta_{21}), (\theta_{12}, \theta_{22}), \ldots, (\theta_{1N}, \theta_{2N})$ as

$$\begin{aligned} \ell_c =& N \log 2 - \frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{n=1}^{N} w_n^2 + \lambda \sum_{n=1}^{N} w_n \sin(\theta_{1n} - \mu_1)\sin(\theta_{2n} - \mu_2) \\ &- \frac{\lambda^2}{2} \sum_{n=1}^{N} \sin^2(\theta_{1n} - \mu_1)\sin^2(\theta_{2n} - \mu_2) + \sum_{n=1}^{N} \log f_1(\theta_{1n} - \mu_1) + \sum_{n=1}^{N} \log f_2(\theta_{2n} - \mu_2). \end{aligned}$$

In this expression, changing w_n and w_n^2 with thier respective expectations

$$\widehat{w_n} = \lambda \sin(\theta_{1n} - \mu_1)\sin(\theta_{2n} - \mu_2) + \frac{\phi\left(\lambda \sin(\theta_{1n} - \mu_1)\sin(\theta_{2n} - \mu_2)\right)}{\Phi\left(\lambda \sin(\theta_{1n} - \mu_1)\sin(\theta_{2n} - \mu_2)\right)} \quad \text{and}$$

$$\widehat{w_n^2} = 1 + \lambda^2 \sin^2(\theta_{1n} - \mu_1)\sin^2(\theta_{2n} - \mu_2) + \lambda \sin(\theta_{1n} - \mu_1)\sin(\theta_{2n} - \mu_2)\frac{\phi\left(\lambda \sin(\theta_{1n} - \mu_1)\sin(\theta_{2n} - \mu_2)\right)}{\Phi\left(\lambda \sin(\theta_{1n} - \mu_1)\sin(\theta_{2n} - \mu_2)\right)},$$

we obtain the Q function for the EM algorithm.

Application of the EM algorithm to the hidden Markov model with the introduced distribution and fitting to the time series data about wind direction are shown in the talk.

Acknowledgement

This research was supported in part by JSPS KAKENHI Grant Number 24K06849.

References

[1] T. Abe, T. Imoto, T. Shiohama, and Y. Miyata. On some flexible models for circular, toroidal, and cylindrical data. In Ashis SenGupta and Barry C. Arnold, editors, *Directional Statistics for Innovative Applications: A Bicentennial Tribute to Florence Nightingale*, pages 229–243. Springer, 2022.

シンポジウム報告書（タイトル:円柱上の統計モデルを出力分布に持つ隠れマルコフモデルについて)

宮田　庸一 [*]　　塩濱　敬之 [†]　　阿部　俊弘 [‡]

1　はじめに

Θ は $-\pi$ 以上 π 未満の値を取る角度変数とし, X は非負値もしくは実数値をとる確率変数とする．組 (Θ, X) の確率分布は, シリンダー上の統計モデルと呼ばれている．このため, シリンダー値を取るデータとは, $[-\pi, \pi) \times \mathbb{R}_+$ もしくは $[-\pi, \pi) \times \mathbb{R}$ 上の値をとるデータのことを指している．実際, このようなデータは環境学, 生物学, スポーツなど様々な分野で見ることができ, 様々な研究者により解析がなされている．風向と二酸化硫黄 SO_2 の集中度に関しては, García-Portugués et al. (2014) により研究されている．またアドリア海の海流の方向とスピードに対しては, Lagona et al. (2015) により, Abe and Ley (2017) により提案されたシリンダー上の統計モデルをコンポーネントに持つ隠れマルコフモデルを通じて, 解析がなされている．Abe and Ley (2017) では以下の (X, Θ) の確率密度関数を提案した:

$$f_{AL}(\theta, x) = \frac{\alpha\beta^\alpha}{2\pi\cosh(\kappa)}\left(1 + \lambda\sin(\theta - \mu)\right)x^{\alpha-1}\exp\left\{-(\beta x)^\alpha\left(1 - \tanh(\kappa)\cos(\theta - \mu)\right)\right\}, \quad (1)$$

ただし $x \geq 0$, $\cosh(\kappa) = \{\exp(\kappa) + \exp(-\kappa)\}/2$, and $\tanh(\kappa) = \{\exp(\kappa) - \exp(-\kappa)\}/\{\exp(\kappa) + \exp(-\kappa)\}$ とする．またパラメーターの空間は $\alpha > 0$, $\beta > 0$, $\kappa > 0$, $-\pi \leq \mu < \pi$, $-1 \leq \lambda \leq 1$ となる．このモデルは, 正規化定数が非常にシンプルなものになっており, 取り扱いが容易である．(1) の密度関数においては, Θ の周辺分布が λ を歪みパラメータに持つ正弦関数摂動巻き込みコーシー分布 (Abe and Pewsey, 2011) になっているため, 単峰非対称なデータにも対応できる形になっている．しかしながら, 実際には, 正弦関数摂動巻き込みコーシー分布ではとらえきれない強い歪みを持つシリンダー上のデータが存在し, そのときの歪みパラメータに対する最尤推定量は, パラメーター空間の境界上の値を取ることが知られている．

[*] 高崎経済大学　経済学部
[†] 南山大学 理工学部・データサイエンス学科
[‡] 法政大学 経済学部

2　主な報告内容

この問題に対処するために, シンポジウムでは Miyata et al. (2024) により提案された以下の形の (Θ, X) の結合密度関数を提案した:

$$f^{(q)}_{WeiESSvM}(\theta,x)=\frac{\alpha\beta^{\alpha}}{\pi\cosh(\kappa)}G_q(\lambda\sin(\theta-\mu))x^{\alpha-1}\exp\left\{-(\beta x)^{\alpha}\left(1-\tanh(\kappa)\cos(\theta-\mu)\right)\right\}, \tag{2}$$

ただし $x\geq 0$, $\alpha>0$, $\beta>0$, $\kappa>0$, $-\pi\leq\mu<\pi$, $-1\leq\lambda\leq 1$ とし, $\boldsymbol{H}=\{(\mu,\kappa,\lambda,\alpha,\beta)|-\pi\leq\mu<\pi,\kappa>0,-1\leq\lambda\leq 1,\alpha>0,\beta>0\}$ をパラメーター空間とする. また, $[-1,1]$ 上の密度関数

$$g_q(x)=\frac{\Gamma(2(q+1))}{2^{2q+1}\Gamma(q+1)^2}(1-x^2)^q \qquad (-1\leq x\leq 1),$$

の分布関数を $G_q(x)=\int_{-1}^{x}g_q(t)dt$ とする. ただし $q\geq 0$ は事前に決めておくべき超パラメーターとし, ここではオーダーと呼ぶ. この分布の Θ および X の周辺分布, Θ を与えたときの X の条件付き分布, X を与えたときの Θ の条件付き分布は陽に表すことができる. さらに, 角度変数 Θ 側の観測データの分布が比較的強い歪みを示すときにも, オーダーを適切に選ぶことにより, よいデータフィッティングを与えることができる. この良さを示すために, 京都大学防災研究所潮岬風力実験所で計測された標本サイズ $T=426$ の風向, 風速データに対して, この分布を出力分布に持つ隠れマルコフモデルを最尤法によりフィットさせた. 従来の Lagona et al. (2015) のモデルでは, いくつかの出力分布における歪みパラメーター λ に対する最尤推定量の値がパラメーター空間の境界上の値を取るのに対して, 今回提案したモデルではこれらの最尤推定がパラメーター空間の内点におさまり, さらにモデルのフィットの良さを表す対数尤度が改善されることを報告した.

参考文献

Abe, T. and C. Ley (2017). A tractable, parsimonious and flexible model for cylindrical data, with applications. *Econom. Stat. 4*, 91–104.

Abe, T. and A. Pewsey (2011). Sine-skewed circular distributions. *Statist. Papers 52*(3), 683–707.

García-Portugués, E., A. M. Barros, R. M. Crujeiras, W. González-Manteiga, and J. Pereira (2014). A test for directional-linear independence, with applications to wildfire orientation and size. *Stochastic environmental research and risk assessment 28*(5), 1261–1275.

Lagona, F., M. Picone, and A. Maruotti (2015). A hidden markov model for the analysis of cylindrical time series. *Environmetrics 26*(8), 534–544.

Miyata, Y., T. Shiohama, and T. Abe (2024). Cylindrical models motivated through extended sine-skewed circular distributions. *Symmetry 16*(3).

（ハイパー）シリンダー上の分布構築に向けて

統計数理研究所 大学統計教員育成センター
清水 邦夫

シリンダー上の分布 (distribution on the cylinder, cylindrical distribution) とは，線形な (linear) 変数 X と円周上 (circular) もしくは角度の (angular) 変数 Θ の同時（結合）分布のことをいう．X もしくは Θ は連続型のこともあるし離散型のこともある．典型的な応用例としては,（風速，風向）,（波高，風向）,（動物の移動距離，角度）,（1 日の各時間の救急搬送件数，時間）がある．いくつかの線形変数といくつかの角度変数からなる多変量のシリンダー上の分布はトーラス上の分布 (distribution on the torus) を拡張しており，ハイパーシリンダー上の分布 (distribution on the hyper-cylinder) と呼ばれる．シンポジウムでは,（ハイパー）シリンダー上の分布の定義，分布生成のための方法，二変量線形と一変量角度の Wicksell–Kibble 型分布の構築（清水・Dou・Liu（2025, 統計関連学会連合大会）について紹介を行った．

（ハイパー）シリンダー上の分布の生成法：

（ハイパー）シリンダー上の分布を生成するためのいくつかの方法が知られている．それらのうちの代表的な方法を上げると

- 周辺分布指定法
- 条件付き分布指定法
- エントロピー最大化法
- 回帰指定法
- その他

のようになる．

Wicksell–Kibble 型分布:

$\mathbb{R}_+^2$ 上の Wicksell–Kibble ガンマ分布を基礎にして，二つの線形変数 $(X_1, X_2)^\top \in \mathbb{R}_+^2$ と一つの角度変数 $\Theta \in [0, 2\pi)$ の同時分布を提案し，分布の諸性質について述べた．$(X_j, \Theta)^\top$ $(j = 1, 2)$ の周辺分布は Abe and Ley (2017) の GGSSVM (generalized Gamma sine-skewed von Mises) に相当し，X_j の周辺分布は重み付き Stacy 一般ガンマ，Θ の周辺分布は sine-skewed negative exponent power cardioid となる．

分布の提案

$\mathbb{R}_+^2$ 上の Wicksell–Kibble ガンマ分布を基にして，$\gamma \in \mathbb{R}_+$, $\lambda \in [-1, 1]$, $\sigma_j, \xi_j \in \mathbb{R}_+$ $(j = 1, 2)$, $\delta \in [-1, 1]$, $\rho \in [0, 1)$, $\mu \in [0, 2\pi)$ に対し，$(X_1, X_2, \Theta)^\top$ の同時確率密度関数

$$\begin{aligned} f(x_1, x_2, \theta) &= C^{-1}\{1 + \lambda \sin(\theta - \mu)\} \frac{\xi_1 \xi_2}{(1-\rho)\Gamma(\gamma)\sigma_1\sigma_2(\rho B)^{(\gamma-1)/2}} \\ &\quad \times \left(\frac{x_1}{\sigma_1}\right)^{\xi_1(\gamma+1)/2-1} \left(\frac{x_2}{\sigma_2}\right)^{\xi_2(\gamma+1)/2-1} \exp\left\{-\frac{\left(\frac{x_1}{\sigma_1}\right)^{\xi_1} + \left(\frac{x_2}{\sigma_2}\right)^{\xi_2}}{1-\rho}\right\} \\ &\quad \times I_{\gamma-1}\left(\frac{2}{1-\rho}\sqrt{\rho B \left(\frac{x_1}{\sigma_1}\right)^{\xi_1} \left(\frac{x_2}{\sigma_2}\right)^{\xi_2}}\right), \quad (x_1, x_2)^\top \in \mathbb{R}_+^2,\ \theta \in [0, 2\pi) \end{aligned} \tag{1}$$

を持つ 9 パラメータ分布を提案した．ここで，$B = 1 + \delta(1-\rho)\cos(\theta - \mu)$，$I_\nu(\cdot)$ は次数 ν の第 1 種変形 Bessel 関数であり，その積分表現と無限級数表現は

$$I_\nu(z) = \frac{1}{2\pi}\int_0^{2\pi} \cos(\nu\theta) e^{z\cos\theta} \mathrm{d}\theta = \left(\frac{z}{2}\right)^\nu \sum_{r=0}^{\infty} \frac{1}{r!\Gamma(\nu + r + 1)} \left(\frac{z^2}{4}\right)^r$$

で与えられる．正規化定数 C は，Gauss 超幾何関数 ${}_2F_1$ を用いて

$$C = 2\pi\, {}_2F_1(\gamma/2, (\gamma+1)/2; 1; (\delta\rho)^2)$$

と表現される．もしも $\delta = 0$ ならば，$(X_1, X_2)^\top$ と Θ は独立となる．

分布の諸性質とその利用の概略:

- $(X_j, \Theta)^\top$ の周辺分布は Abe and Ley (2017) の generalized Gamma sine-skewed von Mises (GGSSVM) に相当する．
- X_j の周辺分布は重み付き Stacy ガンマとでも呼べる分布を表し，Stacy ガンマの混合として解釈できる．$\delta = 0$ もしくは $\rho = 0$ のときは Stacy ガンマに帰着する．
- Θ の周辺分布は，負指数 $-\gamma$ の sine-skewed power cardioid distribution と呼べる分布を表す．
- $\Theta = \theta$ が与えられたときの X_2 の条件付き分布と $(X_1, \Theta)^\top = (x_1, \theta)^\top$ が与えられたときの X_2 の条件付き分布を求めて，“chain rule”

$$f_{X_1,X_2,\Theta}(x_1, x_2, \theta) = f_{X_1|X_2,\Theta}(x_1|x_2, \theta) f_{X_2|\Theta}(x_2|\theta) f_\Theta(\theta)$$

  により疑似乱数を生成するアルゴリズムを与えた．
- 尤度の分解公式

$$L(\gamma, \lambda, \sigma_1, \sigma_2, \xi_1, \xi_2, \delta, \rho, \mu) = \prod_{i=1}^{n} f_{X_2|X_1,\Theta}(x_{2i}|x_{1i}, \theta_i) \prod_{i=1}^{m} f_{X_1,\Theta}(x_{1i}, \theta_i)$$

  により，missing-completely-at-random データセット

$$\left\{ \begin{array}{c} x_{1j} \\ x_{2j} \\ \theta_j \end{array} \right\} (j = 1, \ldots, n) \quad \left\{ \begin{array}{c} x_{1j} \\ \text{(missing)} \\ \theta_j \end{array} \right\} (j = n+1, \ldots, m)$$

  に対してパラメータを最尤推定可能とであることを示した．
- Θ に関する離散化について示した．

以上の諸性質を利用して，$\delta = \xi_1 = \xi_2 = 1$ としたサブモデルの下で，2017 年 6 月 1 日から同 30 日まで 30 日間の Tokyo（緯度 35°41.5′N・経度 139°45.0′E・高度 25.2m）の（朝 6:00，昼 12:00）における風速 $(X_1, X_2)^\top$ と朝 6:00 における 16 方位風向 Θ データの解析を行い，$(X_1, X_2)^\top$，$(X_1, \Theta)^\top$，$(X_2, \Theta)^\top$ の散布図と相関係数を図として与え，パラメータを最尤推定した．

参考文献

[1] Abe, T. and Ley, C. (2017). A tractable, parsimonious and flexible model for cylindrical data, with applications, *Econometrics and Statistics*, **4**, 91–104.

[2] 清水邦夫，Dou Xiaoling and Liu Shuangzhe（2025). A Wicksell–Kibble type distribution with bivariate linear and univariate circular variables, 統計関連学会連合大会（関西大学）.