

## 科研費シンポジウム

「多様な分野における統計科学の総合的研究」

日時：2017年11月17日（金）～11月19日（日）  
(Date: Friday, 17, November – Sunday, 19, November, 2017)

場所：コープシティ花園4F ガレッショホール a  
(Place: Coop City Hanazono 4F Garesso Hall a)  
(TEL: 025-248-7511)

科学研究費・基盤研究（A）（課題番号：15H01678）  
「大規模複雑データの理論と方法論の総合的研究」  
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（新潟大学）  
Junichi Hirukawa (Niigata University)

## Program

Friday, 17, November

Reception 13:00-13:10

Opening 13:10-13:15 Junichi HIRUKAWA (Niigata University)

Afternoon Session I (in Japanese) 13:15-15:15

Chair Takuma Yoshida (Kagoshima University)

1. 13:15-13:55 阿部 文貴 (Abe Fumitaka), 白石 博 (Shiraishi Hiroshi)

慶應義塾大学理工学研究科 (Keio University Graduate School of Science and Technology)

慶應義塾大学理工学部数理科学科 (Keio University Department of Mathematics)

「長期記憶性を持った高次元ポートフォリオの分散に対する収束性の比較」 (Comparison of convergence rate for estimated variance of high dimensional portfolio with long-memory)

2. 13:55-14:35 泉澤 佑 (Izumisawa yu), 白石 博 (Shiraishi Hiroshi)

慶應義塾大学理工学研究科 (Keio University Graduate School of Science and Technology)

慶應義塾大学理工学部数理科学科 (Keio University Department of Mathematics)

「Hawkes 過程によるシステミックリスク評価」 (Systemic Risk Assessment by Hawkes process)

3. 14:35-15:15 藤森 洸 (Kou Fujimori), 西山陽一 (Yoichi Nishiyama)

早稲田大学大学院基幹理工学研究科 (Waseda University, Graduate School of Fundamental Science and Engineering)

早稲田大学国際教養学部 (School of International Liberal Studies, Waseda University)

The Dantzig selector for diffusion processes with high-dimensional parameter

Coffee Break 15:15-15:30

Afternoon Session II (in Japanese) 15:30-17:30

Chair Yoichi Nishiyama (School of International Liberal Studies, Waseda University)

4. 15:30-16:10 吉田拓真 (Takuma Yoshida),

鹿児島大学 (Kagoshima University)

Nonparametric smoothing for extremal quantile regression

5. 16:10-16:50 Masanari, MOTOYAMA and Yoshihiko MAESONO,

National Federation of Agricultural Cooperative Associations

Faculty of Mathematics, Kyushu University

On direct kernel estimator of density ratio

6. 16:50-17:30 柿沢佳秀 (Yoshihide Kakizawa),

北海道大学 (Hokkaido University)

「対称分布ベース  $q$ -MIG カーネル、及び、歪分布ベース  $q$ -BS カーネルに基づく、非負データに対する密度推定」 (Symmetrical-based  $q$ -MIG kernel density estimation, and skew-based  $q$ -BS kernel density estimation for nonnegative data)

**Saturday, 18, November**

**Morning Session (in Japanese or in English) 9:10-11:50**

Chair **Takeshi Kato** (Faculty of Science and Technology, Sophia University)

7. 9:10-9:50 **張元宗 (Chang Yuan-Tsung)**, 篠崎 信雄 (Shinozaki Nobuo), William E. Strawderman

目白大学 (Mejiro University)

慶應義塾大学 (Keio University)

Rutgers University

「未知で等しい分散を持つ  $p$  個の非負な正規母平均の同時推定」 (Simultaneous estimation of  $p$  positive normal means with common unknown variance)

8. 9:50-10:30 **島谷健一郎 (Ichiro Ken Shimatani)**,

統計数理研究所 (The Institute of Statistical Mathematics)

「方向の piece-wise regression model とその魚群データへの応用」 (Circular piece-wise regression model and its application to movement patterns for school of fishes)

9. 10:30-11:10 **石井 晶 (Ishii Aki)** ,

東京理科大学 理工学部 情報科学科 (Department of Information Sciences, Tokyo University of Science)

Asymptotic properties of classification procedures based on eigenstructures in high-dimensional context

10. 11:10-11:50 **Junichi Hirukawa** and Sangyeol Lee

Niigata University

Seoul National University

Asymptotic Properties of Mildly Explosive Processes with Locally Stationary Disturbance

**Lunch 11:50-13:15**

**Afternoon Session I (in English) 13:15-15:15**

Chair **Shiraishi Hiroshi** (Keio University Department of Mathematics)

11. 13:15-13:55 **Daichi Inoue** and Takeshi Kato

Graduate School of Science and Technology, Sophia University

Faculty of Science and Technology, Sophia University

Activation Function Selection in Recurrent Neural Network

12. 13:55-14:35 **阿部俊弘 (Abe, Toshihiro)**, 塩濱 敬之 (Takayuki Shiohama), 宮田 庸一 (Yoichi Miyata)

南山大学理工学部 (Nanzan university, Faculty of Science and Engineering)

東京理科大学工学部 (Department of Information and Computer Technology, Tokyo University of Science)

高崎経済大学経済学部 (Faculty of Economics, Takasaki City University of Economics)

On mode and antimode preserving circular distributions

13. 14:35-15:15 **宮田 庸一 (Yoichi Miyata)**, 塩濱 敬之 (Takayuki Shiohama), 阿部 俊弘 (Toshihiro Abe)

高崎経済大学経済学部 (Faculty of Economics, Takasaki City University of Economics)

東京理科大学工学部 (Department of Information and Computer Technology, Tokyo University of Science)

南山大学理工学部 (Department of Systems and Mathematical Science, Nanzan University)

「超球面上の非対称な分布の有限混合モデルに対する推定について」 (On estimating finite mixtures of skew-rotationally-symmetric distributions)

**Coffee Break** 15:15-15:30

**Afternoon Session II (Guest speakers session)** 15:30-17:30

Chair **Junichi Hirukawa** (Niigata University)

14. 15:30-16:30 **Giuseppe Cavaliere** and Iliyan Georgiev

University of Bologna

BOOTSTRAP INFERENCE UNDER RANDOM DISTRIBUTIONAL LIMITS

15. 16:30-17:30 **Shiqing Ling**, Ruey S. Tsay and Yaxing Yang

Hong Kong University of Science and Technology

University of Chicago

Xiamen University

Testing for Series Correlation and ARCH Effect of High-Dimensional Time Series Data

**Conference Dinner** 18:00-21:00

ぜんてい 新潟駅南店

(Zentei Niigata Ekinanten)

5,500 yen

(TEL: 025-244-8807)

**Sunday, 19, November**

**Morning Session (in Japanese) 9:10-11:50**

Chair **Abe, Toshihiro** (Department of Systems and Mathematical Science, Nanzan University)

**16. Shuichi Shinmura,**

Seikei University, Prof. Emeritus

Cancer Gene Analysis by Singh et al. Microarray Data

**17. 9:50-10:30 佐藤 駿 (Shun Sato), 合田 新平 (Shimpei Goda), 三枝 祐輔 (Yusuke Saigusa), 生亀 清貴 (Kiyotaka Iki), 山本 紘司 (Kouji Yamamoto)**

東京理科大学大学院 理工学研究科 (Graduate School of Science and Technology, Tokyo University of Science)

東京理科大学大学院 理工学研究科 (Graduate School of Science and Technology, Tokyo University of Science)

横浜市立大学 医学部 (Faculty of Medicine, Yokohama City University)

) 東京理科大学 理工学部 (Faculty of Science and Technology, Tokyo University of Science)

大阪市立大学大学院 医学研究科 (Graduate School of Medicine, Osaka City University)

「二元分割表における独立性および対称性に関するモデルと分解」 (Models and decompositions of independence and symmetry for two-way contingency tables)

**18. 10:30-11:10 安藤宗司 (Shuji Ando), 田畑耕治 (Kouji Tahata), 富澤貞男 (Sadao Tomizawa)**

東京理科大学工学部 (Faculty of Engineering, Tokyo University of Science)

東京理科大学理工学部 (Faculty of Science and Technology, Tokyo University of Science)

東京理科大学理工学部 (Faculty of Science and Technology, Tokyo University of Science)

「正方分割表における対称性の尺度について」 (On measures for symmetry in square contingency tables)

**19. 11:10-11:50 野間久史 (Hisashi Noma) ,**

統計数理研究所 (The Institute of Statistical Mathematics)

「ネットワークメタアナリシスによる Comparative Effectiveness Research と高次漸近理論に基づく推測手法」 (Comparative effectiveness research using network meta-analysis and effective inference methods via higher order asymptotics)

**Closing 11:50-11:55 Junichi HIRUKAWA** (Niigata University)

# 長期記憶性を持った高次元ポートフォリオの分散 に対する収束性の比較

## Comparison of convergence rate for estimated variance of high dimensional portfolio with long-memory

慶應義塾大学理工学研究科 阿部 文貴  
慶應義塾大学理工学部 白石 博

### 1 はじめに

ポートフォリオ理論とは分散投資を行う際にリスクを回避し、収益率を高めるには、どう最適化すればよいかを決定する理論であり、平均-分散基準では、個々の投資資産の期待収益率  $\boldsymbol{\mu}$  と分散共分散行列  $\Sigma$  の関数で、最適配分比率が決定する。古典的な手法では、 $\boldsymbol{\mu}$  と  $\Sigma$  の代わりに過去の収益率の標本平均ベクトル  $\hat{\boldsymbol{\mu}}$  と標本分散共分散行列  $\hat{\Sigma}_{sam}$  を用いるが、近年の市場規模の拡大により投資する資産の数が増加し、 $\hat{\Sigma}_{sam}$  が良い推定量ではなくなっている。そこで Fan et al(2008) は収益率にファクター構造を考え、最小二乗法による  $\Sigma$  の推定量  $\hat{\Sigma}_{LS}$  の方がより収束レートが良い事を示した。しかし、Fan et al(2008) では誤差項が i.i.d. と仮定していたが、その仮定は現実的ではないと思われる。本発表では、誤差項が長期記憶過程に従うと仮定し、その下でのポートフォリオの収束性について議論する。

### 2 GMV ポートフォリオとファクターモデル

$Y_{it}$  を資産  $i$  の時刻  $t$  における収益率とする。ここで  $\mathbf{y}_t = (Y_{1t}, \dots, Y_{pt})^\top$  は平均ベクトル  $E(\mathbf{y}_t) = \boldsymbol{\mu}$ 、分散共分散行列  $\text{Cov}(\mathbf{y}_t) = \Sigma > \mathbf{0}$  の分布に従っていると仮定する。さらに配分比率を  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)^\top$  とする。配分比率はその総和が 1 となるようにする。この時、ポートフォリオの収益率の分散は  $\text{Var}(\boldsymbol{\xi}^\top \mathbf{y}_t) = \boldsymbol{\xi}^\top \Sigma \boldsymbol{\xi}$  となる。この分散が最も小さくなる GMV ポートフォリオの配分比率  $\boldsymbol{\xi}^*$  とその時の分散  $\boldsymbol{\xi}^{*\top} \Sigma \boldsymbol{\xi}^*$  は

$$\boldsymbol{\xi}^* = \frac{\Sigma^{-1} \mathbf{1}_p}{\mathbf{1}_p^\top \Sigma^{-1} \mathbf{1}_p} \quad \boldsymbol{\xi}^{*\top} \Sigma \boldsymbol{\xi}^* = \frac{1}{\mathbf{1}_p^\top \Sigma^{-1} \mathbf{1}_p} \quad (*)$$

これより GMV ポートフォリオの分散を推定する事は  $\Sigma^{-1}$  を推定することに帰着する。

ファクターモデルは経済工学や金融工学の分野で幅広く用いられている概念である。ファクターモデルの下で収益率  $Y_{it}$  は、次のように定義される。また次の仮定を置く。

(仮定 1)  $p$  は  $n$  と共に増加し  $p \gg K$ 。 (仮定 2)  $\mathbf{f}_t = (f_{1t}, \dots, f_{Kt})$  と  $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \dots, \epsilon_{pt})^\top$  は無相関。

定義 2.1 (誤差項が長期記憶性を持つファクターモデル)

$$Y_{it} = \mathbf{b}_i^\top \mathbf{f}_t + \epsilon_{it} \quad i = 1, \dots, p$$

ここで  $f_{1t}, \dots, f_{Kt}$  はファクター (確率変数)、 $b_{i1}, \dots, b_{iK}$  ( $i = 1, \dots, p$ ) はファクターローディングと呼ばれる係数である。特に誤差項  $\{\epsilon_{it}\}$  は長期記憶過程に従い、 $E(\epsilon_{it}) := 0, \text{Var}(\epsilon_{it}) := \sigma_i^2, \text{Cov}(\epsilon_{i0}, \epsilon_{it}) := \sigma_i^2 \gamma_t$ , ( $t = 1, 2, \dots$ ) とおくと  $\sum_{t=0}^{\infty} |\gamma_t| = \infty$  を満たす。

さらに行列形式にまとめて書くと

$$\mathbf{y}_t = \mathbf{B} \mathbf{f}_t + \boldsymbol{\epsilon}_t \quad (2.1)$$

ここで、 $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_p)^\top$  である。これから  $\mathbf{y}_t$  の分散共分散行列は

$$\begin{aligned}\Sigma &= \text{Cov}[\mathbf{y}_t] = \text{Cov}[\mathbf{B}\mathbf{f}_t] + \text{Cov}[\boldsymbol{\epsilon}_t] \\ &= \mathbf{B}\text{Cov}[\mathbf{f}_t]\mathbf{B}^\top + \Sigma_0\end{aligned}\quad (2.2)$$

と書けることがわかる。さらに  $\mathbf{X} = (\mathbf{f}_1, \dots, \mathbf{f}_n)$ ,  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ ,  $\mathbf{E} = (\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n)$  と定める。(2.2) 式から、 $\Sigma$  の推定量  $\widehat{\Sigma}_{LS}$ ,  $\widehat{\Sigma}_{BLU}$ ,  $\widehat{\Sigma}_{sBLU}$ ,  $\widehat{\Sigma}_{FGLS}$  を定める。しかし、自己共分散行列  $\Omega$  を推定し、そこから逆行列を考える時、誤差が大きくなることが知られている。この誤差を小さくするために、 $\Omega^{-1}$  を直接推定する。

**定義 2.2** (閾値  $\kappa$  を使った  $\Omega^{-1}$  の近似)

$\Omega^{-1}$  を修正コレスキー分解によって  $\Omega^{-1} = \mathbf{T}^\top \mathbf{D}^{-1} \mathbf{T}$  となるような  $\mathbf{T}, \mathbf{D}$  を定める。ここで  $\mathbf{T} = (t_{jk})$  : 下三角行列,  $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$  : 対角行列 である。この時、閾値  $\kappa$  を用いて、次のように  $\mathbf{T}(\kappa), \mathbf{D}(\kappa)$  を定める。

$$\mathbf{T}(\kappa) = (t_{jk}(\kappa)) = \begin{cases} 0 & \text{if } \kappa + 1 < j \leq n, 1 \leq k \leq j - \kappa - 1 \\ t_{\kappa k} & \text{if } \kappa + 1 \leq i \leq n, j - \kappa \leq k \leq j - 1 \\ t_{jk} & \text{if otherwise} \end{cases}$$

$$\mathbf{D}(\kappa) = \text{diag}(\sigma_1, \dots, \sigma_\kappa, \dots, \sigma_n)$$

そしてこれらの値を用いて  $\Omega^{-1}(\kappa) = \mathbf{T}(\kappa)^\top \mathbf{D}^{-1}(\kappa) \mathbf{T}(\kappa)$  と定める。

**定義 2.3** (FGLSE 実現可能一般化最小二乗推定量)

$$\widehat{\Sigma}_{FGLS} = \widehat{\mathbf{B}}_{FGLS}^\top \widehat{\text{Cov}}(\mathbf{f}) \widehat{\mathbf{B}}_{FGLS} + \widehat{\Sigma}_0 \quad (2.3)$$

ただし、 $\widehat{\mathbf{B}}_{FGLS} = (\mathbf{X}^\top \tilde{\Omega}^{-1}(\kappa) \mathbf{X})^{-1} \mathbf{X}^\top \tilde{\Omega}^{-1}(\kappa) \mathbf{Y}$ ,  $\widehat{\Sigma}_{0,FGLS} = \text{diag}(n^{-1} \widehat{\mathbf{E}}_{FGLS}^\top \widehat{\mathbf{E}}_{FGLS})$ ,  $\widehat{\mathbf{E}}_{FGLS} = \mathbf{Y} - \mathbf{X} \widehat{\mathbf{B}}_{FGLS}$  として、 $\tilde{\Omega}^{-1}(\kappa)$  は定義 2.2 の  $\Omega$  に  $\tilde{\Omega}$  を代入したものである。また  $\kappa$  は Ing et al(2016) に従う。

### 3 数値結果

紹介した推定量と真の値がどの程度近くなっているかをシミュレーションによって確かめる。シミュレーションにはまず 4083 銘柄の対数収益率に対して JASDAQ, 日経平均, TOPIX をファクターとしたファクターモデルを作る。そしてそこから得られた、ファクターローディングの標本平均  $\boldsymbol{\mu}_b$  と標本分散共分散行列  $\Sigma_b$  を計算する。同時に、ファクターの標本平均  $\boldsymbol{\mu}_f$  と標本分散共分散行列  $\Sigma_f$  を計算する。そしてこれらをパラメータとする多変量正規分布からそれぞれファクターローディングとファクターの標本を生成する。そしてこれらの値と ARFIMA(0,d,0) 過程から生成した誤差項の標本を使い、定義 (2.1) で定めたファクターモデルに従って対数収益率の標本を計算する。これらの生成した標本を用いて、紹介した GMV ポートフォリオの分散の推定量を作成し、パラメータから作成した真の値との差を RMSE によってどの程度乖離しているか評価する。実際にシミュレーションを行った詳細な結果については当日報告した。

### 参考文献

- [1] Ledoit, O., and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5), 603-621.
- [2] Bai, Z., Liu, H., and Wong, W. K. (2009). Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. *Mathematical Finance*, 19(4), 639-667.
- [3] Bodnar, T., Parolya, N., and Schmid, W. (2014). Estimation of the global minimum variance portfolio in high dimensions. arXiv preprint arXiv:1406.0437.
- [4] Goldfarb, D., and Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of operations research*, 28(1), 1-38.
- [5] Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1), 186-197.
- [6] Ing, C. K., Chiou, H. T., and Guo, M. (2016). Estimation of inverse autocovariance matrices for long memory processes. *Bernoulli*, 22(3), 1301-1330.
- [7] 刘屋武昭, 矢島美寛, 田中勝人, and 竹内啓. (2003). 経済時系列の統計. *J 金融 時系列 分析入門*, 岩波書店, 3, 47-67.

# Hawkes 過程によるシステミックリスク評価 Systemic Risk Assessment by Hawkes process

慶應義塾大学理工学研究科 泉澤 佑  
慶應義塾大学理工学部 白石 博

## 1 Introduction

近年、システミックリスクと呼ばれる個別の金融機関のショックが金融システム全体の危機を引き起こすリスクの顕在化が憂慮されている。システミックリスクの重要な要素の一つとして金融危機の伝播・波及というメカニズムがある。金融システム全体の安定性を脅かすようなショックの波及の評価は、システミックリスク研究において重視される。しかし、個々の金融機関でのショックの波及のしやすさ、およびその波及の経路を判断することは困難が生じる。これは波及が取引ネットワークの複雑化、レバレッジなど様々な要因によることに起因し、システミックリスクを評価する際の問題となる。

システミックリスクを定量化するため、金融機関の関連性を取引などでモデル化するなどの研究が行われてきた。[2]では、ショック発生時の悪影響の程度を金融機関同士の取引などでモデル化することでシステミックリスクを定量化している。しかしながらこのようなモデルは悪影響をもたらす直接的、間接的な金融システムの状態・要因全てを捉えられているとは限らない。本研究では Hawkes 過程と呼ばれる点過程モデルを利用することで実経済上に現れる損失発生 の波及をモデル化する。このモデルを用いてショックの波及を実経済上に現れる損失の動きから直接、波及経路や波及のしやすさを読み取り、システミックリスクを評価する。

## 2 Hawkes 過程によるシステミックリスク評価

Hawkes 過程は [3] によって提案された点過程確率モデルである。Hawkes 過程はポアソン型のモデルで、強度パラメータに過去のイベントの影響による励起を加えた形で定義 2.1 のように定義される。

**定義 2.1.**  $d \in \mathbb{N}$  に対して  $\mathbf{N} = (N_1, \dots, N_d)^T$  を  $d$  変量 Hawkes 過程とする。このときある  $d$  変量ベクトル  $\boldsymbol{\eta} \in \mathbb{R}_{\geq 0}^d$ 、可測な  $d$  変量正方形行列値関数  $H = (h_{i,j})_{1 \leq i,j \leq d} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}^{d \times d}$  に対して  $d$  変量 Hawkes 過程の条件付き強度は

$$\Lambda_{\mathbf{N}}(t) := \lim_{\delta \downarrow 0} \frac{\mathbb{E}[\mathbf{N}((t, t + \delta]) | \mathcal{H}_t^{\mathbf{N}}]}{\delta} = \boldsymbol{\eta} + \int_{-\infty}^t H(t-s)\mathbf{N}(ds), \quad t \in \mathbb{R}. \quad (2.1)$$

ここで各  $i \in \{1, \dots, d\}$  に対し

$$\left( \int_{-\infty}^t H(t-s)\mathbf{N}(ds) \right)_i := \left( \sum_{j=1}^d \int_{-\infty}^t h_{i,j}(t-s)N_j(ds) \right)_i, \quad (2.2)$$

とし、また  $\mathcal{H}_t^{\mathbf{N}} := \sigma(\{\omega \in \Omega : \mathbf{N}((a, b]) = \mathbf{n}, \mathbf{n} \in \mathbb{N}_0^d, a < b \leq t\})$  とする。このような  $\boldsymbol{\eta} \in \mathbb{R}_{\geq 0}^d$  を基底強度、 $H = (h_{i,j})_{1 \leq i,j \leq d} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}^{d \times d}$  を励起関数行列と呼ぶ。

Hawkes 過程の強度は基底強度と過去のイベント発生からの励起の和として表される。 $H$  は各成分に関数が入った行列で、例えば  $H_{i_2 i_1}$  は  $i_1$  でイベントが発生した場合の  $i_2$  の強度の励起度合いを表す。

Hawkes 過程の定義は個別のイベント発生の影響が波及し、他でのイベント発生を引き起こすという考えをモデル化したもので、システミックリスク顕在化のメカニズムと非常に合致したものである。金融システムにおけるショックの波及ネットワークを Hawkes 過程によりモデル化することで、システミックリスクを評価する。Hawkes 過程の波及を評価するためその可視化である Hawkes グラフ、そしてそれを元にした二つの指標を導入する。Hawkes グラフとは [1] で提案された Hawkes 過程の波及構造を重み付き有向グラフにより可視化したもので定義 2.2 のように定義される。

**定義 2.2.**  $\mathbf{N}$  を基底強度  $\boldsymbol{\eta}$ 、励起関数行列  $H$  の  $d$  変量 Hawkes 過程とし、 $K$  を次のように定義する。

$$K := \left( \int_0^{\infty} H_{j,i}(t) dt \right)_{1 \leq i,j \leq d}$$

このとき  $\mathbf{N}$  に対する Hawkes グラフ  $G$  は次の式を満たす頂点集合  $V_{\mathbf{N}}$ 、辺集合  $E_{\mathbf{N}}$  を持つ重み付き有向グラフとして定義される。

$$V_{\mathbf{N}} := \{(j; \eta_j) : j \in \{1, 2, \dots, d\}; \eta_j \text{ 重み}\}, \\ E_{\mathbf{N}} := \{(i, j; K_{i,j}) : (i, j) \in V_{\mathbf{N}} \times V_{\mathbf{N}}, K_{i,j} > 0; K_{i,j} \text{ 重み}\}.$$

定義 2.2 において  $K$  は Hawkes グラフの隣接行列にあたる。 $K_{i,j} > 0$  であり Hawkes グラフに  $i$  から  $j$  への辺が存在



するとき、つまり  $j$  でのイベント発生確率が  $i$  の過去の情報に依存するとき  $i$  から  $j$  の影響が存在することを表し、辺の重みは影響の大きさを表す。頂点の重みは各変量の基底強度であるのでイベント発生時の基底となる確率を表す。Hawkes 過程からそれに対する Hawkes グラフを構成することで、波及ネットワークの可視化ができる。

また、Hawkes 過程の特徴を表す指標として cascade 係数、feedback 係数を導入する。それぞれ定義 2.3 のように定義される。

**定義 2.3.**  $\mathbf{N}$  を subcritical な  $d$  変量 Hawkes 過程とし、 $K, G_{\mathbf{N}}$  をそれぞれ対応した隣接行列と、それに基づく Hawkes グラフとする。 $(e_{i,j}) := \sum_{g \geq 0} K^g \in \mathbb{R}_{\geq 0}^{d \times d}$  として  $e_{i,j}$  を定め、これを用いて次のように定義をする。

$$c_i := \frac{\eta_i \sum_{j=1}^d e_{i,j}}{\sum_{l=1}^d \eta_l \sum_{j=1}^d e_{l,j}}, \quad f_j := \frac{\eta_j e_{j,j}}{\sum_{i=1}^d \eta_i e_{i,j}}, \quad j \in \{1, 2, \dots, d\}.$$

$\{c_i\}_{i=1, \dots, d}$  を cascade 係数、 $\{f_j\}_{j=1, \dots, d}$  を feedback 係数と呼ぶ。

各  $i$  について cascade 係数の大きさは Hawkes 過程全体でのイベント発生と影響の波及全体に対する頂点  $i$  の寄与を表す。対して  $j$  での feedback 係数の  $j$  が受ける波及による影響への  $j$  自身の寄与を表している。cascade 係数、feedback 係数をそれぞれ金融機関ごとの金融システム全体でのリスクの顕在化への寄与度、金融ネットワークにおける自身へのショックの伝播に対する寄与度と見ることによってシステムリスク評価の指標とすることができる。

### 損失波及モデル

Hawkes グラフはイベントの波及を評価するモデルであった。システムリスクにおけるショックの伝播による損失の定量的な評価を考える。損失の発生が Hawkes 過程に従うとし、発生した損失に確率分布を考えるマーク付き Hawkes, Poisson 過程を利用する。定義した損失を、ショックの波及により生じる損失の定量的な評価に用いる。

$\mathbf{L}(B) := (L_1(B), \dots, L_d(B))^T, d \in \mathbb{N}, B \in \mathcal{B}(\mathbb{R})$  をある期間  $B$  の経済指標・金融商品の損失を表す確率変数とし、 $\mathbf{L}$  を次のように定義する。

$$\mathbf{L}(B) := \begin{pmatrix} L_1(B) \\ \vdots \\ L_d(B) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{N_1(B)} X_{1,k} \\ \vdots \\ \sum_{k=1}^{N_d(B)} X_{d,k} \end{pmatrix} - \begin{pmatrix} \sum_{k=1}^{N_1^*(B)} X_{1,k} \\ \vdots \\ \sum_{k=1}^{N_d^*(B)} X_{d,k} \end{pmatrix}. \quad (2.3)$$

ここで  $\mathbf{N}, \mathbf{N}^*$  をそれぞれ  $d$  次元 Hawkes 過程、 $d$  次元 Poisson 過程に従う確率変数とする。また、 $X_{i,k}$  を  $\mathbf{N}, \mathbf{N}^*, i, k$  について独立な損失を表す確率変数とする。Hawkes 過程はイベント発生時の影響による強度パラメータの励起を Poisson に加えたモデルであることを考慮すると、この  $\mathbf{L}$  は全体の損失から基底強度による、つまり過去の損失の影響を除いた波及によって生じる損失のみを表している。 $\mathbf{L}$  に対してリスク尺度を適用することによりシステムリスクにおける波及による損失の定量的な評価とする。

## 3 数値結果

紹介した指標を金融機関・指標に適用することで、金融システムの脆弱性によるリスクを評価する。経済指標、金融商品の対数成長率の符号を逆にしたものを負のリターンとする。評価の対象となる期間における負のリターンが適当な閾値を超えた値を損失と定義し、その発生が Hawkes 過程に従い、大きさが一般化パレート分布に従うとする。推定した Hawkes 過程のパラメータから Hawkes グラフ、cascade 係数、feedback 係数を推定する。または波及による損失  $\mathbf{L}$  の期待値をとることでシステムリスクの定量的な評価を行う。実際の金融市場のデータを解析した詳細な結果については当日報告した。

### 参考文献

- [1] Embrechts, Paul, and Matthias Kirchner. "Hawkes graphs." arXiv preprint arXiv:1601.01879 (2016).
- [2] Adrian, Tobias, and Markus K. Brunnermeier. "CoVaR." The American Economic Review 106.7 (2016): 1705-1741.
- [3] Hawkes, A. G., & Oakes, D. (1974). A cluster process representation of a self-exciting process. Journal of Applied Probability, 11(3), 493-503.

# The Dantzig selector for diffusion processes with high-dimensional parameters

Kou Fujimori    Yoichi Nishiyama

Waseda University

In this talk, statistical models of diffusion processes with high-dimensional parameters were discussed. To estimate high-dimensional parameters, the Dantzig selector which was proposed by Candés and Tao in 2007 is applied. Then, we proved two types of consistency of the estimator; one is the  $l_q$  consistency for every  $q \in [1, \infty]$  and the other is the variable selection consistency. Moreover, we talked about the construction of an asymptotically normal estimator by using the variable selection consistency of the Dantzig selector.

For example, we considered the following linear model of diffusion process:

$$X_t = X_0 + \int_0^t \Theta^T \phi(X_s) ds + \sigma W_t, \quad (1)$$

where  $\{W_t\}_{t \geq 0} := \{(W_t^1, \dots, W_t^p)\}_{t \geq 0}$  is a  $p$ -dimensional standard Brownian motion,  $\Theta$  is a  $p \times p$  sparse deterministic matrix, and  $\sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$  is a  $p \times p$  diagonal matrix and  $\phi(x) = (\phi_1(x_1), \dots, \phi_p(x_p))$  for  $x = (x_1, \dots, x_p) \in \mathbb{R}^p$  is a smooth  $\mathbb{R}^p$ -valued function. Note that  $\{X_t^i\}_{t \geq 0}$  for each  $i = 1, 2, \dots, p$  satisfies the following equation.

$$X_t^i = X_0^i + \int_0^t \Theta_i^T \phi(X_s) ds + \sigma_i W_t^i,$$

where  $\Theta_i$  is an  $i$ -th row of matrix  $\Theta$ . In this talk, we considered the estimation problem of  $\Theta$  and  $\sigma$  using the observed data of the process  $\{X_t\}_{t \geq 0}$  at  $n+1$  discrete time points  $0 =: t_0^n < t_1^n < \dots < t_n^n$ , where  $t_k^n = kt_n^n/n$ ,  $k = 0, 1, \dots, n$ . Write  $T_0^i$  for the support index set of the true value  $\Theta_i^0$  for every  $i \in \{1, 2, \dots, p\}$ , i.e.,  $T_0^i = \{j : \Theta_{ij}^0 \neq 0\}$ . Let  $S_i$  be the number of elements in the index set  $T_0^i$ . The estimation problem is considered in the following high-dimensional and sparse setting for the true value  $\Theta^0$ .

$$p = p_n \gg n, \quad \sup_{1 \leq i < \infty} S_i =: S^* \ll n,$$

where  $S^* > 0$  is a constant which does not depend on  $n$ . It is well known that we can ignore the influence of  $\Theta$  when we estimate the diffusion coefficients  $\sigma$ .

Therefore, we define the estimator  $\hat{\sigma}_i$  for  $\sigma_i^0$  for every  $i = 1, 2, \dots, p_n$  by the maximum quasi-likelihood estimator, which satisfies the consistency uniformly in  $i$ . Then, we define the Dantzig selector type estimator for  $\Theta_i^0$  by plugging  $\hat{\sigma}_i$  in the log-quasi-likelihood as follows:

$$\hat{\Theta}_{n,i} := \hat{\Theta}_i := \arg \min_{\Theta_i \in \mathcal{C}_n^i} \|\Theta_i\|_1, \quad \mathcal{C}_n^i := \{\Theta_i \in \mathbb{R}^{p_n} : \|\psi_n(\Theta_i)\|_\infty \leq \gamma_n\},$$

where  $\gamma_n$  is a tuning parameter and  $\psi_n(\Theta_i) = \psi_n(\Theta_i, \sigma_i)$  is the gradient of normalized log-quasi-likelihood with respect to  $\Theta_i$ . We discussed the asymptotic property of this estimator. The next theorem states the  $l_q$  consistency of  $\hat{\Theta}_i$ .

**Theorem 1.** *Suppose that  $p_n \gg n$ ,  $\log p_n = o(\sqrt{n\Delta_n})$ ,  $n\Delta_n = t_n^n \rightarrow \infty$ ,  $\Delta_n \rightarrow 0$ , and  $n\Delta_n^2 \rightarrow 0$  as  $n \rightarrow \infty$ . Put  $\gamma_n = O(\log(1 + p_n^2)/\sqrt{n\Delta_n})$ . For global positive constants  $K_1, K_2, K_3$ , the following hold true under some appropriate conditions.*

(i)

$$\lim_{n \rightarrow \infty} P \left( \sup_{1 \leq i \leq p_n} \|\hat{\Theta}_i - \Theta_i^0\|_1 \geq K_1 S^* \gamma_n \right) = 0.$$

(ii)

$$\lim_{n \rightarrow \infty} P \left( \sup_{1 \leq i \leq p_n} \|\hat{\Theta}_i - \Theta_i^0\|_q \geq K_2 S^{*\frac{1}{q}} \gamma_n \right) = 0, \quad q \in (1, \infty).$$

(iii)

$$\lim_{n \rightarrow \infty} P \left( \sup_{1 \leq i \leq p_n} \|\hat{\Theta}_i - \Theta_i^0\|_\infty^2 \geq K_3 \gamma_n \right) = 0.$$

Then, we talked about the next theorem which states the variable selection consistency of  $\hat{\Theta}_i$ .

**Theorem 2.** *The estimator  $\hat{T}_n^i$  for  $T_0^i$  defined by*

$$\hat{T}_n^i = \{j : |\hat{\Theta}_{ij}| > \gamma_n^{\frac{1}{2}}\}$$

*satisfies the following condition under the same assumption as Theorem 1*

$$\lim_{n \rightarrow \infty} P \left( \hat{T}_n^i = T_0^i \text{ for all } i \in \{1, 2, \dots, p_n\} \right) = 1.$$

Finally, we talked about the construction of the asymptotically normal estimator.

**Theorem 3.** *Define the estimator  $\hat{\Theta}_i^{(2)}$  by the solution to the following equation:*

$$\psi_n(\Theta_{i\hat{T}_n^i})_{\hat{T}_n^i} = 0, \quad \Theta_{\hat{T}_n^i c} = 0. \quad (2)$$

*Then, under some appropriate assumptions, it holds for every  $i$  that*

$$\sqrt{t_n^n} (\hat{\Theta}_{i\hat{T}_n^i}^{(2)} - \Theta_{iT_0^i}^0) 1_{\{\hat{T}_n^i = T_0^i\}} \rightarrow^d N(0, Q_{T_0^i, T_0^i}^{i-1}),$$

*where  $Q_{T_0^i, T_0^i}$  is some  $S_i \times S_i$  invertible matrix.*

# Nonparametric smoothing for extremal quantile regression

吉田拓真 (TAKUMA YOSHIDA)<sup>1</sup>  
<sup>1</sup> 鹿児島大学 (Kagoshima University)  
E-mail: yoshida@sci.kagoshima-u.ac.jp

## 1 はじめに

分位点回帰は外れ値に対するロバスト性を有するなど平均回帰にはない利点を持つが、分位点  $\tau$  が 0 や 1 に近い場合 (極値分位点と呼ぶ) はその限りではない。極値分位点では、周辺のデータ数が中央に比べ少なくなることから推測が困難となる。その困難さを定式化するためには、漸近理論と極値理論の融合的な議論を行う必要がある。Chernozhukov (2005), Chernozhukov and Fernández-val (2011), Wang et al. (2012) では線形モデルについて極値分位点回帰の理論結果を与えている。しかし、中央値が線形構造を有する場合でも、最大値や最小値に近い分位点が線形構造を有するケースは稀である。このような背景から、講演者は極値分位点回帰のノンパラメトリック推定の議論が重要であると考えた。本講演では、極値分位点回帰のノンパラメトリック推定の収束レートと関連する漸近理論結果を報告した。

## 2 極値分位点回帰

目的変数  $Y$  と説明変数  $x$  に対して、 $F$  を  $Y$  の分布関数、 $F_Y(\cdot|x)$  を  $x$  を与えた下での  $Y$  の条件付分布関数とする。また、 $q(\tau) = F^{-1}(\tau)$ ,  $q_Y(\tau|x) = F_Y^{-1}(\tau|x)$  をそれぞれ  $Y$  の 100 $\tau$ % 分位点、 $x$  を与えた下での  $Y$  の条件付 100 $\tau$ % 分位点とする。極値分位点の理論評価をするためには、分布の裾の推定の困難さを数学的に扱う必要がある。そこで、データ数  $n$  が大きくなる漸近理論の枠組みで、 $\tau$  が 1 に収束するような設定を考える。今回は  $n(1-\tau) \rightarrow \infty$  と  $n(1-\tau) \rightarrow c > 0$  の 2 通りの設定でノンパラメトリック推定量の漸近収束レートを導出する。

まず、分布の裾について、以下の仮定をする:

1.  $y \rightarrow \infty$  の下で、 $F(y) = y^{-1/\gamma}L(1 + o(1))$  を満たす  $\gamma > 0$ ,  $L > 0$  が存在する。
2. ある関数  $f, H$  が存在して、 $V = Y - f(x)$  に対して、 $y \rightarrow \infty$  の下で

$$1 - F_V(y|x) = H(x)\{1 - F(y)\}(1 + o(1))$$

が成立する。ここで、 $F_V(\cdot|x)$  は  $x$  を与えた下での  $V$  の分布関数である。

すると分位点に対して  $\tau \rightarrow 1$  の下で  $q(\tau) = (1-\tau)^{-\gamma}\bar{L}(1 + o(1))$  と

$$q_Y(\tau|x) = f(x) + H(x)^\gamma q(\tau)(1 + o(1))$$

が成立する。この  $q_Y(\tau|x)$  を  $B$ -スプライン関数で近似することを考える。いま、

$$s(\tau|x) = \mathbf{B}(x)^T \mathbf{b}(\tau)$$

を  $p$  次  $B$ -スプラインモデルとする。ただし、 $\mathbf{B}(x)$  は  $B$ -スプラインベクトル、 $\mathbf{b}(\tau)$  はパラメータである。このとき、ある  $\mathbf{b}^*(\tau)$  が存在して、 $s^*(\tau|x) = \mathbf{B}(x)^T \mathbf{b}(\tau)$  が  $|s^*(\tau|x) - q_Y(\tau|x)| =$

$O(K^{-m}(1-\tau)^{-\gamma})$ を満たすことが示せる. これは,  $B$ -スプラインモデルが持つ性質の簡単な応用である. 具体的には, 同様に  $f(x)$  と  $H(x)^\gamma$  を近似する  $B$ -スプラインモデル  $s_f(x) = \mathbf{H}(x)^T \mathbf{b}_f$  と  $s_H(x) = \mathbf{H}(x)^T \mathbf{b}_H$  を用いて,  $\mathbf{b}^*(\tau) = \mathbf{b}_f + q(\tau)\mathbf{b}_H$  と構成すればよい. つまり, 上記の2つの仮定は極値分位点関数をノンパラメトリック推定を行うために極めて本質的な仮定である.

### 3 ノンパラメトリック推定

以下は報告した研究の主結果である.  $B$ -スプラインモデル  $s(x) = \mathbf{B}(x)^T \mathbf{b}$  に対して, 推定量を  $\tilde{q}_Y(\tau|x) = \mathbf{B}(x)^T \tilde{\mathbf{b}}(\tau)$  とする. ただし,  $\tilde{\mathbf{b}}(\tau)$  は

$$\sum_{i=1}^n \rho_\tau(Y_i - s(x_i)) + \lambda \int_a^b \{s^{(m)}(x)\}^2 dx,$$

の最小化として求められるものである. ここで,  $\lambda > 0$  は平滑化パラメータである.

このとき, 簡単な条件と  $n \rightarrow \infty$ ,  $\tau \rightarrow 1$ ,  $n(1-\tau) \rightarrow \infty$  の下で

$$E \left[ \left\{ \frac{\tilde{q}_Y(\tau|x)}{q_Y(\tau|x)} - 1 \right\}^2 \right] = O(\{(1-\tau_I)n\}^{-2m/(2m+1)})$$

が成立する. 次に,  $n(1-\tau) \rightarrow c > 0$  のケースに有用な推定量を構成する. まずは,  $\tau_1 > \dots > \tau_k$  に対して,

$$\hat{\gamma}(x) = \frac{1}{k-1} \sum_{j=1}^{k-1} \frac{\tilde{q}_Y(\tau_j|x)}{\tilde{q}_Y(\tau_k|x)}$$

とし, 分位点関数の推定量を

$$\hat{q}_Y(\tau|x) = \left( \frac{1-\tau_I}{1-\tau} \right)^{\hat{\gamma}(x)} \tilde{q}_Y(\tau|x)$$

とする. ただし,  $\tau_I$  は  $n(1-\tau_I) \rightarrow \infty$  を満たすとする.

このとき, 簡単な条件と  $n \rightarrow \infty$ ,  $\tau \rightarrow 1$ ,  $n(1-\tau) \rightarrow c$  の下で

$$E \left[ \left\{ \frac{\hat{q}_Y(\tau|x)}{q_Y(\tau|x)} - 1 \right\}^2 \right] = O \left( \max \left\{ k^{-\frac{2m}{2m+1}} \log^2 \left( \frac{1-\tau_I}{1-\tau} \right), \{(1-\tau_I)n\}^{-\frac{2m}{2m+1}} \right\} \right),$$

が成立する.

### 4 数値例

本講演では, 上記の提案した推定量の精度評価をモンテカルロ・シミュレーションで確認した. また, 実データに対して提案手法を適用した結果を報告した.

#### References

- Chernozhukov, V. (2005). Extremal quantile regression. *Ann. Statist.* **33** 806–839.  
 Chernozhukov, V. and Fernández-val. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *The Review of Economic Studies.* **78** 559–589.  
 Wang, H.J., Li, D. and He, X. (2012). Estimation of high dimensional conditional quantiles for heavy-tailed distributions. *J. Amer. Statist. Assoc.* **107** 1453–1464.

## On direct kernel estimator of density ratio

全国農業協同組合連合会 本山真誠  
九州大学大学院数理学研究院 前園宜彦

本講演では密度関数の比のカーネル型推定量について議論する。  $X_1, \dots, X_m$  と  $Y_1, \dots, Y_n$  をそれぞれ分布関数  $F(x)$ ,  $G(x)$  及び密度関数  $f(x)$ ,  $g(x)$  を持つ母集団からの無作為標本とする。このとき、点  $x_0 \in \mathbf{R}$  での密度関数比  $f(x_0)/g(x_0)$  の推定を考察する。ただし  $g(x_0) \neq 0$  とする。簡単のために  $m = n$  と仮定する。自然な推定量としては、分子・分母にそれぞれのカーネル型密度関数推定量を代入するものが提案されている。カーネル関数  $K$  を

$$\int K(u)du = 1$$

を満たし、 $K(-u) = K(u)$  とする。このときそれぞれの密度関数のカーネル型推定量は

$$\hat{f}(x_0) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_0 - X_i}{h}\right), \quad \hat{g}(x_0) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_0 - Y_i}{h}\right)$$

で与えられる。ただし  $h$  はバンド幅で  $h \rightarrow 0, nh \rightarrow \infty$  と仮定する。ここでは簡単のために同じカーネル関数を使用する。これを利用すると密度比の推定量として

$$\frac{\widehat{f(x_0)}}{\widehat{g(x_0)}} = \begin{cases} \frac{\hat{f}(x_0)}{\hat{g}(x_0)}, & \hat{g}(x_0) \neq 0 \\ 0, & \hat{g}(x_0) = 0 \end{cases}$$

が利用できる。

これに対して密度関数の比  $f(x)/g(x)$  は点  $G(x)$  における  $G(X_1)$  の密度関数となることを利用して Ćwik & Mielniczuk (1989) は直接型推定量

$$\frac{\widetilde{f(x_0)}}{\widetilde{g(x_0)}} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{G_n(x_0) - G_n(X_i)}{h}\right)$$

を提案している。ここで  $G_n(\cdot)$  は経験分布関数

$$G_n(y) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq y)$$

で、 $I(\cdot)$  は定義関数である。この推定量の漸近平均二乗誤差 (AMSE) は  $h = O(n^{-d})(1/6 \leq d \leq 1/4)$  のとき

$$\begin{aligned} AMSE \left\{ \frac{\widetilde{f(x_0)}}{\widetilde{g(x_0)}} \right\} &= \left\{ \frac{f''(x_0)}{2g(x_0)} - \frac{fg''(x_0)}{2g^2(x_0)} \right\}^2 \mu_2^2 h^4 \\ &+ \frac{1}{nh} \left\{ \frac{f(x_0)}{g^2(x_0)} + \frac{f^2(x_0)}{g^3(x_0)} \right\} \int_{-1}^1 K^2(u)du + O(h^5) + o\left(\frac{1}{nh}\right) \end{aligned}$$

で与えられる。  $G_n$  は階段関数であるから、上記の直接型推定量は離散的な値を取るものになる。カーネル型推定量は滑らかな推測結果が得られることが重要な利点である。この弱点を補うために、経験分布関数をカーネル型分布関数推定量で置き換えた

$$\frac{\widetilde{f(x_0)}^*}{\widetilde{g(x_0)}^*} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\widehat{G}(x_0) - \widehat{G}(X_i)}{h}\right)$$

を議論する。ここで

$$\widehat{G}(x) = \frac{1}{n} \sum_{i=1}^n W\left(\frac{x - Y_i}{h}\right), \quad W(t) = \int_{-\infty}^t K(u) du$$

である。この推定量は滑らかなものであり、適当な条件の下で下記の漸近平均二乗誤差を得ることができる。

**定理 1** (i)  $g'(u)$  は存在し  $|g'(u)| \leq M$  ( $M > 0$ ),  $u \in \mathbf{R}$ . (ii)  $K$  は対称カーネルでサポートは  $[-1, 1]$  で,  $\int u^2 K(u) du < \infty$ ,  $\int K''(u) du = 0$ ,  $|K^{(4)}(u)| \leq M$  とする。もし  $h = O(n^{-d})$  ( $1/6 < d < 1/4$ ) ならば提案する推定量の  $AMSE$  は

$$\begin{aligned} AMSE \left\{ \frac{\widetilde{f(x_0)}^*}{g(x_0)} \right\} &= \left\{ \frac{r''(G(x_0))}{2} - \frac{f g''}{2g^2}(x_0) \right\}^2 \mu_2^2 h^4 \\ &+ \frac{1}{nh} \left\{ \frac{f(x_0)}{g(x_0)} + \frac{f^2(x_0)}{g^3(x_0)} \right\} \int_{-1}^1 K^2(u) du + O(h^5) + o\left(\frac{1}{nh}\right) \end{aligned}$$

となる。ただし

$$r(u) = \frac{f(G^{-1}(u))}{g(G^{-1}(u))}$$

である。

また次の漸近正規性も成り立つ。

**定理 2** 定理 1 の仮定が成り立ち、もし  $h = n^{-1/5}$  ならば

$$\sqrt{nh} \left[ \frac{\widetilde{f(x_0)}^*}{g(x_0)} - \frac{f(x_0)}{g(x_0)} \right] \xrightarrow{L} N(\mu_*, \sigma_*^2)$$

が成り立つ。ただし

$$\begin{aligned} \mu_* &= \left\{ \frac{r''(G(x_0))}{2} - \frac{f g''}{2g^2}(x_0) \right\} \mu_2, \\ \sigma_*^2 &= \left\{ \frac{f(x_0)}{g(x_0)} + \frac{f^2(x_0)}{g^3(x_0)} \right\} \int_{-1}^1 K^2(u) du \end{aligned}$$

である。

## 参考文献

- [1] Chen, S. M., Hsu, Y. S., and Liaw, J. T. (2009). On kernel estimators of density ratio. *Statistics*, **43**, 463-479.
- [2] Ćwik, J., and Mielniczuk, J. (1989). Estimating density ratio with application to discriminant analysis. *Communications in Statistics-Theory and Methods*, **18**, 3057-3069.
- [3] Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability and Its Applications*, 10, 186-196.
- [4] Watson, G.S. (1964). Smooth regression analysis, *Sankhyā Ser. A*, 26, 359-372.

# 対称分布ベース q-MIG カーネル, 及び, 歪分布ベース q-BS カーネル に基づく, 非負データに対する密度推定 柿沢 佳秀 (北大経済)

1. はじめに  $X_1, \dots, X_n \sim \text{iid } f$  とする. Rosenblatt (1956) の (標準的な) カーネル密度推定量  $(nh)^{-1} \sum_{i=1}^n K(\frac{x-X_i}{h})$  の漸近的性質はよく知られている (Silverman (1986), Wand and Jones (1995)) が, 推定密度  $f$  の台が  $\mathbb{R}$  でないならば一般に境界バイアスがあり (一致性をもたない), この対処法も様々に議論された (リノーマライゼーション法, リフレクション法, 一般化ジャックナイフ法は Jones (1993) により詳しく解説された; なお, このカテゴリーには “局所多項式法による密度推定” も含まれる).

一方, 推定密度  $f$  が区間  $[0, 1]$ , 半無限区間  $[0, \infty) = \mathbb{R}_+$  の台を持つ場合, Chen (1999, 2000) がベータ, ガンマカーネル密度推定量を提案して以来, 『非対称カーネル法』が脚光を浴びている. 平滑化パラメータを  $b = b_n > 0$  とし, この方針から現在までに議論されている, (点  $x \in \mathbb{R}_+$  での) 密度推定量  $n^{-1} \sum_{i=1}^n k_{x,b}(X_i)$  を3つに分類しておく<sup>1</sup>.

(1) ガンマ, 逆ガンマカーネル密度推定量を含む Amoroso カーネル密度推定量族 (Igarashi and Kakizawa (2017)).

(2) バーンバウム・サンダース (BS), 逆ガウス (IG), 相反逆ガウス (RIG) カーネル密度推定量を含む, IG と RIG との  $\epsilon$ -混合 (MIG) カーネル密度推定量族 (Igarashi and Kakizawa (2014)), 及び, 対数正規 (LN), 加重 LN カーネル密度推定量 (Igarashi (2016)); なお, MIG と LN は “q-MIG” として統一的に扱える (すなわち, LN は  $q \rightarrow 0$  に相当).

(3) インデックスが  $\nu \neq \pm 1/2$  の一般化逆ガウス (GIG) カーネル密度推定量 (なお,  $\nu = -1/2$ (IG),  $1/2$ (RIG) を (2) に分類した; Igarashi and Kakizawa (2014) は GIG の代わりに別名 “変形ベッセル” を採用).

ところで, (2) のカーネルが標準正規密度  $\phi$  を元に構成されており, 本報告では次の2点から, 非負データに対する密度推定量族を拡張した<sup>2</sup>:

(A) 標準正規密度  $\phi(u)$  を対称密度  $C_q g(u^2)$  に変更することから, 対称分布をベースにした q-MIG カーネル密度推定量族 (Kakizawa (2018)) への拡張 (なお,  $q \rightarrow 0$  は対数対称カーネル密度推定量族になる).

(B) (A) の部分族である “BS 型” を, ジョンソンシステム (Johnson (1949)) の分布の変換法のアイデアにより拡張 (なお, この拡張は IG 型, RIG 型では不可能である).

---

<sup>1</sup>このような形式は “重み付き推定量” (例えば, Silverman (1986, Subsection 2.9: General weight function estimators)) であるが, Chen (1999, 2000) 以降は “非対称カーネル推定量” がキーワードである. 注意. いくつかの推定量は, オリジナルの提案 (Jin and Kawczak (2003), Scaillet (2004), Koul and Song (2013), Marchant et al. (2013), Saulo et al. (2013) など) において

(i)  $\hat{f}_b(0) = 0$  の制約 (明らかに,  $f(0) > 0$  の場合には適さない)

(ii)  $O(1)$  の境界バイアスの問題

(iii) 積分分散の主要項の存在性に関わる問題 (その存在を仮定すると,  $f(0) > 0$  なる  $\mathbb{R}_+$  上の有界な密度関数が排除される)

などがあり, いずれかの問題があるものについては, それらが回避された “再定義” も議論されている.

<sup>2</sup>報告では非負データに対するノンパラメトリック密度推定問題への応用を念頭にしているが, 関連する分布論の検討でもあることに注意したい.



2. 分布論 (A) Díaz-García and Leiva-Sánchez (2005) は BS 密度において,  $\phi(u)$  を対称密度  $C_g g(u^2)$  に替え, GBS 分布  $BS_g(\alpha, \beta)$  を定義した (報告者は, 対称分布をベースにした BS 分布と名付けた). 一方, Yang (2006) は, dual transformation

$$a^{(q)}(t) = \begin{cases} \frac{t^q - t^{-q}}{2q}, & q \neq 0, \\ \log t, & q = 0 \end{cases} \quad (\text{strictly increasing from } (0, \infty) \text{ onto } \mathbb{R})$$

(see Slymen and Lachenbruch (1984)) に基づいて, 古典的な (正規分布をベースにした) BS 密度を拡張することから,  $q \rightarrow 0$  として LN 密度へ繋いでいる. 両者のアイデアを組み合わせることで, “対称分布をベースにした  $q$ -BS 密度” が得られる:

$$k_g^{\text{BS}(q)}(s; \alpha, \beta) = C_g g\left(\frac{\{a^{(q)}(s/\beta)\}^2}{\alpha^2}\right) \frac{A^{(q)}(s/\beta)}{\alpha\beta}, \quad s \in \mathbb{R}_+. \quad [\text{BS}_g^{(q)}]$$

ここに,  $1/C_g = \int_{\mathbb{R}_+} g(u^2) du = \int_{\mathbb{R}_+} u^{-1/2} g(u) du$ ,  $A^{(q)}(t) = (d/dt)a^{(q)}(t)$ .

この  $[\text{BS}_g^{(q)}]$  を経由して, 以下のような  $[\text{IG}_g^{(q)}]$ ,  $[\text{RIG}_g^{(q)}]$  を導入するには, 積分の等式

$$\int_{\mathbb{R}_+} g((s^q - s^{-q})^2/(2q\alpha)^2) s^{q-1} ds = \int_{\mathbb{R}_+} g((t^q - t^{-q})^2/(2q\alpha)^2) t^{-(q+1)} dt \quad (s = 1/t)$$

が役立つであろう. すなわち,

$$k_g^{\text{IG}(q)}(s; \alpha, \beta) = C_g g\left(\frac{\{a^{(q)}(s/\beta)\}^2}{\alpha^2}\right) \frac{1}{\alpha\beta} \left(\frac{\beta}{s}\right)^{q+1}, \quad [\text{IG}_g^{(q)}]$$

$$k_g^{\text{RIG}(q)}(s; \alpha, \beta) = C_g g\left(\frac{\{a^{(q)}(s/\beta)\}^2}{\alpha^2}\right) \frac{1}{\alpha\beta} \left(\frac{s}{\beta}\right)^{q-1} = \left(\frac{s}{\beta}\right)^{2q} k_g^{\text{IG}(q)}(s; \alpha, \beta). \quad [\text{RIG}_g^{(q)}]$$

なお,  $g(y) = e^{-y/2} = g_N(y)$  (say),  $q = 1/2$  の場合が, Tweedie (1957; (1a)–(1d)) による古典的な (正規分布をベースにした) IG, RIG 密度であり, 古典的な (正規分布をベースにした) BS 密度は, これらの  $1/2$ -混合としても得られる (Jørgensen et al. (1991)).

(B)  $\mathbb{R}$  上の密度  $h$  をもつような確率変数  $Z_{(h)} \sim h$  に対し,  $\xi_{(h)} = \beta a^{-1}(\alpha(Z_{(h)} + \theta))$  の従う分布として『BS 型分布』を定義すると, その密度は

$$k_{(h)}^{\text{BS}}(s; \alpha, \beta, \theta) = h\left(\frac{a(s/\beta)}{\alpha} - \theta\right) \frac{A(s/\beta)}{\alpha\beta}, \quad s \in \mathbb{R}_+$$

になる (この  $q$  バージョンも同様に定義). 本報告では,  $h$  として, Azzalini (1985) 型の skew 密度, 及び, ツーピース型の epsilon-skew 密度の 2 種類を適用した (現在進行形).

### 3. 非負データに対する密度推定 対称分布ベースの $q$ -MIG カーネル

$$k_g^{\text{MIG}_\epsilon(q)}(s; \alpha, \beta) = (1 - \epsilon) k_g^{\text{IG}(q)}(s; \alpha, \beta) + \epsilon k_g^{\text{RIG}(q)}(s; \alpha, \beta), \quad \epsilon \in [0, 1]$$

及び, 歪分布ベースの  $q$ -BS 型のカーネル  $k_{(h)}^{\text{BS}(q)}(s; \alpha, \beta, \theta)$  に付随するパラメータ  $(\alpha, \beta)$  を推定位置  $x \in \mathbb{R}_+$  に依存させて, ノンパラメトリックな (可変的な) 対称分布ベース  $q$ -MIG カーネル, 及び, 歪分布ベース  $q$ -BS カーネルに基づいた密度推定量族を提案し, それらの漸近的性質と数値実験の結果を報告した.

# 未知で等しい分散を持つ $p$ 個の非負な正規母平均の同時推定 (Simultaneous Estimation of $p$ Positive Normal Means with Common Unknown Variance)

目白大学 張 元宗, 慶應義塾大学 篠崎 信雄, Rutgers 大学 William E. Strawderman

1. はじめに  $X_1, \dots, X_p$  を互いに独立に正規分布  $N(\theta_i, \sigma^2), i = 1, \dots, p$  にしたがう確率変数とし,  $\theta_i \geq 0$  とする。等分散  $\sigma^2$  が未知で、標準化 2 乗誤差損失関数を基準としたとき、母平均  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$  の同時推定問題を考える。

分散が既知で、1次元の場合、Katz(1961) は  $\theta$  の事前分布を非負な区間の上の一様分布とすると、 $\theta$  の一般化 Bayes 推定量

$$\delta(X) = X + \sigma\psi(X/\sigma) \quad (1.1)$$

を提案し、 $\delta(X)$  は  $\theta$  のミニマックス推定量であり、許容的な推定量でもあることを示した。ここで、

$$\psi(x) = \frac{e^{-x^2/2}}{\int_{-\infty}^x e^{-t^2/2} dt}, \quad \psi(0) = \sqrt{2/\pi},$$

である。全ての  $-\infty < x < \infty$  に対して、 $\delta(x), \psi(x)$  の性質は

- a)  $x + \psi(x) \geq 0$ ,
- b)  $\psi'(x) = -\psi(x)(x + \psi(x)) \leq 0$ ,
- c)  $x + \psi(x)$  は  $x$  の増加関数である。

である。

$p$  次元の場合に対しても  $\boldsymbol{\delta}(\mathbf{X}) = (\delta(X_1), \dots, \delta(X_p))$  は  $\boldsymbol{\theta}$  の一般化 Bayes 推定量である。しかし、 $\boldsymbol{\theta}$  を同時推定するとき、標準化 2 乗誤差損失関数の下で、張 (1982) は  $\boldsymbol{\delta}(\mathbf{X})$  を改良する縮小推定量

$$\delta_i^c(\mathbf{X}) = \left(1 - \frac{c\sigma^2}{\|\boldsymbol{\delta}(\mathbf{X})\|^2}\right) \delta(X_i) \quad (1.2)$$

を提案し、 $\boldsymbol{\delta}^c(\mathbf{X}) = (\delta_1^c(\mathbf{X}), \dots, \delta_p^c(\mathbf{X}))$  は  $\boldsymbol{\delta}(\mathbf{X})$  よりよくなるための十分条件  $p \geq 3, 0 \leq c \leq 2(p-2)$  を与えた。ここで  $\|\boldsymbol{\delta}(\mathbf{X})\|^2 = \sum_{i=1}^p \delta^2(X_i)$  である。

本研究の目的は分散  $\sigma^2$  が未知の場合、分散が既知の場合の推定量 (1.1) に分散の推定量を代入して得られる推定量

$$\delta(X_i, S) = X_i + a\sqrt{S}\psi(X_i/(a\sqrt{S}))$$

に基づく縮小推定量

$$\delta_i^c(\mathbf{X}, S) = \left(1 - \frac{cS}{\|\boldsymbol{\delta}(\mathbf{X}, S)\|^2}\right) \delta(X_i, S), \quad (1.3)$$

を提案し、 $\delta(X_i, S), i = 1, \dots, p$  を改良するための十分条件を与える。ここで、 $S \sim \sigma^2 \chi_n^2$  は  $X_i$  と互いに独立とする。さらに、 $\delta_i^c(\mathbf{X}, S)$  を含む 2 つの Baranchik タイプ縮小推定量のクラスを提案する。証明するため、下記の予備定理が有用である。

予備定理 2.1 全ての  $-\infty < X < \infty$  に対して、

- a)  $S\psi(X/S)$  は  $S$  の増加関数である。
- b)  $\delta(X, S)$  は  $S$  の増加関数である。(よって、 $\|\boldsymbol{\delta}(\mathbf{X}, S)\|^2$  は  $S$  の増加関数である)。
- c)  $-S\psi'(X/(a\sqrt{S}))$  は  $\sqrt{S}$  の増加関数である。
- d)  $S^{-1}(X + \sqrt{S}\psi(X/\sqrt{S}))$  は  $\sqrt{S}$  の減少関数である。(よって、 $\|\boldsymbol{\delta}(\mathbf{X}, S)\|^2/S^2$  は  $\sqrt{S}$  の減少関数である。)

が成立する。

予備定理 2.2 ( $\chi^2$  identity)

$S \sim \sigma^2 \chi_n^2$  とする。

$$E[h(S)S\sigma^2] = \frac{E[S^2h(S)] - 2\sigma^2E[S^2h'(S)]}{n+2}.$$

が成立する。

予備定理 2.3  $g(s)$  は  $s_0$  で負の値から正の値に変わって、 $h(s)$  は非減少関数であるとする

$$E[g(S)h(S)] \geq E[g(S)]h(s_0).$$

が成立する。

## 2. 結果

定理 2.1. 標準化 2 乗誤差損失のもとで、 $\delta^c(\mathbf{X}, S)$  が  $\delta(\mathbf{X}, S)$  を改良するための十分条件は

$$p \geq 3, 0 < c < 2\frac{p-2}{n+2}, a^2 \geq \frac{1}{n-4}, n \geq 5.$$

上の結果を下記のような 2 つの Baranchik タイプ推定量のクラスに拡張する。  
タイプ 1 :

$$F = \|\delta(\mathbf{X}, S)\|^2/S = a^2 \sum_{i=1}^p \left( \frac{X_i}{a\sqrt{S}} + \psi\left(\frac{X_i}{a\sqrt{S}}\right) \right)^2 \quad \text{とし、}$$

$$\delta_i^B(\mathbf{X}, S) = \left( 1 - \frac{r(F/S)}{F} \right) \delta(X_i, S)$$

とする。次の定理が得られる。

定理 2.2.  $\delta^B(\mathbf{X}, S)$  が  $\delta(\mathbf{X}, S)$  を改良するための十分条件は下記のようなものである。

- i)  $r(F/S)$  は  $F/S$  の単調非減少関数
- ii)  $r(F/S)/(F/S)$  は  $F/S$  の単調非増加関数
- iii)  $0 \leq r(F/S) \leq 2\frac{p-2}{n+2}, a^2 \geq 1/(n-4), n \geq 5.$

標本の大きさに対する条件を改良するため、次のような推定量のクラスを提案する。  
タイプ 2 :

$$\tilde{F} = a^2 \sum_{i=1}^p \max \left\{ \frac{X_i}{a\sqrt{S}} + \psi\left(\frac{X_i}{a\sqrt{S}}\right), \sqrt{2/\pi} \right\}^2 \geq F$$

を定義し、

$$\delta_i^{B\tilde{F}}(\mathbf{X}, S) = \left( 1 - \frac{r(\tilde{F})}{\tilde{F}} \right) \delta(X_i, S)$$

とする。次の定理が得られる。

定理 2.3.  $\delta^{B\tilde{F}}(\mathbf{X}, S)$  が  $\delta(\mathbf{X}, S)$  を改良するための十分条件は下記のようなものである。

- i')  $r(\tilde{F})$  は  $\tilde{F}$  の単調非減少関数
- ii')  $r(\tilde{F})/\tilde{F}$  は  $\tilde{F}$  の単調非増加関数
- iii')  $0 \leq r(\tilde{F}) \leq 2\frac{p-2}{n+2}, a^2 \geq 1/(n-2), n \geq 3.$

## 参考文献

- (1) Baranchik, A. J., 1970. A Family of Minimax Estimation of the Mean of a Multivariate Normal Problems. *Annals of Math. Stat.*, Vol.41, pp. 642-645.
- (2) Chang Y.-T., 1982. Stein-Type Estimators for Parameters in Truncated Space. *Keio Science and Technology Reports*, Vol. 35, No.11, pp. 185-193.
- (3) Birnbaum, Z.W., 1942. An Inequality for Mill's Ratio. *Ann. Math. Stat.* Vol. 13, pp. 245-246.
- (4) Gordon, R.D., 1941. Values of Mill's Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument, *Annals of Math. Stat.*, Vol.12, pp. 364-366.
- (5) Katz, M. W., 1961. Admissible and Minimax of Parameter in Truncated Spaces. *Annals of Math. Stat.*, Vol. 32, pp 136-1142.
- (6) Sampford, M. R., 1953. Some Inequalities on Mill's Ratio and Related Functions. *Annals of Math. Stat.*, Vol. 24, pp 130-132.
- (7) Stein, C. M., 1981. Estimation of the Mean of Multivariate Normal Distribution. *The Annals of Statistics*, Vol. 9, No.6, pp. 1135-1151.

## 方向の piece-wise regression model とその魚群データへの応用

島谷健一郎 (統計数理研究所)

### **Circular piece-wise regression model and its application to movement patterns for school of fishes**

Ichiro Ken Shimatani (The Institute of Statistical Mathematics)

When we investigate relationships between two variables, drawing a scattering diagram is a common first step, followed by an application of a regression model. When variables are both circular, this first step does not always give us good visualization, which facilitates the importance of the second step. Currently, couple of regression models are proposed and have been applied to circular-circular data. Among them, Downs and Mardia (2002) proposed the following model;

$$y = \alpha + \beta + 2 \tan^{-1}\{w \tan((x - \alpha)/2)\} + e, \quad (1)$$

where  $x$  and  $y$  are circular variables ( $-\pi \leq x, y < \pi$ ),  $w$  ( $-1 \leq w \leq 1$ ) is the regression coefficient,  $\beta$  is the intercept (practically, rotation), and  $\alpha$  is the “focal direction” ( $-\pi \leq \alpha, \beta < \pi$ ) (because directions do not have a specific origin,  $\alpha$  is included for playing a role of an origin).  $e$  is a random variable from a circular probability distribution such as the von Mises and the wrapped Cauchy distribution

This model is an example of a generalized linear model (GLM), where  $\tan((x - \alpha)/2)$  is used as a link function for transforming circular variables to linear variables. In fact, substituting  $X = \tan((x - \alpha)/2)$  and  $Y = \tan((y - \alpha)/2)$  into the linear regression equation  $Y = wX$ , we obtained  $\tan((y - \alpha)/2) = w \tan((x - \alpha)/2)$ , which is equivalent to equation (1). A difference from common GLMs is that the circular-circular regression uses the link function twice because both independent and dependent variables are circular. Consequently, dependent variables tend to be attracted to  $\alpha + \beta$ , and the attraction becomes stronger if  $w$  is close to 0. If  $w = 1$ , (1) reduces to

$$y = x + \beta + e, \quad (2)$$

which expresses a rotation of angle  $\beta$ .

Occasionally, circular-circular data exhibit more complex, irregular patterns. Here, we focus on when a basic pattern is just a rotation, but  $(x, y)$ s in some parts do not follow it (even including stochastic uncertainty), and show concentration on each specific angles.

One method for producing a well-fitting model is to connect several regression models continuously at the border (piece-wise regression). For circular-circular regression, however, because of its periodicity, it is not so straightforward.

In this study, I propose the circular-circular regression model that is defined as a weighted sum of model (1). The weights are given by functions on a circle that have high peaks in each

sepcific parts.

The expected direction can be expressed by their weighted average over circular variables (the argument of a weighted sum over unit vectors)

$$y = \arg\left(\sum_{j=0}^J g_j(x) f_j(x)\right) + e, \quad (3)$$

where  $f_j(x) = \alpha_j + \beta_j + 2a \tan(w_j \tan(\frac{x-\alpha_j}{2}))$  ( $j = 1, \dots, J$ ),  $g_j(x)$  is a probability density function of the von Mises or wrapped Cauchy distribution (subtracted by the minimum so that  $\min(g_j(x)) = 0$ ); high around  $\alpha_j$  and low (almost = 0) when distant from  $\alpha_j$ . In addition, we prepare a basic rotation;  $f_0(x) = x - \beta_0$  (rotation) and  $g_0(x) = c_0$  (constant). Dependent variable  $y$  is deformed toward  $\alpha_j + \beta_j$  when  $x$  is around  $\alpha_j$ , and the degree of deformation is strong if  $w_j$  is close to 0. If  $x$  is distant from any of  $\alpha_j$ , it is closed to the rotation of  $\beta_0$ .

The unknown parameters are ( $C_x, C_y, \rho$  in Eq. (2),  $c_0, \{\alpha_j, \beta_j, \rho_j, w_j\}$ ).

The model was formulated under a Bayesian model (non-informative prior distributions were used), and we estimated the unknown parameters by Markov chain Monte Carlo (MCMC) method.

The integer  $J$ , the number of components, expresses how many abnormal parts there were in a swarm. Because model (3) is a kind of “mixture model”, and it is well known that for such a “singular model”, commonly used information criterion such as AIC (Akaike information criterion) is not applicable. This study used the recently proposed information criterion, widely applicable Bayesian information criterion (WBIC, Watanabe (2013)), which is for Bayesian models and applicable to singular models.

What is important is that a model and parameter estimates should have interpretations associating with given data and the background. The proposed model is for analyzing spatio-temporal data about a rotating school of fishes in a aquarium, which was produced and digitalized by Terayama et al. (2015). I have demonstrated how the model and its parameter estimates are interpreted from zoological viewpoints.

T.D. Downs and K.V. Mardia. (2002). Circular Regression. *Biometrika*, Vol. 89, pp. 683-697.

K. Terayama, H. Hioki, and M. Sakagami. (2015). *Int. J. Semantic Computing*, Vol 9, pp.143-168.

S. Watanabe. (2013). A Widely Applicable Bayesian Information Criterion. *J. Machine Learning Research*, Vol 14, pp. 867-897.

# Asymptotic properties of classification procedures based on eigenstructures in high-dimensional context

Aki Ishii

Department of Information Sciences, Tokyo University of Science, Chiba, Japan

## 1 Introduction

Suppose we have two classes  $\pi_i$ ,  $i = 1, 2$ , and define independent  $p \times n_i$  data matrices,  $\mathbf{X}_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}]$ ,  $i = 1, 2$ , from  $\pi_i$ ,  $i = 1, 2$ , where  $\mathbf{x}_{ij}$ ,  $j = 1, \dots, n_i$ , are independent and identically distributed (i.i.d.) as a  $p$ -dimensional distribution with a mean vector  $\boldsymbol{\mu}_i$  and covariance matrix  $\boldsymbol{\Sigma}_i (\geq \mathbf{O})$ . We assume  $n_i \geq 3$ ,  $i = 1, 2$ . The eigen-decomposition of  $\boldsymbol{\Sigma}_i$  is given by

$$\boldsymbol{\Sigma}_i = \mathbf{H}_i \boldsymbol{\Lambda}_i \mathbf{H}_i^T = \sum_{s=1}^p \lambda_{s(i)} \mathbf{h}_{s(i)} \mathbf{h}_{s(i)}^T,$$

where  $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{1(i)}, \dots, \lambda_{p(i)})$  having  $\lambda_{1(i)} \geq \dots \geq \lambda_{p(i)} (\geq 0)$  and  $\mathbf{H}_i = [\mathbf{h}_{1(i)}, \dots, \mathbf{h}_{p(i)}]$  is an orthogonal matrix of the corresponding eigenvectors. Let  $\mathbf{X}_i - [\boldsymbol{\mu}_i, \dots, \boldsymbol{\mu}_i] = \mathbf{H}_i \boldsymbol{\Lambda}_i^{1/2} \mathbf{Z}_i$  for  $i = 1, 2$ . Then,  $\mathbf{Z}_i$  is a  $p \times n_i$  sphered data matrix from a distribution with the zero mean and identity covariance matrix. Let  $\mathbf{Z}_i = [\mathbf{z}_{1(i)}, \dots, \mathbf{z}_{p(i)}]^T$  and  $\mathbf{z}_{j(i)} = (z_{j1(i)}, \dots, z_{jn_i(i)})^T$ ,  $j = 1, \dots, p$ , for  $i = 1, 2$ . Note that  $E(z_{jk(i)} z_{j'k(i)}) = 0$  ( $j \neq j'$ ) and  $\text{Var}(\mathbf{z}_{j(i)}) = \mathbf{I}_{n_i}$ , where  $\mathbf{I}_{n_i}$  denotes the  $n_i$ -dimensional identity matrix. Also, note that if  $\mathbf{X}_i$  is Gaussian,  $z_{jk(i)}$ s are i.i.d. as the standard normal distribution,  $N(0, 1)$ . We assume that the fourth moments of each variable in  $\mathbf{Z}_i$  are uniformly bounded for  $i = 1, 2$ . Let  $\mathbf{z}_{oj(i)} = \mathbf{z}_{j(i)} - (\bar{z}_{j(i)}, \dots, \bar{z}_{j(i)})^T$ ,  $j = 1, \dots, p$ ;  $i = 1, 2$ , where  $\bar{z}_{j(i)} = n_i^{-1} \sum_{k=1}^{n_i} z_{jk(i)}$ . We assume that  $P(\lim_{p \rightarrow \infty} \|\mathbf{z}_{o1(i)}\| \neq 0) = 1$  for  $i = 1, 2$ , where  $\|\cdot\|$  denotes the Euclidean norm.

Let  $\mathbf{x}_0$  be an observation vector of an individual belonging to  $\pi_i$  ( $i = 1, 2$ ). We assume  $\mathbf{x}_0$  and  $\mathbf{x}_{ij}$ s are independent. We estimate  $\boldsymbol{\mu}_i$  and  $\boldsymbol{\Sigma}_i$  by  $\bar{\mathbf{x}}_{in_i} = \sum_{j=1}^{n_i} \mathbf{x}_{ij} / n_i$  and  $\mathbf{S}_{in_i} = \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_{in_i})(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{in_i})^T / (n_i - 1)$ . A typical classification rule is that one classifies an individual into  $\pi_1$  if

$$\begin{aligned} & (\mathbf{x}_0 - \bar{\mathbf{x}}_{1n_1})^T \mathbf{S}_{1n_1}^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_{1n_1}) - \log \left\{ \frac{\det(\mathbf{S}_{2n_2})}{\det(\mathbf{S}_{1n_1})} \right\} \\ & < (\mathbf{x}_0 - \bar{\mathbf{x}}_{2n_2})^T \mathbf{S}_{2n_2}^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_{2n_2}), \end{aligned} \quad (1.1)$$

and into  $\pi_2$  otherwise. However, the inverse matrix of  $\mathbf{S}_{in_i}$  does not exist in the HDLSS context ( $p > n_i$ ).

Aoshima and Yata (2011, 2015) considered substituting  $\{\text{tr}(\mathbf{S}_{in_i})/p\} \mathbf{I}_p$  for  $\mathbf{S}_{in_i}$  by using the difference of a geometric representation of HDLSS data from each  $\pi_i$ . Aoshima and Yata (2015) gave the misclassification rate adjusted classifier for multiclass, high-dimensional data whose misclassification rates are no more than specified thresholds under the following condition for eigenvalues:

$$\frac{\lambda_{1(i)}^2}{\text{tr}(\boldsymbol{\Sigma}_i^2)} \rightarrow 0 \text{ as } p \rightarrow \infty \text{ for } i = 1, 2. \quad (1.2)$$

Recently, Aoshima and Yata (2017) considered the ‘‘strongly spiked eigenvalue (SSE) model’’ as follows:

$$\liminf_{p \rightarrow \infty} \left\{ \frac{\lambda_{1(i)}^2}{\text{tr}(\boldsymbol{\Sigma}_i^2)} \right\} > 0 \text{ for } i = 1 \text{ or } 2. \quad (1.3)$$

On the other hand, Aoshima and Yata (2017) called (1.2) the ‘‘non-strongly spiked eigenvalue (NSSE) model’’.

## 2 A new classification under the SSE model

As for the two-sample test, Ishii (2017a, 2017b) gave new test procedures for high dimensional data under the SSE model. In this talk, we considered the distance-based classifier under one of the SSE models. Aoshima and Yata (2014) considered a classification rule given by using the identity matrix  $\mathbf{I}_p$  instead of  $\mathbf{S}_{in_i}$  in (1.1). They showed the asymptotic normality of the classifier and provide a sample size determination so as to control misclassification rates being no more than a prespecified value. They further developed the classifier to multiclass classification. In order to remove the bias, we considered a data transformation and gave the following new classification rule. One classifies an individual into  $\pi_1$  if

$$\left(\mathbf{x}_0 - \frac{\bar{\mathbf{x}}_{1n_1} + \bar{\mathbf{x}}_{2n_2}}{2}\right)^T \mathbf{A}(\bar{\mathbf{x}}_{2n_2} - \bar{\mathbf{x}}_{1n_1}) + \sum_{i=1}^2 (-1)^i \frac{\text{tr}(\mathbf{S}_{in_i} \mathbf{A})}{2n_i} < 0 \quad (2.1)$$

and into  $\pi_2$  otherwise. Here, the second term of (2.1) is a bias-correction term. If  $\mathbf{h}_{1(1)} = \mathbf{h}_{1(2)} (= \mathbf{h}_1, \text{ say})$ , we gave  $\mathbf{A} = \mathbf{I}_p - \mathbf{h}_1 \mathbf{h}_1^T$ . We used the noise-reduction (NR) methodology given by Yata and Aoshima (2012) to estimate the first eigenvalues and eigenvectors. By using the asymptotic properties of the NR estimator given by Ishii, Yata and Aoshima (2016), we gave the asymptotic properties of our new classification rule. We also compared classification procedures by using computer simulations.

## References

- [1] Aoshima, M. and Yata, K. (2011). Two-Stage Procedures for High-Dimensional Data. *Sequential Analysis* (Editor's special invited paper) 30: 356-399.
- [2] Aoshima, M., Yata, K. (2014). A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. *Annals of the Institute of Statistical Mathematics*, 66, 983–1010.
- [3] Aoshima, M. and Yata, K. (2015). Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. *Methodology and Computing in Applied Probability* 17: 419-439.
- [4] Aoshima, M. and Yata, K. (2017). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica*, to appear.
- [5] Ishii, A., Yata, K., and Aoshima, M. (2016). Asymptotic properties of the first principal component and equality tests of covariance matrices in high-dimension, low-sample-size context. *Journal of Statistical Planning and Inference*, 170: 186-199.
- [6] Ishii, A. (2017a). High-dimensional two-sample tests for non-gaussian strongly spiked eigenvalue models, to appear in *Journal of the Japan Statistical Society*.
- [7] Ishii, A. (2017b). A two-sample test for high-dimension, low-sample-size data under the strongly spiked eigenvalue model, *Hiroshima Mathematical Journal*, 47, in print.
- [8] Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.* 105, 193-215.

# Asymptotic Properties of Mildly Explosive Processes with Locally Stationary Disturbance

Junichi Hirukawa and Sangyeol Lee  
Niigata University and Seoul National University

**1. Introduction** Autoregressive processes of the form  $y_t = \rho y_{t-1} + \varepsilon_t$  with an explosive root  $|\rho| > 1$ , where  $\varepsilon_t$  are iid standard normal random variables, were first studied in White (1958) and Anderson (1959). By assuming a zero initial value for  $y_t$ , a Cauchy limit theory is derived for the least squares estimate  $\widehat{\rho}_n = (\sum_{t=1}^n y_{t-1} y_t) (\sum_{t=1}^n y_{t-1}^2)^{-1}$ :

$$(1) \quad \frac{\rho^n}{\rho^2 - 1} (\widehat{\rho}_n - \rho) \Rightarrow C,$$

where  $C$  denotes a Cauchy random variable and  $\Rightarrow$  denotes convergence in distribution when  $n$  tends to infinity. The Gaussian assumption imposed on the innovation sequence  $\{\varepsilon_t\}$  plays an important role and cannot be relaxed to obtain the same asymptotic distribution as in (1): see Anderson (1959) who demonstrate that the limit distribution of the least squares estimate depends upon the distributional assumptions imposed on the error sequence.

However, this difficulty can be avoided when the explosive root approaches unity as the sample size tends to infinity. Phillips and Magdalinos (2007a) consider autoregressive processes with root  $\rho_n = 1 + c/k_n$ , where  $k_n$  is a positive real sequence with  $k_n = o(n)$ . When  $c > 0$ , such roots are explosive in finite samples and approach unity at the rate slower than  $O(n^{-1})$ . It is well known that the asymptotic behavior of such mildly explosive autoregressions is more uniform than their purely explosive counterparts. Under the second moment condition on the iid innovations, Phillips and Magdalinos (2007a) establish limit theorems for sample moments generated by mildly explosive processes and obtain the following Cauchy limit result:

$$(2) \quad \frac{1}{2c} k_n \rho_n^n (\widehat{\rho}_n - \rho_n) \Rightarrow C.$$

This limit result is unaffected by both the distribution of the initial condition  $y_0$  as far as  $y_0 = o(k_n)$ . The result was extended by Phillips and Magdalinos (2007b) to a class of weakly dependent innovations. Aue and Horváth (2007) relax the moment conditions on the innovations by considering an iid innovation sequence that belongs to the domain of attraction of a stable law. Magdalinos and Phillips (2008) give multivariate extensions and Magdalinos (2012) considers mildly explosive autoregressions generated by a linear process that may exhibit long-range dependence. Oh et al. (2017) recently study mildly explosive autoregressions with strong mixing innovation sequence, showing that the least squares estimate has the same limit distribution as the iid innovation case.

However, the time homogenous assumption on the residuals seems to be restrictive. The analysis of relatively long stretches of time series data that may contain either slow or rapid changes in the spectrum is of interest in a number of areas. Although the idea of having locally approximately a stationary process was also the starting point of Priestley's theory of processes with evolutionary spectra (Priestley (1965)), recently one of the most important classes of non-stationary processes has been formulated in a rigorous asymptotic framework by Dahlhaus (1996a, 1996b, 1996c, 1997), called locally stationary processes. Locally stationary processes have time varying spectral densities whose spectral structures smoothly change in time. Dahlhaus (2012) also gave the extensive review about locally stationary processes. In this study, we investigate the asymptotic distribution of the LSE for the autoregression with locally stationary error process. The limit behavior of the LSE from the unit root and near stationary autoregressions with locally stationary disturbance was considered by Hirukawa and Skdakata (2012). For the iid unit root process case, see Chan and Wei (1980) and Lee and Wei (1999).



Although we mainly focus on the limit behavior of the LSE, we develop a method for identifying the onset and the end of a bubble period of an econometric time series as an application, originally considered in Phillips and Yu (2009) and Phillips et al. (2011). For this task, we investigate the limiting distribution of the Dickey-Fuller tests when the underlying process is either a unit-root process or explosively mild process and finally demonstrate that they are consistent.

**2. Limit distribution of LSE** In this section, we consider the following mildly explosive process with locally stationary disturbance:

$$\begin{aligned} y_{t,n} &= \rho_{t,n} y_{t-1,n} + u_{t,n}, \quad t = 1, \dots, n, \\ &= \left( \prod_{k=1}^t \rho_{k,n} \right) y_{0,n} + \sum_{j=1}^t \left( \prod_{k=j+1}^t \rho_{k,n} \right) u_{j,n}, \end{aligned}$$

where  $\{u_{j,n}\}$  is generated from the time varying MA ( $\infty$ ) model:

$$u_{t,n} = \sum_{l=0}^{\infty} \alpha_l \left( \frac{t}{n} \right) \varepsilon_{t-l} = \sum_{l=0}^{\infty} \alpha_l \left( \frac{t}{n} \right) L^l \varepsilon_t := \alpha \left( \frac{t}{n}, L \right) \varepsilon_t,$$

where  $\varepsilon_t \sim i.i.d. (0, \sigma^2)$ ,  $\rho_{t,n} = 1 + \frac{1}{k_n} \beta \left( \frac{t}{n} \right)$  with  $\beta \in C[0, 1]$ , the class of continuous real-valued functions on  $[0, 1]$ , satisfying  $0 < \beta(u) < \infty$ , the MA coefficients  $\alpha \left( \frac{t}{n}, L \right) := \sum_{l=0}^{\infty} \alpha_l \left( \frac{t}{n} \right) L^l$  with lag operator  $L$ , satisfying

$$\sum_{l=0}^{\infty} l \sup_{0 \leq u \leq 1} |\alpha_l(u)| < \infty \quad \text{and} \quad \sum_{l=0}^{\infty} l \sup_{0 \leq u \leq 1} \left| \frac{\partial}{\partial u} \alpha_l(u) \right| < \infty,$$

and  $y_{0,n} = o_P(\sqrt{k_n})$  with  $k_n = o(n)$  being a sequence of positive real numbers.

We then consider the normalized serial correlation coefficient:

$$(3) \quad S_n := \frac{\frac{1}{k_n \left( \prod_{k=1}^n \rho_{k,n} \right)} \sum_{t=1}^n y_{t-1,n} u_{t,n}}{\frac{1}{k_n^2 \left( \prod_{k=1}^n \rho_{k,n} \right)^2} \sum_{t=1}^n y_{t-1,n}^2} := \frac{U_n}{V_n}.$$

Note that if  $\rho_{t,n} \equiv \rho_n = 1 + \frac{\beta}{k_n}$ ,  $t = 1, \dots, n$ , we have

$$(4) \quad S_n = k_n \left( \prod_{k=1}^n \rho_{k,n} \right) (\widehat{\rho}_n - \rho_n),$$

where  $\widehat{\rho}_n := \frac{\sum_{t=1}^n y_{t-1,n} y_{t,n}}{\sum_{t=1}^n y_{t-1,n}^2}$  is the least squares estimator (LSE) of the AR(1) process with constant coefficient:

$$y_{t,n} = \rho_n y_{t-1,n} + u_{t,n}, \quad t = 1, \dots, n, \quad \rho_n = 1 + \frac{\beta}{k_n}.$$

Therefore, the  $\widehat{\rho}_n$  has bias for the estimation of  $\rho_{t,n}$ . We then obtain the following theorem.

**Theorem 1.** Let  $\kappa = 2 \sqrt{\beta(0)\beta(1)} \left| \frac{\alpha(1,1)}{\alpha(0,1)} \right|$ . Then,

$$S_n \Rightarrow \kappa C.$$

Therefore, if  $\rho_{t,n} \equiv \rho_n = 1 + \frac{\beta}{k_n}$ ,  $t = 1, \dots, n$ ,

$$k_n \left( \prod_{k=1}^n \rho_{k,n} \right) (\widehat{\rho}_n - \rho_n) \Rightarrow \kappa C$$

from (4).

# Activation function selection in Recurrent Neural Network

Daichi Inoue\*

Takeshi Kato†

## 1 Introduction

The problem of solving the inverse of a time-varying matrix arises in many fields, science, engineering and business, for instance. Even when the matrix is time-invariant, however, it is well known that the actual computation is not easy if the matrix size is large or sparse. To say nothing of time-varying matrices.

A new approach by applying recurrent neural network was proposed by Zhang *et al* [1] to compute the inverse of a time-varying matrix, and a series of research works has been made in several papers such as [3], [4], and [5]. Recurrent neural network (RNN) is one type of neural networks. The advantage of RNN is that it can handle continuous information such as a sentence, so that it has originally applied to natural language processing and shown high performance. The research works by Zhang and others, [1], [2], [3], [4], and [5], make use of what it calls "super-sigmoid function" as a good activation function in RNN, but the mathematically rigorous argument has been skipped in those papers. Then the aim of our research work is to clarify how to select good activation functions of RNN in theory and propose a new activation function with higher performance than the super-sigmoid function. The argument is made in view of the convergence of error terms to zero and robustness caused by the differentiation error and the model-implementation. Numerical examples are also given in addition to theories.

## 2 The summary of previous studies

### 2.1 Problem formulation

We consider computing the inverse of a smooth time-varying matrix  $A(t) \in R^{n \times n}$ . In other words, we will find  $X(t) \in R^{n \times n}$  satisfying the equation

$$A(t)X(t) - I = 0, \quad t \in [0, \infty) \quad (1)$$

where  $I \in R^{n \times n}$  is the identity matrix.

### 2.2 General RNN Model

We define a matrix-valued error function by

$$E(X(t), t) = (e_{ij}(t)) := A(t)X(t) - I$$

to monitor the matrix-inversion process. The error function derivative  $\dot{E}(X(t), t)$  with respect to  $t$  should be made such that every entry  $e_{ij}(t)$  of  $E(X(t), t)$  converges to zero. Specifically,  $E(X(t), t)$  can be described in the general form

$$\frac{dE(X(t), t)}{dt} = -\Gamma \mathcal{F}(E(X(t), t)) \quad (2)$$

where  $\mathcal{F}(\cdot) : R^{n \times n} \rightarrow R^{n \times n}$  denotes a matrix mapping of neural networks, and the matrix  $\Gamma$  consisting of positive elements scales the convergence rate of the solution. Each element of  $\mathcal{F} = (f_{ij})$  is called an activation function. The RNN design formula (2) leads to the following implicit dynamic equation of the generalized neural model

$$A(t)\dot{X}(t) = -\dot{A}(t)X(t) - \Gamma \mathcal{F}(A(t)X(t) - I) \quad (3)$$

where  $X(t)$ , starting from an initial condition  $X(0) = X_0 \in R^{n \times n}$ , is the activation state matrix corresponding to the theoretical solution  $X^*(t)$  of (1).

\*Graduate School of Science and Technology, Sophia University (Master's Program)

†Faculty of Science and Technology, Sophia University

### 3 Comparison of activation functions

We assume below that  $\Gamma = \gamma I$  for some  $\gamma > 0$  and  $f_{ij} \equiv f$  in  $\mathcal{F} = (f_{ij})$  for simplicity. Zhang *et al.* [2] discusses the performance of several activation functions in RNN, and shows that the following power-sigmoid activation function  $f_1$  is the best among them in the sense of convergence of errors to zero and the robustness:

$$f_1(u) = \begin{cases} u^p & : |u| \geq 1 \\ \frac{e^\xi + e^{-\xi}}{e^\xi - e^{-\xi}} \tanh(\xi u) & : |u| < 1 \end{cases}$$

where  $\xi > 1/2$  is a real number and  $p \geq 3$  is an odd integer.

We propose a new activation function

$$f_2(u) = \begin{cases} \frac{1}{e^p - 1} \operatorname{sgn}(u)(e^{p|u|} - 1) & : |u| \geq 1 \\ \frac{e^\xi + e^{-\xi}}{e^\xi - e^{-\xi}} \tanh(\xi u) & : |u| < 1 \end{cases}$$

with  $\xi > 1$  and an odd integer  $p \geq 3$ .

Solving the differential equation (2) for  $f = f_1, f_2$  and noting the difference between the power function and the exponential function in  $|u| \geq 1$ , we can show that that  $f_2$  is slightly better than  $f_1$  in the sense of the decay of  $u = e_{ij}(t)$  for large  $t$ . For the perturbed version of (3) by the model-implementation and differentiation errors, it is also shown that  $f_2$  is more robust against those errors than  $f_1$ .

### 4 Conclusion

In this presentation, we formulated rigorously the structure of activation functions in RNN which have exponential decay of errors and robustness. Moreover, we propose a slightly better activation function than the super-sigmoid function, which is the best among several candidates discussed in Zhang *et al.* [2]. Recently, however, Zhang *et al.* [6] proposed a new approach to find good activation function in RNN. The argument by taking the new idea into account has not been made, and it is left for further research.

### References

- [1] Y. Zhang and S. S. Ge, "A general recurrent neural network model for time-varying matrix inversion," *42nd IEEE Conf. Decis. Cont.* pp. 6169-6174, 2003.
- [2] Y. Zhang and S. S. Ge, "Design and analysis of a general recurrent neural network model for time-varying matrix inversion," *IEEE Trans. Neural Netw.* vol. 16, no. 6, pp. 1477-1490, 2005.
- [3] D. Guo, Z. Nie and L. Yan, "Novel discrete-time Zhang neural network for time-varying matrix inversion," *IEEE Trans. Syst., Man., Cybern., Syst.*, vol. 47 no. 8, pp. 2301-2310, 2017.
- [4] Y. Zhang, G. Ruan, K. Li and Y. Yang, "Robustness analysis of the Zhang neural network for online time-varying quadratic optimization," *J. Phys. A: Math. Theor.*, vol. 43, pp. 245-202, 2010.
- [5] Y. Zhang, Y. Yang, N. Tan and B. Cai, "Zhang neural network solving for time-varying full-rank matrix Moore-Penrose inverse," *Springer Link Comput.* vol. 92, no. 2, pp. 97-121, 2011.
- [6] Y. Zhang, L. Jin and S. Li, "Integration-enhanced Zhang neural network for real-time-varying matrix inversion in the presence of various kinds of noises," *IEEE Trans. Neural Netw. Learn. Syst.* vol. 27, no. 12, pp. 2615-2627, 2016.

# On mode and antimode preserving circular distributions

Toshihiro Abe<sup>1</sup>, Yoichi Miyata<sup>2</sup> and Takayuki Shiohama<sup>3</sup>

<sup>1</sup> Nanzan University, Japan.

abetosh@ss.nanzan-u.ac.jp

<sup>2</sup> Takasaki City University of Economics, Japan.

y Miyatagbt@tcue.ac.jp

<sup>3</sup> Tokyo University of Science, Japan.

shiohama@rs.tus.ac.jp

Azzalini & Capitanio (2003) constructed an extension of the skew-normal distribution given by Azzalini (1985) as follows: Let  $f_0(x)$  ( $x \in \mathbb{R}$ ) be a symmetric pdf about  $x = 0$ . Let  $G$  be the distribution function such that  $G(-x) = 1 - G(x)$ . Then their perturbation of the symmetric pdf is given by

$$2f_0(x)G\{w(x)\}, \quad (1)$$

where the weighting function  $w : \mathbb{R} \rightarrow \mathbb{R}$  is taken to be an odd function, i.e.,  $w(-x) = -w(x)$  for  $x \in \mathbb{R}$ . It is noted that there exists an odd function  $h(x)$  ( $-1 \leq h(x) \leq 1$ ) such that  $2G(w(x)) = 1 + h(x)$ . Then, it is easily shown that

$$\int_{\mathbb{R}} (1 + h(x))f_0(x)dx = 1.$$

This point of view is seen in some literature (cf. Huang & Chen, 2007).

Their skew family of the symmetric distributions is not always unimodal even if the base symmetric distribution  $f_0(x)$  is unimodal. On the other hand, Jones (2014) proposed another member of skew distributions with pdf

$$f(x) = f_0(s^{-1}(x)), \quad x \in S_f, \quad (2)$$

where  $s'(x) + s'(-x) = 2$ , the function  $s : D \rightarrow S_f$  ( $D \supseteq S_{f_0} \ni 0$ ) is a monotone and bijection function and  $S_f$  is a support of the function  $s$ . A noteworthy point is that the resulting density  $f(x)$  is always unimodal regardless of the parameter values of the base symmetric unimodal density.

For a construction of the circular version of the skew distributions by Azzalini & Capitanio (2003), Umbach & Jammalamadaka (2009) adapted the pdf (1) in order to obtain a general means of skewing symmetric circular models. Let  $f_0(\theta)$  and  $g(\theta)$  ( $\theta \in [-\pi, \pi)$ ) be circular pdf's symmetric about  $\theta = 0$  and  $G(\theta) = \int_{-\pi}^{\theta} g(\phi)d\phi$  is the distribution function of the latter. The weighting function  $w$  is an odd periodic function with  $|w(\theta)| \leq \pi$ . Then

$$f(\theta) = 2f_0(\theta)G\{w(\theta)\}, \quad -\pi \leq \theta < \pi \quad (3)$$

is a circular pdf. As a tractable and simple case of the circular pdf (3), Abe & Pewsey (2011) also introduced the  $k$ -sine-skewed circular distributions (SSCD $_k$ ) with the density

$$f(\theta) = (1 + \lambda \sin k\theta)f_0(\theta) \quad -\pi \leq \theta < \pi,$$

where  $\lambda$  ( $-1 \leq \lambda \leq 1$ ) is the skewing parameter. For a related topic for  $\text{SSCD}_k$  with  $k = 1$ , Abe & Pewsey (2011) investigated its properties and proposed the sine-skewed Jones-Pewsey family as well as its three special cases. As they showed, if  $f_0(\theta)$  is the wrapped Cauchy density, the resulting density is always unimodal. Otherwise, the resulting density is not always unimodal.

Combining with (2) and (3), we obtain another family of unimodal skew distributions on the circle given by

$$f(\theta) = f_0(s^{-1}(\theta; w)), \quad \theta \in [-\pi, \pi), \quad (4)$$

which is an extension of the model in Jones & Pewsey (2012), where  $f_0$  is a symmetric unimodal density about  $\theta = 0$  and the function  $s$  is given by

$$s(\theta; w) = 2 \int_{-\pi}^{\theta} G(w(t)) dt - \pi. \quad (5)$$

It is not easy to get the explicit expressions for the moments of the family (4) in general, except for the Jones & Pewsey (2012)'s inverse Batschelet distributions (see Abe, 2015). The advantage of the family of the distributions is always unimodal. Note that this function is a bijection function from  $[-\pi, \pi)$  to  $[-\pi, \pi)$ , and can be rewritten as  $s(\theta) = \theta + H(\theta)$ , where  $H(\theta)$  is an even function with  $|H'(\theta)| \leq 1$ . Further, let us confine our attention to the function  $s$  which maps zero to zero. Then the distribution with pdf (4) has the mode-preserving property (cf. Fujisawa & Abe, 2015).

## References

- [1] Abe, T. & Pewsey, A. (2011). Sine-skewed circular distributions. *Statistical Papers*, **52**, 683–707.
- [2] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.
- [3] Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew  $t$ -distribution. *Journal of the Royal Statistical Society, Series B*, **65**, 367–389.
- [4] Fujisawa, H. & Abe, T. (2015). A family of skew distributions with mode-invariance through transformation of scale. *Statistical Methodology*, **25**, 89–98.
- [5] Jones, M.C. & Pewsey, A. (2012). Inverse Batschelet distributions for circular data. *Biometrics*, **68**, 183–193.
- [6] Umbach, D. & Jammalamadaka, S. R. (2009). Building asymmetry into circular distributions. *Statistics & Probability Letters*, **79**, 659–663.

# シンポジウム報告書 (タイトル: On estimating finite mixtures of skew-rotationally-symmetric distributions)

宮田 庸一\*, 塩濱 敬之†, 阿部 俊弘‡

## 1 はじめに

円周データ, 方向データにおいては, しばしば分布の形が多峰型になったり, モードに関して非対称となることがある. そのようなデータに対しては, 通常回転に対して対称 (rotationally symmetric, 以後単に対称という) な分布の有限混合モデルを用いることが多く, 様々な研究者 (例えば Holzmann, Munk, and Stratmann [3], Banerjee Dhillon, Ghosh and Sra [1]) によりアルゴリズムが提案され, 理論的な性質が明らかにされている. しかしながら, 一方でコンポーネントが非対称な分布の有限混合モデルにおいては, あまり研究がなされていない. このため, 本報告においては, 以下の sine-skewed wrapped Cauchy 分布の有限混合モデルを考え, その最尤推定量を求めるための EM アルゴリズムおよび漸近的な性質に関して報告を行った:

$$f(\theta|\gamma) = \sum_{j=1}^g \alpha_j f_{swc}(\theta|\mu_j, \rho_j, \lambda_j), \quad (1)$$

ただし  $g$  はコンポーネントの数を表し,  $\alpha_j$  ( $j = 1, \dots, g$ ) は混合比率,

$$f_{swc}(\theta|\mu, \rho, \lambda) = \frac{1 - \rho^2}{2\pi(1 + \rho^2 - 2\rho \cos(\theta - \mu))} \{1 + \lambda \sin(\theta - \mu)\} \quad (2)$$

は sine-skewed wrapped Cauchy 分布の確率密度関数を表す. また  $\gamma_j = (\mu_j, \rho_j, \lambda_j)^T$ ,  $\gamma = (\alpha_1, \dots, \alpha_g, \gamma_1^T, \dots, \gamma_g^T)^T$  をパラメーターベクトルとする.

## 2 EM アルゴリズム

大きさ  $n$  の観測値  $\Theta_1, \dots, \Theta_n$  が互いに独立に, sine-skewed wrapped Cauchy 分布の有限混合モデル (1) に従うとする. 我々の提案する EM アルゴリズムにおける E-step は, 通常の E-step と同じである. 一方で M-step

\*高崎経済大学経済学部. E-mail:ymiyatagbt@tcue.ac.jp

†東京理科大学工学部

‡南山大学理工学部

は通常の方法とは異なり,  $Q_j$  関数と呼ばれる  $j$  番目のコンポーネントのパラメーターの関数を最大化するために, それぞれのコンポーネントにおけるパラメーター空間の領域を 2 つに分割する. そして, それぞれの領域において  $Q_j(\gamma_j, \gamma^{(k-1)})$  を最大化する  $\gamma_j$  の値を求め, 大きい  $Q_j$  を持つ  $\gamma_j$  の値を  $\gamma_j^{(k)}$  として更新する方法であることを報告した. この修正した EM アルゴリズムと通常の EM アルゴリズムをシミュレーションにより比較したところ, 20 回中 8 回はアルゴリズムの収束において改善が見られ, 12 回は通常の方法と同じパフォーマンスであることを確認した. またこのアルゴリズムを実データ (風向データ) に対して適用した例を報告した.

### 3 一貫性

有限混合モデルにおいては, 通常の見別可能性が成り立たないため, 真のモデルを表すパラメーターの集合と, 最尤推定量をプラグインした密度関数の表すパラメーターの集合との”集合間の距離”において一貫性を示す必要がある (詳しくは例えば, Cheng and Liu [2] を参照のこと). 本報告においては, この一貫性を示すために必要となる, sine-skewed wrapped Cauchy 分布 (2) の見別可能性が成り立つことを報告した. さらにこの結果と Cheng and Liu [2] の定理を用いることで, 集合間の距離の下での最尤推定量の強一貫性が成り立つことを報告した.

### 4 質疑応答について

会場に寄せられたコメントとしては, 我々の提案する EM アルゴリズムの計算時間に関する質問および EM アルゴリズムの収束に関する質問がされた.

### 参考文献

- [1] Banerjee, A., Dhillon, I. S., Ghosh, J. P. and Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions, *Journal of Machine Learning Research*, 6, Sep, 1345–1382.
- [2] Cheng, R.C.H. and Liu, W.B. (2001). The consistency of estimators in finite mixture models, *Scandinavian Journal of Statistics*, 28, 4, 603–616.
- [3] Holzmann, H., Munk, A. and Stratmann, B. (2004). Identifiability of finite mixtures-with applications to circular distributions, *Sankhyā*, 440–449.

## BOOTSTRAP INFERENCE UNDER RANDOM DISTRIBUTIONAL LIMITS

Speaker: *Giuseppe Cavaliere*, University of Bologna  
(this is a joint work with Iliyan Georgiev, University of Bologna )

Asymptotic bootstrap validity is usually understood and established as consistency of the distribution of a bootstrap statistic, conditional on the data, for the unconditional limit distribution of a statistic of interest. In many applications, however, the bootstrap statistic may possess, conditionally on the data, a random limit distribution. Among others, cases of random bootstrap limit distributions are documented for infinite variance processes (Athreya, 1987; Knight, 1989; Aue et al., 2008; Cavaliere et al., 2016), time series with unit roots (Basawa et al., 1991; Cavaliere et al., 2015), and parameters on the boundary of the parameter space (Andrews, 2000). In most cases, the occurrence of a random limit distribution for a bootstrap statistic given the data – in contrast to a necessarily non-random limit of the unconditional distribution of the corresponding statistic, computed on the original sample – is taken as evidence of failure of the bootstrap.

In this paper we show that randomness in the limiting distribution of a bootstrap statistic need not invalidate bootstrap inference, as the bootstrap may still deliver confidence intervals (or hypothesis tests) with the desired coverage probability (or size) when the sample size diverges. Moreover, in such cases the bootstrap may also have the appealing asymptotic interpretation of a conditional inferential procedure, and may deliver efficiency (or power) gains over unconditional inference.

To see why, it is useful to recall that, apart from an unconditional limit distribution, a statistic in general possesses a family of (random) conditional limit distributions, depending on the choice of the conditioning  $\sigma$ -algebra. If one of these conditional (random) limit distributions matches the (random) limit distributions of a chosen bootstrap statistic, then – under regularity conditions that will be discussed in the paper – inference based on the bootstrap is asymptotically valid and, importantly, conditional in nature. This observation was initially made by Lepage and Podgorski (1996) but has not been pursued further in the bootstrap literature, in particular because its development requires probabilistic tools that are not widely popular in this field.

Conditional inference can be justified by the ‘conditionality principle’, according to which “the evidential meaning of any outcome of any mixture experiment is the same as that of the corresponding outcome of the corresponding component experiment, ignoring the overall structure of the mixture experiment” (Birnbaum, 1962, p.271). Whenever for a statistic of interest the bootstrap estimates consistently a component of the limit unconditional distribution



viewed as a mixture of conditional distributions, the bootstrap can be regarded as a large-sample implementation of the conditionality principle.

In such cases the bootstrap replicates asymptotically the property of conditional tests and confidence intervals to have conditionally constant size and coverage probability, respectively. Regarding tests, this property has been argued to be necessary for test optimality in the special case of conditioning on a complete sufficient statistic; see Lockhart (2012). More generally, gains in power and precision can be expected to occur when the reference population is effectively restricted to outcomes that share statistically relevant features with the actual sample. For instance, in the case of confidence intervals, Lepage and Podgorski (1996, Figure 2) provide numerical evidence of substantial precision gains in a particular implementation of a permutation bootstrap with a random limit of the bootstrap statistic, where conditioning is on the order statistics of regression residuals (ancillary in that context).

Following a practice in the literature (see, e.g., Lockhart, 2012), we recast the constant conditional size and coverage probability property into the requirement that bootstrap p-values should be uniformly distributed conditionally, at least asymptotically. One of our main results is a general sufficient condition for this to be the case. We also provide conditions for the more basic property of unconditional asymptotic distributional uniformity of p-values; this property implies asymptotic control of the frequency of wrong inferences on average over the conditioning variables but no longer warrants a conditional interpretation of the bootstrap inferential procedure.

When dealing with random limiting distributions, the usual convergence concept employed to establish bootstrap validity, i.e. weak convergence in probability, can only be employed in some special cases. Therefore, in this paper we discuss asymptotic bootstrap validity also in cases where consistency of the bootstrap distribution for a conditional (null) limit distribution of an original statistic holds in a sense weaker than the usual weak convergence in probability.

To show the practical relevance of our results, we include an analysis of the well-known and much applied (but also misunderstood) bootstrap tests of parameter constancy in regression models where the design matrix could be random but be conditioned upon; see Hall (1991,p.170). In the resampling process forming the bootstrap sample, it appears natural to take the design matrix as fixed, i.e. it does not vary across the bootstrap repetitions. Accordingly, bootstrap algorithms with this feature are sometimes labelled as ‘fixed design’, ‘fixed regressor’ or ‘conditional’ bootstrap. Under a set of assumptions proposed by Hansen (2000), we argue that the fixed-regressor bootstrap test statistics have random limit distributions, thus invalidating previous claims that the bootstrap is consistent for the unconditional limit distribution of the original parameter constancy test statistics. Then we provide conditions under which the fixed-regressor bootstrap entails conditional asymptotic inference.

# Testing for Series Correlation and ARCH Effect of High-Dimensional Time Series Data

Speaker: Shiqing Ling \*

*Hong Kong University of Science and Technology*

(this is a joint work with Ruey S. Tsay in University of Chicago  
and Yaxing Yang in Xiamen University)

The small  $n$ /large  $p$  problem has been extensively studied in the last decade for independent and identically distributed (iid) data. Various methods have been developed for variable selection in a large data set with a sparse covariance matrix, such as the lasso-type estimation procedure in Cai et al. (2011) and random matrix theory in Bai et al. (2015); see also Paul and Aue (2014), and the principal component procedure in Shen and Huang (2007). There have been several papers that tried to explore the small  $n$ /large  $p$  problem in the time series setting. For example, Song and Bickel (2011) and Han and Liu (2016) proposed two different approaches to estimate the parameters of a large vector autoregressive model.

To make our discussion clear, we first define the high-dimensional time series (HDTS) as follows.

**Definition 1.1.** A stochastic process  $X_t = (x_{1t}, \dots, x_{pt})'$  is called a high dimensional (HD) time series if  $X_t \rightarrow (x_{1t}, \dots, x_{it}, \dots)'$  as  $p \rightarrow \infty$ , where  $t = 0, \pm 1, \pm 2, \dots$ . We say that a HDTS  $\{X_t\}$  is weakly stationary if, for any given  $p$ ,

- (a).  $EX_t = \mu^{(p)}$ ,
- (b).  $E(X_t - \mu^{(p)})(X_{t+l} - \mu^{(p)})' = \Gamma_l^{(p)}$ ,  $l = 0, \pm 1, \pm 2, \dots$ ,

where  $\mu^{(p)}$  is a  $p$ -dimensional constant vector and  $\Gamma_l^{(p)}$  is a  $p \times p$  constant matrix independent of  $t$ . Similarly, we say that a HDTS  $\{X_t\}$  is strictly stationary if it is strictly stationary any given  $p$ . As a special case, if  $X_t$  is serially independent of a given  $p$ , then  $X_t$  is a HD white noise. When  $\Gamma_l^{(p)} = 0$  for all  $l(\neq 0)$  and  $p$ , we say that the HD time series  $\{X_t\}$  is serially uncorrelated. Let  $\mathcal{F}_t$  be the  $\sigma$ -field generated by  $\{X_s : s \leq t\}$  and denote

$$E[(X_t - \mu^{(p)})(X_t - \mu^{(p)})' | \mathcal{F}_{t-1}] = \Omega_{t-1}^{(p)}.$$

---

\*E-mail address: maling@ust.hk

The HDTS  $\{X_t\}$  is said to have no ARCH effect if, for any given  $p$ ,  $\Omega_t^{(p)} = \Gamma_0^{(p)}$ , a constant matrix; otherwise, it has ARCH effect.

For ease in notation, we drop the superscript of  $\mu^{(p)}$  and  $\Gamma_l^{(p)}$  in the sequel. Keep in mind, however, that the dimensions of  $\mu$  and  $\Gamma_l$  are increasing with  $p$ .

Using the random matrix theory, Li, Yao, Lam and Yao (2016) proposed a test for serial correlation of a HD time series. Change, Yao and Zhou (2016) used the maximum absolute auto- and cross-correlation to construct a test for the same purpose, but critical values of the test statistic need to be obtained via a bootstrap procedure. Recently, Tsay (2017) proposed a rank-based test using Spearman's rank correlation and the extreme value theory. The asymptotic distribution is derived under the null hypothesis and contemporaneous independence. The moment condition of the data is not needed and the asymptotic critical values of test statistics have a close-form solution. This is a nonparametric approach, see more review on this in Section 3.

Testing for the ARCH effect in a time series has become an important issue in analysis of univariate and multivariate time series since Engle (1982) proposed the ARCH model. The classical methods for detecting ARCH effect are the Lagrange multiplier test of Engle (1982) and the Portmanteau test of Li and Mak (1986), which are based on the squares of random variables. Li and Li (2008) considered two Portmanteau tests by using the absolute random variables. Zhu and Ling (2011) proposed a sign based test statistic. Ling and Li (1997) extended Li-Mak test for multivariate ARCH models by using a quadratic form of random vectors. We refer to Tsay (2014) for more discussions on the subject. To the best of our knowledge, there is no test statistic available for detecting the ARCH effect in a HD time series.

This paper proposes two Portmanteau tests for detecting serial correlation and ARCH effect of a HD time series  $\{X_t\}$  allowing the dimension of data  $p = p(n) \rightarrow \infty$  when the sample size  $n \rightarrow \infty$ . We first show that the sample autocorrelation function of the  $L_1$ -norm of data is asymptotically normal and a norm-based Portmanteau test statistic can be constructed. When the cross-sectional variables are  $s$ -dependent (i.e., at most  $s$  elements are dependent), the test statistic still works well in the case with  $p > n$ . Using a suitable function of data, the norm-based test can be applied for the heavy-tailed time series. We next show that the sample rank autocorrelation function, i.e. Spearman's rank correlation, of the  $L_1$ -norm of data is asymptotically normal and the norm-based rank Portmanteau test statistic can also be constructed. Surprisingly, the norm-based rank test is dimension-free, i.e. independent of  $p$ , and without requiring any moment condition of data or the covariance structure condition required in the literature. Two standardized norm-based tests are further discussed. Simulation results show that both test statistics have satisfactory size and are very powerful even for small  $n$  and large  $p$ .

# Cancer Gene Analysis by Singh et al. Microarray Data

成蹊大学 名誉教授 新村秀一 (しんむら しゅういち)

## 1. はじめに

Fisher が計算機のない時代、Fisher の仮説を考えることで Fisher の線形判別関数 (F-LDF) を定式化し、判別分析の世界を開き多くの分野で成果を上げた。彼は Iris データで検証し、さらに「現実のデータが仮説を満たさない場合、推測学的結論は信用できないことがある」と明確に述べている。筆者は多くの医学データなどの研究を通して 5 つの問題を見つけた。そして最小誤分類数 (Minimum Number of Misclassification, MNM) 基準による最適線形判別関数 (Revised IP-OLDF, RIP) [2] と「小標本のための 100 重交差検証法 (新手法 1)」と「癌の遺伝子データの Matryoshka 特徴抽出法 (新手法 2)」を開発して 6 種の実データと 6 種の Microarray データで実証研究し 2017 年 3 月の特任教授退任を前に全ての問題を解決し、昨年末に Springer から解説書を出版した [3]。本発表では、応用研究として行った米国の 6 研究グループが公開している Microarray データ (Dataset) を用いて行った「癌の遺伝子解析から遺伝子診断」の重要点を Singh ら [5] の結果で報告する。

## 2. Microarray データ

2015 年 10 月 26 日に、「多様な分野における統計科学の新展開」で石井ら [1] の発表で、米国の 6 研究グループが Microarray データを集め論文を書いている、そのデータが HP (<http://www.bioinf.ucd.ie/people/ian/>) [8] から入手できることを知った。2015 年 10 月 28 日にそれをダウンロードし 8 種の LDF で判別した。3 種の OLFDF は MNM=0 で、k 以下の遺伝子の判別係数が 0 でなく残り全て 0 になった。即ち自然に k 個以下の遺伝子で癌遺伝子が特定できる。これを Small Matryoshka (SM1) と呼ぶ。この SM1 を全遺伝子から省いて再度判別すると別の SM2 が求まる。そして Dataset は複数個の SM の排他的和集合の信号と、高次元の MNM ≥ 1 の雑音に分かれる事が分かった。30 年以上癌の遺伝子解析が行われ、有用な結果が得られなかった (問題 5)。これは、統計手法で雑音を含んだ Dataset の有効な分析ができない事と、SM の排他的和集合という特殊な構造が分からなかったためと考えられる。そして、12 月 20 日までに「6 種のデータ全てが、表のように SM の排他的和集合である」ことが分かった。

表 6 個の Dataset の新手法 2 の結果 (JMP 列まで) と遺伝子診断 (SM 列以降)

章	Dataset	2 群と患者数	JMP	SM	Max Ratio	Min Ratio	>=5%	PCA
2	Alon et al.	Normal (22) vs. tumour cancer (40)		BGS130	0.90%	0.00%	0	4.50%
3	Alon et al.	Normal (22) vs. tumour cancer (40)	5 (8.0)	64	26.76%	2.35%	63	30.40%
4	Singh et al.	Normal (50) vs. tumour prostate (50)	2 (1.6)	179	11.67%	0.28%	38	14.35%
5	Golub et al.	All (47) vs. AML (25)	8 (11.6)	69	15.69%	0.00%	13	34.88%
6	Tien et al.	False (36) vs. True (137)	3 (3.9)	159	19.13%	0.63%	27	24%
7	Chiaretti et al.	B-cell (95) vs. T-cell (33)	10 (9.8)	95	38.98%	10.73%	95	51.46%
8	Shipp et al.	Follicular lymphoma (19) vs. DLBCL (58)	29 (16.8)	130	30.67%	4.99%	129	31.70%

## 3. 癌の遺伝子解析から癌の遺伝子診断へ

今年に入ってから、遺伝子数が 2000 個と少ない Alon らのデータが筆者が BGS (Basic Gene Set) と呼ぶ癌遺

伝子 130 組の排他的和集合であることが分かった。BGS は、1 個の遺伝子を省くと MNM が 1 以上になる最小の SM である。Datasets は統計手法でよい結果が出なかったが、

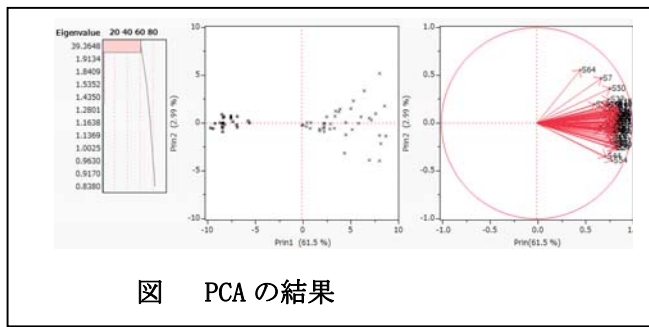


図 PCA の結果

各 SM は小標本のため統計的にアプローチできるので、F-LDF、2 次判別関数、ロジスティック回帰、一元配置の分散分析、クラスター分析、主成分分析で分析を行った。しかし、ロジスティック回帰だけが BGS と SM のすべてが NM=0 であるが、他の手法は 2 群が線形分離可能な兆候を示さなかった。しかし、BGS と SM を RIP で判別した判別スコアを新しい変数としたデータを作成した。2 群の平均の t 検定は、t 値が正から、ほぼ 0、そして負のものが得られた。クラスター分析と PCA は 2 群がきれいに分かれた。そのうえ、PCA の第 1 固有値は、青島らの []PCA による

第 1 固有値がスパイク状になる研究と同じである。また、図のスコアプロットのように正常症例は第 1 主成分軸上の負に、がん症例が正にきれいに分かれた。また、線形分離可能な判別分析の評価に SV 間の距離 2 を判別スコアの範囲に対する比を RatioSV という統計量で評価することにした。表の章列は、[4]の章番号を示す。6 種の数理計画法による LDF は、SV で -1 以下に正常患者、1 以上に癌患者が判別される。しかし判別結果の評価基準がなかった。SV の距離の 2 が判別スコアの範囲の長さの何%であるかを表す RatioSV (=2/判別スコアの範囲\*100) を考えた。これが表の Max Ratio と Min Ratio 列である。Alon 他の 130 個の RIP の中で最大は 0.90%で最小は 0.00%

(0.005 未満)である。これで新規の患者の検証標本でよい結果が出るとは考えられない。しかし SM では最大が 26.76%で、最小が 2.35%である。この値が 5%以上のものが 63 個あり、1 個だけが 2.35%である。この閾値は今後の課題であるが、仮に 5%以上とすれば、63 個の RIP が癌診断に利用できる。Chiaretti の場合、約 40%の大きな窓が開き残りの 60%に正常と異常が散らばっている。また Ward 法と PCA の結果は綺麗に 2 群に分かれる。図は Alon 他の 64 個の判別スコアデータの PCA の結果である。左の固有値を見ると第 1 主成分 (Prin1) の固有値が 39.4 と Spike 状に大きい。これは 2 群が大きく離れ、2 群の群内分散が小さいためと考える。右の因子負荷量は、4 象限から 1 象限に布置している。真ん中のスコアプロットは、正常群が Prin1 の負にほぼ直線上に、癌は原点から正の方向に扇のように布置している。右の癌患者は悪性で原点近くの癌は比較的軽度の癌患者であり、正常の原点に近い患者は癌になりやすいか否かは、病理学者が患者カルテを調べれば分ると考えている。即ち、Prin1 の主成分軸は 63 個の RIP の判別スコアと同じく癌の悪性度の指標に使えることが期待される。もう一つの応用としては、癌患者が治療で正常領域に誤判別されれば 5 年も経過観察なく治癒したと確定できると素人的に考える。主成分軸上の RatioSV は 30.4%で 26.76%よりも約 4%も大きいのは、64 個の合成のためである。

#### 4. まとめ

残念なことに、せっかく求めた癌の悪性度指標が、公開されたデータの変数が真の遺伝子名を他の変数に置き換えているために医師による検証が行えない。真の遺伝子名を知る方法をご存知の方は、教えていただきたい。

#### REFERENCES

1. Ishii A, Yata K, Aoshima M (2014) Asymptotic distribution of the largest eigenvalue via geometric representations of high-dimension, low-sample-size data. Sri Lankan J. Appl. Statist., Special Issue: Modern Statistical Methodologies in the Cutting Edge of Science (ed. Mukhopadhyay, N.): 8194.
2. 新村秀一 (2010). 最適線形判別関数. 日科技連出版.
3. Shinmura, S. (2016). The New Theory of Discriminant Analysis after R Fisher, Springer
4. Shinmura, S. (2017). From Cancer Gene Analysis to Cancer Gene Diagnosis, Amazon, 2017年6月.
5. Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer-Verlag.
6. 新村秀一 (2011). 数理計画法による問題解決法. 日科技連出版.
7. 新村秀一 (2004). JMP活用 統計学とおき勉強法. 講談社.
8. Jeffery, IB. et al. (2006). Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data, BMC Bioinformatics. 7:359: 1-16.

# 二元分割表における独立性および 対称性のモデルと分解

佐藤 駿 (東京理科大学大学院 理工学研究科)  
合田 新平 (東京理科大学大学院 理工学研究科)  
三枝 祐輔 (横浜市立大学 医学部)  
生亀 清貴 (東京理科大学 理工学部)  
山本 紘司 (大阪市立大学大学院 医学研究科)

## 第一部「既知のスコアを用いた儉約な独立モデルの直交分解」

行と列に順序がある  $r \times c$  分割表において、行変数を  $X_1$ 、列変数を  $X_2$  とし、 $P(X_1 = i, X_2 = j) = p_{ij}$  とする ( $i = 1, \dots, r; j = 1, \dots, c$ )。行変数と列変数の独立性が成り立たない場合には、それらの行と列の連関構造に関心がある。行変数と列変数に既知のスコア  $\{u_i\}, \{v_j\}$  をそれぞれ割り振ることができるとき (ただし、 $u_1 < \dots < u_r, v_1 < \dots < v_c$ )、Tomizawa (1992) は儉約な Linear-by-Linear 連関 (PLL) モデルを次のように提案した:

$$p_{ij} = \mu \alpha^{u_i} \beta^{v_j} \theta^{u_i v_j} \quad (i = 1, \dots, r; j = 1, \dots, c).$$

特に、 $\{u_i = i\}, \{v_j = j\}$  のとき、PLL モデルは儉約な一様連関 (PU) モデルである。

ここで、 $g_1(i) = u_i$  ( $i = 1, \dots, r$ )、 $g_2(j) = v_j$  ( $j = 1, \dots, c$ ) とおく。  $Y_1, Y_2$  を  $Y_1 = g_1(X_1), Y_2 = g_2(X_2)$  によって定義し、 $Y_1$  と  $Y_2$  の Pearson 相関係数を  $\rho(Y_1, Y_2)$  とする。また、 $G^2(M)$  をモデル  $M$  の適合度を検定する尤度比カイ二乗統計量とする。本報告では、行と列に既知のスコアを割り振ることができる場合に、儉約な独立モデルを定義した。さらに、PLL モデルおよび、 $\rho(Y_1, Y_2)$ 、Kendall の  $\tau_b$  (Kendall, 1945)、Spearman の  $\rho_s$  (Stuart, 1963; Kendall and Gibbons, 1990, p. 8) を用いて、儉約な独立モデルの分解について考えた。

**定理 1.**  $PI$  モデルが成り立つための必要十分条件は、PLL モデルと  $\rho(Y_1, Y_2) = 0$  の両方が成り立つことである。

**定理 2.**  $PI$  モデルが成り立つための必要十分条件は、PLL モデルと  $\tau_b = 0$  の両方が成り立つことである。

**定理 3.**  $PI$  モデルが成り立つための必要十分条件は、PLL モデルと  $\rho_s = 0$  の両方が成り立つことである。

**定理 4.**  $G^2(PI)$  は、 $G^2(PLL)$  と  $G^2(\rho(Y_1, Y_2) = 0)$  の和に漸近的に同等である。

## 第二部「順序付き正方分割表における非制限正規分布型対称モデル」

行と列が順序のある同じ分類からなる  $r \times r$  正方分割表を考える。正方分割表解析においては、分類間の対称性の解析に関心があり様々なモデルが提案されている。 $(i, j)$  セル確率を  $p_{ij}$  とする ( $i = 1, \dots, r; j = 1, \dots, r$ )。

Yamamoto, Nakane and Tomizawa (2016) は次のような制限正規分布型対称 (RNDS)

モデルを提案した:

$$p_{ij} = \mu \alpha^{i^2+j^2} \beta^{i+j} \gamma^{ij} \quad (i = 1, \dots, r; j = 1, \dots, r).$$

対称性のモデルが成り立たないときには非対称性の構造に関心があり, Tomizawa (1991) は次のような拡張線形対角パラメータ対称 (ELDPS) モデルを提案した:

$$p_{ij} = \begin{cases} \delta^{j-i} \gamma^{(j-i)(j+i)/2} \psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

ただし  $\psi_{ij} = \psi_{ji}$ .

本報告では, RNDS モデルの制限を取り除いたモデルを提案し, そのモデルを用いた RNDS モデルの分解について述べた.

非制限正規分布型対称 (UNDS) モデルを次のように提案した:

$$p_{ij} = \mu \alpha_1^{i^2} \alpha_2^{j^2} \beta_1^i \beta_2^j \gamma^{ij} \quad (i = 1, \dots, r; j = 1, \dots, r).$$

特に  $\alpha_1 = \alpha_2$  かつ  $\beta_1 = \beta_2$  のとき UNDS モデルは RNDS モデルである. また, ELDPS モデルの特別な場合であることに注意する.

行と列が順序のある同じ分類からなる  $r \times r$  正方分割表において, 行変数を  $X_1$ , 列変数を  $X_2$  とし, 平均分散一致 (MV) モデルを次のように定義した:

$$E(X_1) = E(X_2), \quad \text{かつ} \quad \text{Var}(X_1) = \text{Var}(X_2),$$

ただし,

$$\begin{aligned} E(X_1) &= \sum_{k=1}^r k p_{k.}, & E(X_2) &= \sum_{k=1}^r k p_{.k}, \\ \text{Var}(X_1) &= \sum_{k=1}^r (k - E(X_1))^2 p_{k.}, & \text{Var}(X_2) &= \sum_{k=1}^r (k - E(X_2))^2 p_{.k}, \end{aligned}$$

次の定理について述べた:

**定理 5.** RNDS モデルが成り立つための必要十分条件は, UNDS モデルと MV モデルの両方が成り立つことである.

## 参考文献

- Goodman, L. A. (1979). *Journal of the American Statistical Association*, **74**, 537-552.  
 Kendall, M. G. (1945). *Biometrika*, **33**, 239-251.  
 Kendall, M. G. and Gibbons, J. D. (1990). *Rank Correlation Methods*, 5th edition. Edward Arnold, London.  
 Stuart, A. (1963). *American Statistician*, **17**, 23-24.  
 Tomizawa, S. (1991). *Metron*, **49**, 401-409.  
 Tomizawa, S. (1992). *Biometrical Journal*, **34**, 129-140.  
 Yamamoto, K., Nakane, H. and Tomizawa, S. (2016). *Journal of Statistics: Advances in Theory and Applications*, **15**, 71-84.

## 正方分割表における対称性の尺度について

東京理科大学 工学部 安藤宗司  
 東京理科大学 理工学部 田畑耕治  
 東京理科大学 理工学部 富澤貞男

行と列が同じ分類からなる正方分割表データについて考える。表1のデータは、臨床試験における2つの治療群（エソメプラゾール群とプラセボ群）のベースライン時点と最終評価時点のカテゴリカルアウトカム（修正LANZAスコア）に関する変化を表しており、正方分割表データである。修正LANZAスコアは、“0”が最も良い評価，“+4”が最も悪い評価となっている。表1のデータにおいて、プラセボ群よりもエソメプラゾール群の方が治療効果が優れているかどうかに関心がある。

表1: エソメプラゾール群とプラセボ群におけるベースライン時点と最終評価時点での修正LANZAスコアの変化；Sugano, Kinoshita, Miwa and Takeuchi (2012).

(a) エソメプラゾール群							(b) プラセボ群						
最終評価 時点	ベースライン時点					計	最終評価 時点	ベースライン時点					計
	0	+1	+2	+3	+4			0	+1	+2	+3	+4	
0	78	9	26	3	1	117	0	41	2	19	0	0	62
+1	1	5	6	4	0	16	+1	8	0	4	0	0	12
+2	9	1	10	3	1	24	+2	12	4	14	3	0	33
+3	1	0	1	0	0	2	+3	0	1	1	3	0	5
+4	3	0	1	1	2	7	+4	29	7	11	6	0	53
計	92	15	44	11	4	166	計	90	14	49	12	0	165

正方分割表データの解析においては、分類間の独立性に代わり対称性や非対称性に関するモデルが用いられる。対称性に関するモデルとして、対称モデル (Bowker, 1948) が提案されている。表1のデータに対して対称モデルが成り立つとき、ベースライン時点と最終評価時点の修正LANZAスコアは同等であることを指している。また、表1のデータが非対称の構造を持つとき、ベースライン時点に対して最終評価時点の修正LANZAスコアは改善または悪化していることを指している。

分割表データにおいて、モデルの適合度を評価する際に、適合度検定統計量（例えば、尤度比カイ二乗統計量）が用いられる。複数の分割表データに対して対称モデルの当てはまりが悪いとき、非対称性の程度を比較することに関心がある。適合度検定統計量を用いて、複数の分割表データに対して非対称性の程度を比較する際に2つの問題点が挙げられる。



1. 分割表のサイズ及びサンプルサイズに依存する
2. 非対称の方向性（改善または悪化等）を区別できない

臨床試験において、各治療群のサンプルサイズが等しくない場合もあること、治療効果を比較する際に非対称性の構造を区別する必要があるため、適合度検定統計量を用いて治療群間の治療効果を比較をすることはできない。

順序カテゴリ正方分割表において、Tomizawa, Miyamoto and Hatanaka (2001) は対称性からの隔たりを測る尺度  $\gamma$  を提案した。この尺度  $\gamma$  は分割表のサイズ及びサンプルサイズに依存しないが、非対称の方向性（改善または悪化）を区別できないことに注意する。Tahata, Miyazawa and Tomizawa (2010) は、2種類の非対称性の方向性を区別することができる尺度  $\phi$  を提案した。この尺度  $\phi$  は2種類の非対称性の方向性を区別できるが、対称モデルからの隔たりを測っていないことに注意する。

本講演では、分割表のサイズ及びサンプルサイズに依存せず、非対称性の程度と2種類の非対称性の方向性を同時に分析することができるベクトル尺度  $\Phi$  を提案した。

$$\Phi = \begin{pmatrix} \gamma \\ \phi \end{pmatrix}; 2 \times 1 \text{ ベクトル.}$$

提案したベクトル尺度  $\Phi$  を用いることにより、2種類の非対称性の方向性を区別しながら、非対称性の程度を測ることができることを示した。また、提案したベクトル尺度は信頼楕円を用いて、複数の分割表データに対する非対称性の程度を視覚的に比較することに有用であることを示した。

# ネットワークメタアナリシスによる **Comparative Effectiveness Research** と 高次漸近理論に基づく推測手法

統計数理研究所 野間 久史

メタアナリシスとは、過去に行われた複数の臨床試験のエビデンスを統合し、総合的な治療効果の評価を行うための方法論であり、**Evidence-Based Medicine** における重要な研究手法として広く普及している。しかしながら、従来のメタアナリシスの方法は、原則として、対象となる治療の対比較（1対1の比較）の結果を統合するという単純化されたものであり、「有効性についてのエビデンスが確立された複数の治療の選択肢の中から、どの治療を行うのが最善なのか？（有効性、安全性は？費用対効果は？）」などの問いに必ずしも本質的な答えを与えてくれるものではない。折しも、近年、先進諸国における（超）高齢化社会の到来により、医療費・医療資源の節減および効率的な配分のために、比較効果研究（**comparative effectiveness research**）が世界的に大きな関心を集めており、このような複数の治療を対象とした有効性・有用性を総合的に評価するための方法論に対する要請は飛躍的に高まっている。ネットワークメタアナリシス（**network meta-analysis**）は、このような背景のもと、急速に普及しつつある、従来のメタアナリシスの方法を一般化し、複数の治療の有効性・有用性を比較・評価することを目的とした方法論である。ネットワークメタアナリシスでは、複数の対象となる治療を含む臨床試験の結果を系統的に集め統合し、治療間の間接比較の情報も併せたエビデンスの統合が行われる。これにより、直接比較の行われていない治療間の比較も含めて、対象となった治療法すべての有効性・有用性を比較することが可能となる。

しかしながら、ネットワークメタアナリシスは、この数年で急速に普及した新規な方法論でもあり、現在、実践で普及しているスタンダードな方法でも、さまざまな統計学的問題を抱えている可能性がある。その問題のひとつとして、**Brockwell and Gordon (2001)**, **Noma (2011)** などで議論されている推測の妥当性の問題がある。メタアナリシスでは、異なる情報源から得られるエビデンスの異質性を考慮した上で情報の統合を行うために、変量効果モデル（**random effects model**）を用いた解析が行われるのが一般的である。ネットワークメタアナリシスでは、対比に基づく（**contrast-based**）方法として、以下の多変量変量効果モデルが広く用いられている。

$$Y_i \sim MVN(\theta_i, S_i), \theta_i \sim MVN(\mu, \Sigma), \quad (*)$$

$$\mathbf{S}_i = \begin{pmatrix} s_{i1}^2 & \rho_{i12}s_{i1}s_{i2} & \cdots & \rho_{i1p}s_{i1}s_{ip} \\ \rho_{i12}s_{i1}s_{i2} & s_{i2}^2 & \cdots & \rho_{i2p}s_{i2}s_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{i1p}s_{i1}s_{ip} & \rho_{i2p}s_{i2}s_{ip} & \cdots & s_{ip}^2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \tau_1^2 & \kappa_{12}\tau_1\tau_2 & \cdots & \kappa_{1p}\tau_1\tau_p \\ \kappa_{12}\tau_1\tau_2 & \tau_2^2 & \cdots & \kappa_{2p}\tau_2\tau_p \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_{1p}\tau_1\tau_p & \kappa_{2p}\tau_2\tau_p & \cdots & \tau_p^2 \end{pmatrix}$$

$\mathbf{Y}_i$ は、 $i$  番目の試験における、特定の参照レベル（プラセボなど）に対しての対象となる治療の治療効果の指標（ハザード比，オッズ比など）の推定量からなる確率ベクトルであり、 $\boldsymbol{\theta}_i$ はその真値を表すベクトル、 $\mathbf{S}_i$ は試験内共分散行列（ $\mathbf{Y}_i$ の共分散行列）、 $\boldsymbol{\Sigma}$  は試験間共分散行列を表す。パラメータの推定は、制限付き最尤法（restricted maximum likelihood; REML）によって行われるのが一般的である。通常、試験間の異質性を表す  $\boldsymbol{\Sigma}$  に関する情報は、 $\mathbf{Y}_i$  のデータ数に比例することとなり、試験数が十分に大きいもとで、大標本理論による正当化が可能となる。しかしながら、多くのメタアナリシスでは、統合の対象となる試験数が十分に大きくはなく、古典的な2つの治療法の対比較のメタアナリシスでは、20にも満たない試験での統合が行われるのが一般的である。この場合、大標本近似の崩れから、信頼区間の被覆率が名目水準を大きく下回るなどの問題があることが知られている（Noma, 2011）。ネットワークメタアナリシスでは、複数の治療法の比較を対象としていることもあり、相対的に対象となる試験数は大きくなるが、サブグループ解析などの感度解析なども含めれば、数試験程度の条件下で統合解析を行う状況は多く存在する。

本講演では、シミュレーション実験を通して、まず、現状のスタンダードな推測手法である REML 法に、上記のような条件下で、推測の妥当性が成り立たない状況が多く存在することを示す。そして、演者らが近年開発した、ネットワークメタアナリシスの解析モデルに汎用的に用いることができる高次漸近理論に基づく有効な推測手法（Noma *et al.*, 2017）の紹介を行う。また、シミュレーション実験および統合失調症などのネットワークメタアナリシスの事例解析を通して、その実践的有用性について示す。

## 文献

- Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med.* 2001;20(6):825-840.
- Noma H. Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Stat Med.* 2011;30(28):3304-3312.
- Noma H, Nagashima, K, Maruo K, Goshio, M., Furukawa TA. Bartlett-type corrections and bootstrap adjustments of likelihood-based inference methods for network meta-analysis. *Stat Med.* In Press, 2017.