科研費シンポジウム

「統計科学と関連分野における諸問題に関する理論と方法論の革新的展開」

（Innovative developments in theory and methodology regarding various issues in statistical science and related fields）

日時：2023 年 9 月 29 日（金）～10 月 1 日（日）

(Date: Friday, 29, September – Sunday, 1, October, 2023)

場所：新潟大学駅南キャンパスときめいと 講義室 A,B

(Place: Niigata University, Station South Campus Tokimate, Lecture Room A, B)

（TEL: 025-248-8141）

科学研究費・基盤研究（A）（課題番号：20H00576）

「大規模複雑データの理論と方法論の革新的展開」

研究代表者：青嶋 誠（筑波大学）

Makoto Aoshima (University of Tsukuba)

開催責任者：蛭川 潤一（新潟大学）

Junichi Hirukawa (Niigata Universiry)

Program

Friday, 29, September

Reception 13:00-13:15

Opening 13:15-13:20 **Junichi HIRUKAWA** (Niigata University)

Afternoon Session I (in Japanese) 13:20-16:00
Chair **Shimatani, K. Ichiro** (The Institute of Statistical Mathematics)
1. 13:20-14:00 **江頭 健斗**, 矢田 和善, 青嶋 誠
東京理科大学創域理工学部, 筑波大学数理物質系, 筑波大学数理物質系
Hierarchical clustering and its asymptotic behaviors in multiclass HDLSS settings
2. 14:00-14:40 **齊藤 実祥**, 寒河江 雅彦
金沢大学 人間社会研究域 経済学経営学系
多次元 Tensor Product 型密度関数の推定とその漸近的性質
3. 14:40-15:20 **種市 信裕**, 関谷 祐里
元北海道教育大札幌校, 北海道教育大学釧路校
2nd order correction term を最小化する分割表のΦ - ダイバージェンス独立性検定統計量
4. 15:20-16:00 **柿沢 佳秀**
北海道大学大学院経済学研究院
境界バイアスを回避する非対称カーネル法の様々な適用について

Coffee Break 16:00-16:30

Afternoon Session II (in Japanese & English) 16:30-17:50
Chair **Yuichi Goto** (Kyushu University)
5. 16:30-17:10 曾 小強 (Xiaoqiang Zeng), **柿沢 佳秀 (Yoshihide Kakizawa)**
Hokkaido University
Some estimators in the ADCINAR(1) process
6. 17:10-17:50 **竹内 努 (Tsutomu T. TAKEUCHI)**
名古屋大学素粒子宇宙物理学専攻 (Division of Particle and Astrophysical Science, Nagoya University, Japan)
多様体学習による銀河進化の定量化と定式化 (Quantification and Formulation of Galaxy Evolution by Manifold Learning)

Saturday, 30, September

Morning Session (in Japanese) 9:30-10:50
Chair **Yan Liu** (Waseda University)
7. 9:30-10:10 **永井 勇**
中京大学 教養教育研究院
GMANOVA モデルにおけるフルランク仮定が不要な罰則無推定法
8. 10:10-10:50 **藤森 洸**, 佃 康司
信州大学, 九州大学
Two step estimations via the Dantzig selector for models of stochastic processes with high-dimensional parameters

Lunch 10:50-13:20

Afternoon Session I (in English) 13:20-15:20
Chair **Kou Fujimori** (Shinshu University)
9. 13:20-14:00 **島谷健一郎 (Shimatani, K. Ichiro)**
統計数理研究所 (The Institute of Statistical Mathematics)
空間種分布モデルに対する多様性指数の統計数理 (Statistical mathematics of biodiversity indices for spatial species distribution model)
10. 14:00-14:40 **張 元宗 (Chang Yuan-Tsung)**, 篠崎 信雄 (Nobuo Shinozaki)
目白大学 (Mejiro university), 慶應大学 (Keio university)
Simultaneous estimation of Poisson means: recent developments and its applications
11. 14:40-15:20 **田中 勝人 (Katsuto Tanaka)**
Hitotsubashi University (Emeritus Professor)
Extensions of Darling's Formula for the Fredholm Determinant

Coffee Break 15:20-15:50

Afternoon Session II (Guest speakers session) 15:50-17:50

Chair **Junichi Hirukawa** (Niigata University)

12. 15:50-16:30 **Liang-Ching Lin**

National Cheng Kung University

Authors: Liang-Ching Lin (National Cheng Kung Univ.), Hsiang-Lin Chien (National Cheng Kung Univ.), Hao Sung (National Cheng Kung Univ.) and Sangyeol Lee (Seoul National University)

Comprehensive interval-valued time series model with application to the S&P 500 index and PM2.5 level data analysis

13. 16:30-17:10 **Nan-Jung Hsu**

National Tsing Hua University

Authors: Nan-Jung Hsu (National Tsing-Hua Univ), Hsin-Cheng Huang (Institute of Statistical Science, Academia Sinica), Ruey S. Tsay (University of Chicago) and Tzu-Chieh Kao (National Tsing-Hua Univ.)

Matrix Autoregressive Spatio-Temporal Models

14. 17:10-17:50 **Hsin-Cheng Huang**

Institute of Statistical Science, Academia Sinica

Authors: Chien-Chung Wang (Colorado State Univ.), Chih-Hao Chang (National Chengchi Univ.) and Hsin-Cheng Huang (Institute of Statistical Science, Academia Sinica)

Stable p-value Assignment in High-Dimensional Regression via Data Splitting

Conference Dinner 18:30-21:30

いかの墨 新潟駅前店

(Ika-no-sumi)

6.000 yen

(TEL: 025-242-0510)

Sunday, 1, October

Morning Session (in English) 9:30-12:10
Chair **Katsuto Tanaka** (Hitotsubashi University)
15. 9:30-10:10 **塩濱 敬之 (Takayuki Shiohama)**
南山大学理工学部 (Department of Data Science, Nanzan University)
Modeling Joint Cylindrical Distributions and Related Markov Processes
16. 10:10-10:50 **後藤 佑一 (Yuichi Goto)**
九州大学 (Kyushu University)
Authors: Yuichi Goto (Kyushu Univ.), Xuze Zhang (Maryland Univ.), Benjamin Kedem (Maryland Univ.) and Shuo Chen (Maryland Univ.)
Test for the existence of the residual spectrum with application to brain functional connectivity detection
17. 10:50-11:30 **Yan Liu**
Faculty of Science and Engineering, Waseda University
Prediction-based statistical inference for multiple time series
18. 11:30-12:10 **Junichi Hirukawa**, Kou Fujimori
Niigata University, Shinshu University
Weak convergence of the partial sum of $I(d)$ process to a fractional Brownian motion in finite interval representation

Closing 12:10-12:15 **Junichi HIRUKAWA** (Niigata University)

Hierarchical clustering and its asymptotic behaviors in multiclass HDLSS settings

Kento Egashira[a], **Kazuyoshi Yata**[b], **Makoto Aoshima**[b]

[a]**Faculity of Science and Technology, Tokyo University of Science**
[b]**Institute of Mathematics, University of Tsukuba**

Hierarchical clustering is a methodology to group a set of datas by building dendrogram based on a similarity or a dissimilarity between clusters so that datas in a cluster are similar in the sense of pre-determined linkage function. In hierarchical clustering, one can observe a process how a cluster is combined or devided through dendrogram on graphic. Hierarchical clustering has been approved as useful tool for analysis of gene expression microarray data. In fact, applications of hierarchical clustering on gene expression microarray data are given by Eisen et al. [4], Perou et al. [8], Bhattacharjee et al. [2], among others. A characteristic of datas used in Eisen et al. [4], Perou et al. [8] and Bhattacharjee et al. [2] is that the number of variables is much larger than sample size. This type of data represented by gene expression microarray data is called high-dimension, low-sample-size (HDLSS) data. Substantial work about clustering has been performed on HDLSS asymptotics in recent years. For example, Liu et al. [6] proposed a two-way split clustering called "statistical significance of clustering(SigClust)" especially for HDLSS data. Ahn et al. [1] proposed a hierarchical divisive clustering and considered its high dimensional asymptotics. Huang et al. [5] developed the SigClust by Liu et al. [6] with soft thresholding approach. Yata and Aoshima [9] gave consistency properties of sample principal component scores and applied it to clustering under high dimensional settings. Nakayama et al. [7] investigated clustering by kernel principal component analysis for HDLSS data. Borysov et al. [3] studied behaviors of hierarchical clustering under several asymptotic settings from moderate dimension through HDLSS, nevertheless it is considered that theoretical assumptions are strict for HDLSS data due to having discussions on several asymptotic settings at once.

Given this background, we focused on HDLSS settings and considered asymptotic properties of hierarchical clustering with several linkage functions. This study explores practical assumptions to indicate the behavior of hierarchical clustering.

In this talk, we theoretically investigated hierarchical clustering when both the dimension and sample size approach infinity at first. Then, we gave an asymptotic behavior in boundary cases of thresholds which divide asymptotic behaviors. Finally, we examined the hierarchical clustering theoretically in multiclass HDLSS context and reported numerical simulations.

References

[1] Ahn, J., Lee, M.H., Yoon, Y.J. (2012). Clustering high dimension, low sample size data using the maximal data piling distance. *Statistica Sinica*, 22, 443–464.

[2] Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., Meyerson, M. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 13790–13795.

[3] Borysov, P., Hannig, J., Marron, J.S. (2014). Asymptotics of hierarchical clustering for growing dimension. *Journal of Multivariate Analysis*, 124, 465–479.

[4] Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 14863–14868.

[5] Huang, H., Liu, Y., Yuan, M., Marron, J.S. (2015). Statistical Significance of Clustering using Soft Thresholding. *Journal of computational and graphical statistics*, 24, 975–993.

[6] Liu, Y., Hayes, D.N., Nobel, A., Marron, J.S. (2008). Statistical significance of clustering for high-dimension, low-sample size data. *Journal of the American Statistical Association*, 103, 1281–1293.

[7] Nakayama, Y., Yata, K., Aoshima, M. (2021). Clustering by principal component analysis with Gaussian kernel in high-dimension, low-sample-size settings. *Journal of Multivariate Analysis*, 185, 104779.

[8] Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A. L., Brown, P. O., Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, 406, 747–752.

[9] Yata, K., Aoshima, M. (2020). Geometric consistency of principal component scores for high-dimensional mixture models and its application. *Scandinavian Journal of Statistics*, 47, 899–921.

多次元 Tensor Product 型密度関数の推定とその漸近的性質

齊藤実祥 (金沢大学), 寒河江雅彦 (金沢大学)

1 研究目的

多次元ヒストグラムにおいて、各ビン間の不連続性によって推定精度が低下する問題がある。そこで、平滑化によるモデル拡張でヒストグラムの推定精度 $O(n^{-2q/2q+d}),\ (q=1)$, (q は推定効率、d は次元数) の改良を考える。先行研究では、Scott(1985) と Hjort(1986) がヒストグラムの各ビンの中点を線形で接続する Frequency Polygon(FP) を d 次元へ拡張し、その推定精度が $O(n^{-4/d+4}),\ (q=2)$ で、ヒストグラムを改良することを示した。本稿では、曲面の近似で用いられる Tensor Product Spline を密度関数の推定に適用し、その理論面を整備する。2 次元の場合での Tensor Product Spline 密度推定量 (以降、TPS 推定量) について、平均積分二乗誤差 (MISE) に基づく推定精度と漸近正規性を示す。その中で、TPS 推定量で推定精度の維持とパラメータの削減が両立できる方法について提案する。また、TPS 推定量の改良モデルを提案し、その漸近的性質から推定精度が改良できることを示す。

2 Tensor Product Spline 密度推定量

2.1 TPS 推定量の構築と漸近的性質

$\mathbb{R}^2$ 上の確率密度関数 $f(x,y)$ から n 個のデータを与え、ヒストグラムのビンの中点を x 軸側 $x_i,\ (i=1,\ldots,N)$、y 軸側 $y_j,\ (j=1,\ldots,M)$、ビン幅を x 軸側 h_1、y 軸側 h_2、$i,\ j$ 番目のビン $B_{i,j}$ の定義域を $B_{i,j}=\left[x_i-\frac{h_1}{2},\ x_i+\frac{h_1}{2}\right]\times\left[y_j-\frac{h_2}{2},\ y_j+\frac{h_2}{2}\right]$ とする。

ビン $B_{i,j}$ において、提案モデル Tensor Product Spline 密度推定量 (以降、TPS 推定量)$\hat{f}_{i,j}(x,y)$ は次の通りである;

$$\hat{f}_{i,j}(x,y)=\begin{pmatrix}1 & (x-x_i) & (x-x_i)^2\end{pmatrix}\begin{pmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{pmatrix}\begin{pmatrix}1\\ (y-y_j)\\ (y-y_j)^2\end{pmatrix},\quad (x,\ y)\in B_{i,j},\tag{1}$$

ただし、$\boldsymbol{A}=\left[a_{kl}^{(i,j)}\right],(k,l=1,2,3)$ は未知の定数で、簡単のため各要素を a_{kl} と略す。

未知の定数 a_{kl} を求める方程式について次の行列表現が可能で、一意に導出できる ;

$$\boldsymbol{A}=\begin{pmatrix}0 & 1 & 0\\ \frac{1}{2} & 0 & \frac{1}{2}\\ -\frac{1}{2h_1} & 0 & \frac{1}{2h_1}\end{pmatrix}\begin{pmatrix}\gamma_{i-1,j-1} & \alpha_{i-1,j} & \gamma_{i-1,j}\\ \beta_{i,j-1} & u_{i,j} & \beta_{i,j}\\ \gamma_{i,j-1} & \alpha_{i,j} & \gamma_{i,j}\end{pmatrix}\begin{pmatrix}0 & \frac{1}{2} & -\frac{1}{2h_2}\\ 1 & 0 & 0\\ 0 & \frac{1}{2} & \frac{1}{2h_2}\end{pmatrix},\tag{2}$$

ただし、既知の各係数について、$B_{i,j}$ の中点での高さは度数 $\nu_{i,j}$ を用いて $u_{i,j}=\nu_{i,j}/nh_1h_2$、$x$ での偏微分 $\alpha_{\cdot,\cdot}=\dot{\hat{f}}_x(\cdot,\cdot)$、$y$ での偏微分 $\beta_{\cdot,\cdot}=\dot{\hat{f}}_y(\cdot,\cdot)$、$x,\ y$ 両方での偏微分 $\gamma_{\cdot,\cdot}=\ddot{\hat{f}}_{xy}(\cdot,\cdot)$ である。

TPS 推定量 $\hat{f}(x,y)$ の漸近的な MISE(AMISE) は、

$$\mathrm{AMISE}\left\{\hat{f}(x,y)\right\}=\frac{1.1125}{nh_1h_2}+\frac{h_1^4}{576}R\left(\ddot{f}_{xx}\right)+\frac{h_2^4}{576}R\left(\ddot{f}_{yy}\right)+\frac{h_1^2h_2^2}{288}R\left(\ddot{f}_{xy}\right),\tag{3}$$

ただし、$R(\phi)=\int\phi(x,y)^2dxdy$ である。

このとき、最適ビン幅は $h_i^*=O\left(n^{-1/6}\right)$、最小 AMISE は $O\left(n^{-2/3}\right),\ (q=2)$ で 2 次元 FP と漸近同等である。また、$\hat{f}(x,y)$ は漸近正規性が成り立つ。

2.2 TPS 推定量のパラメータ削減法

ここで、$\boldsymbol{A}$ でパラメータを全 9 個から 6 個に減らす改良法を提案する。この理由は、多項式における 3 次以上の項 $(x-x_i)^{k-1}(y-y_j)^{l-1},\ k+l\geq 5$ に対応する 3 項を消去することで、MISE の収束レートは変えずに定数項を改良できるからである。

パラメータを削減した TPS 推定量は次の通り表現される；

$$\hat{f}_{i,j}(x,y) = \begin{pmatrix} 1 & (x-x_i) & (x-x_i)^2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ (y-y_j) \\ (y-y_j)^2 \end{pmatrix}, \quad (x,\ y) \in B_{i,j}. \tag{4}$$

この時の AMISE は次の通りである；

$$\mathrm{AMISE}\left\{\hat{f}(x,y)\right\} = \frac{1.0696}{nh_1h_2} + \frac{h_1^4}{576}R\left(\ddot{f}_{xx}\right) + \frac{h_2^4}{576}R\left(\ddot{f}_{yy}\right) + \frac{h_1^2h_2^2}{288}R\left(\ddot{f}_{xy}\right). \tag{5}$$

また、パラメータを削減した $\hat{f}(x,y)$ についても漸近正規性が成り立つ。

3　TPS 推定量の改良

各ビンで面積相等性を満たすように制約条件を追加することで TPS 推定量を改良することを考える。

ビン $B_{i,j}$ における改良版 TPS 推定量 $\hat{g}_{i,j}(x,y)$ は次の多項式で表現され、制約条件に基づく方程式を解くことで各係数が導出される；

$$\begin{aligned} \hat{g}_{i,j}(x,y) = {} & a_{31}(x-x_i)^2 + a_{13}(y-y_j)^2 + a_{22}(x-x_i)(y-y_j) \\ & + a_{21}(x-x_i) + a_{12}(y-y_j) + a_{11}. \end{aligned} \tag{6}$$

改良版 TPS 推定量 $\hat{g}(x,y)$ の漸近的な MISE(AMISE) は、

$$\begin{aligned} \mathrm{AMISE}(\hat{g}(x,y)) = {} & \frac{4.0210}{nh_1h_2} + \frac{1}{30240}\left(h_1^6R\left(\dddot{f}_{xxx}\right) + h_2^6R\left(\dddot{f}_{yyy}\right)\right) \\ & + \frac{h_1^4h_2^2}{6912}R\left(\dddot{f}_{xxy}\right) + \frac{h_1^2h_2^4}{6912}R\left(\dddot{f}_{xyy}\right). \end{aligned} \tag{7}$$

このとき、最適ビン幅は $h_i^* = O\left(n^{-1/8}\right)$、最小 AMISE は $O\left(n^{-3/4}\right),\ (q=3)$ であり、TPS 推定量の推定精度を改良することが示された。

また、改良版 TPS 推定量 $\hat{g}(x,y)$ は漸近正規性が成り立ち、$h \propto O(n^{-\alpha}),\ (x,\ y) \in B_{i,j}$ に対して、

$$\alpha = \frac{1}{8}\text{のとき}：\sqrt{nh_1h_2}\left\{\hat{g}_{i,j}(x,y) - g(x,y)\right\} \xrightarrow{d} N\left(\mathrm{Bias}[\hat{g}_{i,j}(x,y)],\ 4.0210g(\zeta_i,\eta_j)\right), \tag{8}$$

$$\alpha > \frac{1}{8}\text{のとき}：\sqrt{nh_1h_2}\left\{\hat{g}_{i,j}(x,y) - g(x,y)\right\} \xrightarrow{d} N\left(o(1),\ 4.0210g(\zeta_i,\eta_j)\right), \tag{9}$$

ただし、$g(\zeta_i,\eta_j)$ は $p_{i,j} = \int_{B_{i,j}} g(u,v)dudv = h_1h_2g(\zeta_i,\eta_j),\ (\zeta_i,\eta_j) \in B_{i,j}$ を満たす点である。

4　結論

多次元ヒストグラムを Tensor Product 型で平滑化する 2 次元 TPS 推定量の推定精度は $\mathrm{AMISE}(\hat{f}(x,y)) = O\left(n^{-2/3}\right),\ (q=2)$ で、2 次元ヒストグラムの $O\left(n^{-1/2}\right),\ (q=1)$ を改良できるものの、2 次元 FP の $O\left(n^{-2/3}\right),\ (q=2)$ と同等の収束レートである。2 次元 TPS 推定量でパラメータを全 9 個から 6 個に削減する方法を提案し、収束レートが $O\left(n^{-2/3}\right),\ (q=2)$ で変化せず、分散の定数項が低下することで AMISE が改良できることを示した。また、面積相等性の条件を追加して TPS 推定量を改良することを提案した。その推定精度が $\mathrm{AMISE}(\hat{g}(x,y)) = O\left(n^{-3/4}\right),\ (q=3)$ で、TPS 推定量より優れた推定精度であることを示した。

参考文献

[1] Hjort, N. L. (1986), “On Frequency Polygons and Average Shifted Histograms in Higher Dimensions”, Tech Report 22, Stanford University.

[2] Scott, D. W. (1985), “Frequency Polygons: Theory and Application”, *Journal of the American Statistical Association*, Vol. 80, No. 390, pp.348-354.

2nd order correction term を最小化する分割表の ϕ-ダイバージェンス独立性検定統計量

元北海道教育大札幌校　種市信裕
北海道教育大釧路校　関谷祐里

1. ϕ-ダイバージェンスに基づく分割表独立性検定統計量.

多次元の $J_1 \times J_2 \times \cdots \times J_M$ 分割表において，サンプルサイズ n が固定されている多項分布モデルの場合を考える。すなわち，$(j_1, \ldots, j_M)$ セル $(j_m = 1, \ldots, J_m; m = 1, \ldots, M)$ の観測度数を $X_{j_1,\ldots,j_M}$ とし，$X^* = (X_{11\cdots1}, X_{11\cdots12}, \ldots, X_{11\cdots1J_M}, \ldots, X_{J_1J_2\cdots J_M})'$ とおく。ただし，$\sum_{j_1=1}^{J_1} \cdots \sum_{j_M=1}^{J_M} X_{j_1\cdots j_M} = n$ である。すると，X^* は多項分布 $\mathrm{Mult}_K(n, p^*)$ に従って分布する。ただし，K は分割表のセル数，すなわち $K = \prod_{m=1}^{M} J_m$, であり，$p^* = (p_{11\cdots1}, p_{11\cdots12}, \ldots, p_{11\cdots1J_M}, \ldots, p_{J_1J_2\cdots J_M})'$, $0 < p_{j_1\cdots j_M} < 1$, $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$, $\sum_{j_1=1}^{J_1} \cdots \sum_{j_M=1}^{J_M} p_{j_1\cdots j_M} = 1$ である。M 次元分割表の完全独立性の仮説は，$H_0^{(I)} : p_{j_1\cdots j_M} = P(j_1, \ldots, j_M)$, $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ である。ただし，$P(j_1, \ldots, j_M) = \prod_{m=1}^{M} p_{\cdot(m,j_m)}$, $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$, であり，記号 $a_{\cdot(m,j_m)}$ は $a_{j_1\cdots j_m\cdots j_M}$, $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ を，M 個の添え字をもつ列とするとき，$a_{\cdot(m,j_m)} = \sum_{j_1=1}^{J_1} \cdots \sum_{j_{m-1}=1}^{J_{m-1}} \sum_{j_{m+1}=1}^{J_{m+1}} \cdots \sum_{j_M=1}^{J_M} a_{j_1\cdots j_m\cdots j_M}$, $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ で与えられる。$H_0^{(I)}$ を検定するための ϕ-divergence に基づく統計量 C_ϕ は，$C_\phi = 2n \sum_{j_1=1}^{J_1} \cdots \sum_{j_M=1}^{J_M} \hat{P}(j_1, \ldots, j_M) \phi(\hat{p}_{j_1\cdots j_M} / \hat{P}(j_1, \ldots, j_M))$ で定義される。ここで，$\phi(t)$ は $t > 0$ で定義され $\phi(1) = \phi'(1) = 0, \phi''(1) = 1$ (Pardo et al. [2]) を満たす実凸関数であり，$\hat{P}(j_1, \ldots, j_M) = \prod_{m=1}^{M} \hat{p}_{\cdot(m,j_m)}$, $\hat{p}_{j_1\cdots j_M} = X_{j_1\cdots j_M}/n$, $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$, $\hat{p}_{\cdot(m,j_m)} = X_{\cdot(m,j_m)}/n$, $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ である。Zografos [4] は $M = 2$ のときにこの統計量を示した。完全独立性の仮説 $H_0^{(I)}$ のもとで，統計量 C_ϕ はすべて，自由度 $\eta = K - \sum_{m=1}^{M} J_m + M - 1$ の極限カイ二乗分布をもつ。$\phi(t)$ として，関数 $\phi_a(t) = \{a(a+1)\}^{-1}\{t^{a+1} - t + a(1-t)\}, (a \neq 0, -1); = t\ln t + 1 - t, (a = 0); = -\ln t - 1 + t, (a = -1)$ を選ぶと，C_{ϕ_a} は，Cressie & Read [1] により提案されたパワーダイバージェンスに基づく統計量 $R_{(I)}^a$ になる。$R_{(I)}^0$ は対数尤度比統計量，$R_{(I)}^1$ は Pearson X^2 統計量，$R_{(I)}^{2/3}$ は Cressie & Read [1] によって推奨された統計量である。ϕ として $\phi_a^Q(t) = (t-1)^2/[2\{a + t(1-a)\}], a \in [0, 1]$ を選ぶと Rukhin [3] により提案された統計量 $Q_{(I)}^a$ となる。$Q_{(I)}^1$ は Pearson X^2 統計量である。

2. 2nd-order correction term.

統計量 C_ϕ の帰無仮説 $H_0^{(I)}$ のもとでの原点のまわりの s 次モーメントの評価のための展開式 $E((C_\phi)^s | H_0^{(I)}) = E((F_\eta)^s) + m_\phi^C(s)/n + o(n^{-1}), \quad (s = 1, 2, \ldots)$ を考える。ここで，F_η は，自由度 η のカイ二乗分布に従う確率変数とする。この時，$m_\phi^C(s)$ は，**2nd-order correction term** と呼ばれる。もし，$m_\phi^C(s)$ が 0 に近ければ，その統計量の分布が自由度 η のカイ二乗分布に近いと考えることができる。故に，$m_\phi^C(s)$ の値を計算することによって，どの統計量がカイ二乗分布に近いかを調べることができる。M 次元分割表における完全独立性検定統計量 C_ϕ はモーメントが存在しない場合

がある。しかしながら，$(C_\phi)^s = \infty$ になる確率は n の増加に対して非常に速く 0 に近づく。つまり，大きなサンプルサイズ n に対して，C_ϕ の n^{-1} のオーダーまでの展開式を C_ϕ^* とおくとき，C_ϕ の分布は，C_ϕ^* の分布に非常によく似ていると考えられる。よってここでは，s 次モーメントの評価のための展開式において C_ϕ のかわりに C_ϕ^* とおいた場合の 2nd-order correction term を $m_\phi^C(s)$ とする。

3. 多次元分割表における完全独立性検定の場合の 2nd-order correction term に関する定理および導出される検定統計量.

m_ϕ^C に関して次の定理を得る。

定理 1. C_ϕ^* の分布は連続分布で，$\phi(t)$ を 4 回微分可能，$\phi^{(4)}(t)$ が $t = 1$ で連続で，$p_{j_1 \cdots j_M} = O(K^{-1})$, $(j_m = 1, \ldots, J_m; m = 1, \ldots, M)$ とする。このとき，ϕ-ダイバージェンス統計量に対する展開式 C_ϕ^* に対して，$m_\phi^C(s) = 0$, $(s = 1, 2, \ldots)$ を満たす ϕ について，$H_0^{(I)}$ のもとで，$J_1, \ldots, J_M$ が同程度のオーダーでかつ $K = \prod_{m=1}^M J_m$ が無限大に発散するとき，$4\phi'''(1) + 3\phi^{(4)}(1)$ は 0 に収束する。

$m_\phi^C(s)$ の評価に関して以下の定理を得る。

定理 2. C_ϕ^* の分布は連続分布で，$\phi(t)$ を 4 回微分可能，$\phi^{(4)}(t)$ が $t = 1$ で連続で，$p_{j_1 \cdots j_M} = O(K^{-1})$, $(j_m = 1, \ldots, J_m; m = 1, \ldots, M)$ とする。このとき，$4\phi'''(1) + 3\phi^{(4)}(1) = 0$ を満たす ϕ に対して，$m_\phi^C(s)$ は，$m_\phi^C(s) = -(s/3)\phi'''(1)K^s \sum_{m=1}^M S_m/J_m + O(K^s) + O(K^{s-1}(\sum_{m=1}^M J_m)^2)$, $(s = 1, 2, \ldots)$ と評価される。

定理 1, 定理 2 より，より速くカイ二乗分布に収束する検定統計量を得るために，$4\phi'''(1) + 3\phi^{(4)}(1) = 0$ かつ $\phi'''(1) = 0$，つまり，$\phi'''(1) = \phi^{(4)}(1) = 0$ を満たす関数 ϕ がもし存在するならば，その ϕ により定義される検定統計量 C_ϕ を推奨する。このような性質は，パワーダイバージェンス統計量および Rukhin による統計量の場合には，$R_{(I)}^1$, $Q_{(I)}^1$ (共に Pearson X^2 統計量）のみが満たす。一方，$\psi_a(t) = (3/2)\phi_a(t) - (1/2)\phi_a^Q(t)$ とおき，$a = 2/3$ の場合を $\phi_{NT}(t)$ とする，つまり，$\phi_{NT}(t) = \psi_{2/3}(t) = (27/20)t^{5/3} - 3t - (27/4)(t+2)^{-1} + 39/10$ とおき，これを ϕ とすると上述の性質，つまり「$\phi(1) = \phi'(1) = \phi'''(1) = \phi^{(4)}(1) = 0$, $\phi''(1) = 1$, かつ $t > 0$ において凸」を満たす。よってこの ϕ_{NT} を用いることによって，推奨される検定統計量を構築することができる。本報告では，構築された統計量と Pearson X^2 統計量を含めた他の検定統計量の性能を検討する。

参考文献

[1] Cressie, N., and Read, T. R. C. *JRSS, B,* **46**, (1984), 440–464.

[2] Pardo, L., Pardo, M. C., and Zografos, K. *JJSS,* **29**, (1999), 213–228.

[3] Rukhin, A. L. *12th Prague Conference on Information Theory,* (1994), 214–219.

[4] Zografos, K. *International Journal of Mathematics and Statistical Science* **2**, (1993), 5–12.

境界バイアスを回避する非対称カーネル法の様々な適用について

柿沢佳秀（北海道大学大学院経済学研究院）

1．はじめに　データ $\{X_1,\dots,X_n\}\sim$ iid F に基づいて，$f(=F')$，F，$f^{(r)}$ を推定したい．原点対称で 2 乗可積分なカーネル k (ただし，$\int_{-\infty}^{\infty}k(t)\,dt=1$，$\int_{-\infty}^{\infty}tk(t)\,dt=0$，$\int_{-\infty}^{\infty}t^2k(t)\,dt\neq 0$ を仮定する) に対して，バンド幅を h とした尺度型カーネル $k_h(s)=k(s/h)/h$ を導入する．

標準的なカーネル密度推定量 (Rosenblatt(1956)，Parzen(1962))

$$\widehat{f}_h^{RP}(x)=n^{-1}\sum_{i=1}^{n}k_h(x-X_i)$$

とそれを積分または微分して得られる分布推定量 (Nadaraya(1964))

$$\widehat{F}_h^{RP}(x)=\int_{-\infty}^{x}\widehat{f}_h^{RP}(y)\,dy=n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{x-X_i}k_h(t)\,dt\,,$$

密度微分推定量 (Bhattacharya(1967))

$$\widehat{f}_h^{(r)RP}(x)=\frac{d^r}{dx^r}\,\widehat{f}_h^{RP}(x)=n^{-1}\sum_{i=1}^{n}k_h^{(r)}(x-X_i)$$

の漸近特性は $\mathrm{supp}(f)$ の内点 x でのみ保証される．もし $\mathrm{supp}(f)\neq\mathbb{R}$ なら，一般に，境界点の近くで一致性がない．実際，本報告の非負データ設定 $\mathrm{supp}(f)=[0,\infty)$ において $u=f^{(r)}$ (or F) を考えるとき，畳み込み積分 $(k_h*u)(x)$ は

$$\int_{-\infty}^{\infty}\frac{1}{h}\,k\Big(\frac{x-s}{h}\Big)u(s)\,ds=\int_{0}^{\infty}\frac{1}{h}\,k\Big(\frac{x-s}{h}\Big)u(s)\,ds=\int_{-\infty}^{\frac{x}{h}}k(t)u(x-ht)\,dt$$

のように変形され，非負カーネル関数 k が $[-1,1]$ (say) の台を持つ場合，$h\to 0$ のとき

$$\begin{aligned}&\int_{-1}^{\min(\frac{x}{h},1)}k(t)u(x-ht)\,dt-u(x)\\&\quad\approx\begin{cases}\dfrac{1}{2}\,h^2u''(x)\displaystyle\int_{-1}^{1}t^2k(t)\,dt\,, & x\in[h,\infty)\,,\\ -u(0)\displaystyle\int_{p}^{1}k(t)\,dt+hu'(0)\int_{p}^{1}(-p+t)k(t)\,dt\,, & x=hp\ (p\in[0,1])\end{cases}\end{aligned}$$

となり，$u(0)>0(<0)$ ならば，$x=0$ の近くで深刻な $O(1)$ の下方 (上方) バイアスがある．一方，例外的には，$u(0)=0$ のとき漸近不偏で一致性があるが，それでも，$u(0)=0$，かつ，$u'(0)\neq 0$ ならば，$x=0$ の近くで $O(h)$ のバイアスがある (そのオーダーは，境界点から離れた位置での $O(h^2)$ よりも遅い)．

この “境界バイアス問題” を対処する研究もあり，台の変換法 (Marron and Ruppert(1994)) の他，境界カーネル法 (renormalization 法，reflection 法，一般化ジャックナイフ法は，例えば，Jones(1993) を参照) が議論された．一方，Silverman(1986; p28) は，単純な 1 つのアイディアとして，推定したい関数の台とマッチするような (可変的) カーネルを選ぶべきと指摘している．

有界区間 $[0,1]$，半無限区間 $[0,\infty)$ のデータに対する，このアプローチは Chen(1999,2000) がベータ/ガンマカーネル密度推定量を提案した以降，“非対称カーネル法” と知られている．

2．非負データに対する非対称カーネル法　本報告で，非負データ $\{X_1,\ldots,X_n\}\sim$ iid F に基づく $f(=F')$ と F の非対称カーネル推定法について，先行研究をまずレビューし，次に $f^{(r)}$ の推定量の漸近特性を調べた (この構成のアイディアは先行研究と異なることに注意する)．以下，バイアスと分散を制御する平滑化パラメータを b とする．

2.1．非対称カーネル　Chen(2000) のガンマカーネル以降，種々の非対称カーネルが個別に提案され，現在オプションは豊富になっている．本報告では，例として，対称密度 $C_g\,g(u^2)$ を土台とした非心 qBS カーネル族と qIG/qRIG カーネル族 (Kakizawa(2018,2021)) を紹介した．

2.2．非対称カーネル密度推定量　非対称カーネル密度推定量 $\widehat{f}_b(x)=n^{-1}\sum_{i=1}^n k(X_i;b,x)$ の漸近特性は Kakizawa(2021) などを参照．

2.3．非対称カーネル分布/密度微分推定量のナイーブ版　ナイーブには，“$\widehat{f}_b(\cdot)$ を積分 (or 微分) すればよい” と思うに違いない，すなわち，

$$\widetilde{F}_b(x)=\int_0^x \widehat{f}_b(y)\,dy=n^{-1}\sum_{i=1}^n\int_0^x k(X_i;b,y)\,dy\,,$$
$$\widetilde{f}_b^{(r)}(x)=\frac{d^r}{dx^r}\,\widehat{f}_b(x)=n^{-1}\sum_{i=1}^n\frac{\partial^r}{\partial x^r}\,k(X_i;b,x)$$

という推定量もありえるが，以下の理由から，これらは推奨されない．

- 前者 (Chekkal et al.(2023)) では y 積分が数値積分にならざるをえない；
- ガンマカーネルを用いた $r=1$ の推定量 (Dobrovidov and Markovich(2013)) は高階微分への拡張に不利 (対数ガンマ関数の高階導関数が関与し，バイアス導出も見通しが悪い)．

2.4．非対称カーネル分布/密度微分推定量　$\int_0^x k(s;b,y)\,dy$ の替わりに，$\int_s^\infty k(t;b,x)\,dt$ を使用し，Mombeni et al.(2021) は非対称カーネル分布推定量

$$\widehat{F}_b(x)=n^{-1}\sum_{i=1}^n\int_{X_i}^\infty k(t;b,x)\,dt$$

を考察している．本報告は，柿沢 (2022; 連合大会/日本数学会秋季総合分科会) による非対称カーネル密度微分推定量

$$\widehat{f^{(r)}}_b(x)=n^{-1}\sum_{i=1}^n(-1)^r k^{(r)}(X_i;b,x)$$

に対して，バイアスが 2.2 節の結果で f,f',f'' を $f^{(r)},f^{(r+1)},f^{(r+2)}$ に読み替えて得られることを指摘した．さらに，(少なくとも) 標準正規密度を土台とした中心 qBS カーネルの場合で漸近分散も明示的に与えた．

3．おわりに　積型カーネルの使用から，多次元化は容易であるが，非積型カーネルを使うこともできる (Kakizawa(2022) を参照)．実際，qBS 型密度は，分布論として楕円密度を土台にした定式化が可能で，多変量正規密度を除く楕円密度を採用すれば，必然的に非積型となる．

Some estimators in the ADCINAR(1) process

Xiaoqiang Zeng and Yoshihide Kakizawa
Graduate School of Economics and Business Administration, Hokkaido University

1 Introduction

The analysis of count time series has rapid progress. There is a huge literature about formulation of models, probabilistic aspects, and statistical inference (see Weiß (2008,2018)). Al-Osh and Alzaid (1987) proposed nonnegative integer-valued autoregressive process of the first-order (INAR(1)), based on the binomial thinning operator due to Steutel and van Harn (1979). Du and Li (1991) gave the stationary condition of the pth-order INAR process, and proved its ergodicity. Silva and Oliveira (2004) discussed some moment structures for the stationary INAR(1) process under the Poisson innovation, focusing on Yule–Walker (YW) type equation of the third raw automoment (and autocumulant) function. Schweer and Weiß (2016) derived, for any positve integer r, the $(r+1)$th autocumulant function of the stationary INAR(1) process under the Poisson innovation. Under a general innovation, Zeng and Kakizawa (2022) gave the third, fourth, fifth, and sixth autocumulant functions.

Nastić et al. (2017) (see also Ristić et al. (2013)) not only introduced an alternative dependent counting nonnegative INAR process of the first-order (ADCINAR(1)), with a modification of the original Steutel and van Harn's (1979) thinning operation, but also considered the estimation of the parameter α and the new parameter ϑ. Recently, even for the stationary ADCINAR(1) process, Zeng and Kakizawa (2023) studied higher autocumulant functions, except for the autocovariance function. Since the asymptotic normality of the estimator for ϑ was established by Nastić et al. (2017) under an unrealistic assumption, i.e., α is known, our first goal is to revisit such an asymptotic theory.

In the statistical analysis of the count time series data, the conditional least squares (CLS) method due to Klimko and Nelson (1978) has been widely used. See, e.g., Al-Osh and Alzaid (1987) and Park and Oh (1997) for the stationary INAR(1) process, and Nastić et al. (2017) for the stationary ADCINAR(1) process. The YW estimator for α can be applied easily, since α is the autocorrelation at lag 1 of various INAR(1) type models, as in the usual stationary AR(1) process (e.g., Brockwell and Davis (1987)). These estimators have desirable asymptotic properties, but are biased in a finite-sample. Some authors derived asymptotic expansions of the biases in order to perform an analytical bias-correction for the stationary INAR(1) process (see Bourguignon and Vasconcellos (2015), Weiß and Schweer (2016), and Zeng and Kakizawa (2022)). Our second goal is to develop the bias-corrections of the CLS and YW estimators for α in the stationary ADCINAR(1) process.

2 ADCINAR(1) process

As usual, $B(p)$ is the Bernoulli random variable, i.e., $P[B(p)=1] = 1 - P[B(p)=0] = p(\in [0,1])$. Let $S_j(\alpha,\vartheta) = B_j(\vartheta)B(\alpha/\vartheta)$, where $0 \le \alpha \le \vartheta \le 1(\vartheta \ne 0)$ and $\{B_j(\vartheta)\}$ is a sequence of IID Bernoulli random variables, which is independent of $B(\alpha/\vartheta)$. An alternative generalized binomial thinning operator was recently introduced by Nastić et al. (2017), as follows. Given a nonnegative integer-valued random variable Y, let $\alpha \diamond_\vartheta Y = \sum_{j=1}^{Y} S_j(\alpha,\vartheta)$, $Y = 1,2,\ldots$ (when $Y=0$, $\alpha \diamond_\vartheta Y = 0$), where $\{B_j(\vartheta)\}$ and $B(\alpha/\vartheta)$ are independent of Y. Note that $S_j(\alpha,\vartheta)$ is distributed as the Bernoulli distribution $\mathrm{Bin}(1,\alpha)$, whereas $\{S_j(\alpha,\vartheta)\}$ is, in general, a dependent sequence, i.e., for $i \ne j$, $Cov[S_i(\alpha,\vartheta), S_j(\alpha,\vartheta)] = \alpha(\vartheta - \alpha)$.

Nastić et al. (2017) thus defined the ADCINAR(1) process by $Y_t = \alpha \diamond_\vartheta Y_{t-1} + \varepsilon_t, t = 0, \pm 1, \ldots$, where an innovation $\{\varepsilon_t\}$ is a sequence of IID nonnegative integer-valued random variables, such that ε_t and Y_{t-i} are independent for all integer t and positive integer i (it is implicitly assumed that the operations $\alpha \diamond_\vartheta$ at different times are performed mutually independently, which are also independent of $\{Y_t\}$ and $\{\varepsilon_t\}$). Note that the INAR(1) process is a special case of the ADCINAR(1) process with $\vartheta = \alpha (\in (0, 1])$. As mentioned in Nastić et al. (2017), the ADCINAR(1) process $\{Y_t\}$ is strictly stationary and ergodic when $0 \le \alpha \le \vartheta < 1 (\vartheta \ne 0)$. We assumed this assumption throughout this paper.

3 Estimation of α and ϑ

Minimizing the criterion $J(\alpha, \mu_\varepsilon) = \sum_{t=2}^n (Y_t - E[Y_t|Y_{t-1}])^2 = \sum_{t=2}^n (Y_t - \alpha Y_{t-1} - \mu_\varepsilon)^2$, the CLS estimator $\widehat{\alpha}_{CLS}$ for α can be obtained (see also Klimko and Nelson (1978) and Al-Osh and Alzaid (1987)). Further, a general estimator $\widehat{\alpha}_{c_1,c_2}$ for α is available, where the YW estimator is given by $\widehat{\alpha}_{1,1}$ (see Zeng and Kakizawa (2022)). The estimators $\widehat{\alpha}_{CLS}$ and $\widehat{\alpha}_{c_1,c_2}$ have the desirable asymptotic properties (strong consistency and asymptotic normality). On the other hand, the criterion $J(\alpha, \mu_\varepsilon)$ is free of the parameter ϑ, hence, another tool, referred to as the two-step CLS (2CLS) method, is needed to estimate ϑ, as in Karlsen and Tjøstheim (1988) (see also Nastić et al. (2017)). Suppose that an estimator $\widehat{\alpha}$ for $\alpha (> 0)$ is available. Define $\overline{Y} = n^{-1} \sum_{t=1}^n Y_t$ and $\widehat{\sigma}_Y^2 = n^{-1} \sum_{t=1}^n (Y_t - \overline{Y})^2$. Then, the 2CLS estimator for ϑ can be constructed in the closed-form, denoted by $\widehat{\vartheta}_{(\widehat{\alpha}, \overline{Y}, \widehat{\sigma}_Y^2)}$ (the detail is omitted here). Under suitable conditions, we showed that the 2CLS estimator $\widehat{\vartheta}_{(\widehat{\alpha}, \overline{Y}, \widehat{\sigma}_Y^2)}$ is strongly consistent and asymptotically normal.

4 Bias-corrections of the CLS and YW estimators for α

Let $\mu_Y(u, \ell, \ell) = E[(Y_t - \mu_Y)(Y_{t+u} - \mu_Y)(Y_{t+\ell} - \mu_Y)^2]$ be fourth automoment function at lag (u, ℓ, ℓ) of $\{Y_t\}$, where $u = 0, 1$ and $\ell = 1, \ldots, n-1$. Recall $\widehat{\alpha}_{YW} = \widehat{\alpha}_{1,1}$. In the same way as Zeng and Kakizawa (2022), it can be shown that $\widehat{\alpha}_{CLS} = \widehat{\alpha}_{1,0} + O_p(n^{-2})$, and that the bias of the estimator $\widehat{\alpha}_{c_1,c_2}$ is given by $E(\widehat{\alpha}_{c_1,c_2} - \alpha) = -n^{-1}[1 + (c_1 + c_2)\alpha] - (n\sigma_Y^4)^{-1}\{M(1) - \alpha M(0)\} + o(n^{-1})$, where $M(u) = \sum_{\ell=1}^\infty \mu_Y(u, \ell, \ell)$, $u = 0, 1$. We proceeded to estimate $M(1) - \alpha M(0)$ in two different ways. One is based on the analytical evaluation; the other is the nonparametric method applying the lag window-type kernel (note that the former analytical evaluation needs a further extra tedious calculation of $M(1) - \alpha M(0)$).

5 Conclusion and future works

The asymptotic normality of the 2CLS estimator $\widehat{\vartheta}_{(\widehat{\alpha}, \overline{Y}, \widehat{\sigma}_Y^2)}$ for ϑ was revisited. We proposed the lag window-type bias-correction and the analytical bias-correction of the commonly used CLS and YW estimators for α in the stationary ADCINAR(1) process and then conducted some simulations.

Bootstrap and jackknife procedures are alternative methods of constructing bias-corrected estimators. Jentsch and Weiß (2019) proved the validity of bootstrapping for the stationary INAR process, so that the extension of thier result to the stationary ADCINAR(1) process was left for a future topic.

Quantification and Formulation of Galaxy Evolution by Manifold Learning

Tsutomu T. Takeuchi[1,2]*, Suchetha Cooray[3]†, Hai-Xia Ma[1], Kiyoaki Christopher Omori[1]‡, Wen Shi[1], Ryusei R. Kano[1], Sena A. Matsui[1], and Aina May So[4,1]

1. Division of Particle and Astrophysical Science, Nagoya University, Japan,
2. Research Center for Statistical Machine Learning, Institute of Statistical Mathematics, Japan,
3. National Astronomical Observatory of Japan,
4. Department of Physics, Gakushuin University, Japan

Matter in the early Universe was almost uniform. Slightly dense regions have grown by gravity, and finally turned into galaxies. Astrophysicists have searched for the equations governing the complex physical phenomena of galaxy formation and evolution over 13 billion years of the cosmic age, from the first principles of physics. We, instead, elucidate the physics of galaxy evolution by applying manifold learning, one of the latest methods of data science, to a feature space spanned by galaxy luminosities and cosmic time. We have discovered that galaxies consist a complex but low-dimensional subset in an ultra-high-dimensional feature space (Siudek et al., 2018). We refer to it as the galaxy manifold in a modern sense. The discovered galaxy manifold by Siudek et al. (2018) has a strongly nonlinear one-dimensional structure. Further, surprisingly, the spectra of the sample galaxies were discriminated only by a few broadband luminosities, *not* complicated combinations of quantities. This suggests that the galaxy evolution at optical wavelengths can be described by only a few parameters at most. This is a new characterization of galaxy evolution that could have never been found by conventional methods. We will further pursue the galaxy manifold and elucidate the dependence of parameters (probably several at most) that govern the physics of galaxy evolution from the galaxy manifold, and derive the governing equations for galaxy evolution.

We then applied the unsupervised machine learning to express it even in an easier way. This is a part of the technique referred to as the dimensionality reduction. We used the method topological data analysis to elucidate the (possibly several at most) parameter dependence that governs the physics of galaxy evolution from galaxy manifolds. Especially, a method known as the manifold learning is optimal for parametrizing the galaxy manifold. Multimodality, and dispersions in classical scaling laws are merely a consequence of unsuitable projection of the galaxy manifold. We make use of this method to construct the governing equation(s) of galaxy evolution.

After obtaining the galaxy manifold, for example, observational quantities as luminosities missing for objects, or more physical quantities as SFR and stellar mass M_* can be straightforwardly estimated from the location on the manifold. We adopt the algorithm `Isomap` Tenenbaum et al. (2000) and `UMAP` (Uniform Manifold Approximation and Projection) McInnes et al. (2018, 2020). `Isomap` defines the neighboring points by using input-space distance, and the distant points as a sequence of "short hops" between neighboring points. `Isomap` tries to find shortest paths in a graph with edges connecting neighboring data points By construction, `Isomap` preserves the "surface density" of data points in the feature space. `UMAP` is based on differential geometry and algebraic topology. The algorithm is founded on three assumptions: 1) the data are uniformly distributed on a Riemannian manifold, 2) the Riemannian metric is locally constant (or can be approximated as such),

*E-mail: tsutomu.takeuchi.ttt@gmail.com.
†JSPS Fellow (PD)
‡JSPS Fellow (DC2)

and 3) the manifold is locally connected. From these assumptions it is possible to model the manifold with a fuzzy topological structure. Sine it defines the manifold so that the data points distribute as homogeneously as possible, it does not preserve the surface density of data points. `UMAP` also preserves some important structural properties, and it is more robust against noise than `Isomap`. Manifold learning algorithm can "unfold" a curved and/or rolled manifold in the feature space, and provide a local coordinate system on it Tenenbaum et al. (2000); Roweis & Saul (2000). We also stress that two different algorithms, `Isomap` and `UMAP` yield similar manifolds. Since `Isomap` preserves the density of data point cloud, we observe that the manifold has a density structure, i.e., dense and sparse regions on the manifold.

To demonstrate that the manifolds contain important information on galaxy evolution, we compared SFR and stellar mass as a function on the manifold. The behaviors of SFR and stellar mass are both qualitatively very similar, suggesting that the estimated manifold structure is robust. This means that the manifold learning actually "learns" the important characteristics of galaxy evolution at optical wavelengths. Since we could connect the manifold and some physical quantities such as SFR and stellar mass, we can also parametrize the evolution of galaxies on the manifold. As stars are formed, the stellar mass, namely the total mass of accumulated stars will increase. This is one of the fundamental aspect of galaxy evolution, and we can visualize this evolution as a vector field on the manifold.

Last question is how to describe and interpret the evolutionary track of galaxies on the manifold. We applied a classical theoretical model of the chemical evolution of galaxies. Chemical evolution is a field of galactic physics that treats the formation and evolution of elements in galaxies based on the stellar evolution theory. This strange terminology comes from the fact that the theory has been used to analyze the chemical composition of stars and the interstellar medium (ISM) in a galaxy. The important physical process is the nucleosynthesis via nuclear fusion in the center of stars. We adopt a simple model with infall and outflow of matter proposed by Lilly et al. (2013).

$$M_*(t_{n+1}) = M_*(t_n) + (1-r)\mathrm{SFR}(t_n)\Delta t \ , \tag{1}$$

$$M_{\mathrm{gas}}(t_{n+1}) = M_{\mathrm{gas}}(t_n) - (1-r+\eta)\mathrm{SFR}(t_n)\Delta t \tag{2}$$

where r is the returning mass fraction, η is the mass loading factor, and Δt is the timestep (Cooray et al., 2023). We calculate theoretical evolutionary tracks of galaxies from eqs. (1) and (2). Interpretation of the vector field on the galaxy manifold can be given by comparing eqs. (1) and (2) with the vector field on the galaxy manifold. However, a sophisticated methodology is desired to interpret the observed manifold more directly.

We restricted our discussion to multiwavelength photometric surveys, but this method can in principle be extended to spectroscopic surveys. In this work, we have discussed the optical properties, but also we can include any other properties of galaxies, not only luminosities at other wavelengths but also dynamical, structural, and environmental properties. Manifold learning will provide a fundamentally new insight to the studies on the formation and evolution of galaxies.

References

Cooray, S., Takeuchi, T. T., Kashino, D., et al. 2023, Mon. Not. R. Astron. Soc. , 524, 4976

Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, Astrophys. J. , 772, 119

McInnes, L., Healy, J., & Melville, J. 2020, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, arXiv:1802.03426

McInnes, L., Healy, J., Saul, N., & Großberger, L. 2018, Journal of Open Source Software, 3, 861

Roweis, S. T., & Saul, L. K. 2000, Science, 290, 2323

Siudek, M., Małek, K., Pollo, A., et al. 2018, Astronomy and Astrophysics, 617, A70

Tenenbaum, J. B., Silva, V. d., & Langford, J. C. 2000, Science, 290, 2319

GMANOVA モデルにおけるフルランク仮定が不要な罰則無推定法

(報告書)

中京大学 教養教育研究院 永井 勇

本講演では n 個の各個体に対して，時間と共に p 回測定されたデータは経時測定データの分析を考えていた．このような経時測定データにおいて，全ての個体で測定した時点が揃っている場合，Potthoff and Roy (1964) により提案された次の一般化多変量分散分析（GMANOVA）モデルでの分析がよくされていることを紹介した;

$$\boldsymbol{Y} = \mathbf{1}_n \boldsymbol{\mu}' \boldsymbol{X}' + \boldsymbol{A}\boldsymbol{\Xi}\boldsymbol{X}' + \boldsymbol{\mathcal{E}}, \tag{1}$$

ここで，$\boldsymbol{Y}$ は各行が各個体の経時測定データからなる $n \times p$ 既知行列，$\mathbf{1}_n$ は全ての要素が 1 の n 次元のベクトル，$\boldsymbol{\mu}$ は q 次元未知ベクトル，$\boldsymbol{X}$ は個体内計画行列と呼ばれる測定時点から作られる $p \times q$ 既知行列 (詳細は後述)，$\boldsymbol{A}$ は個体間計画行列と呼ばれる測定時点に無関係な各個体の特徴を表す変数からなる $n \times k$ 既知行列で各列で中心化されているとし，$\boldsymbol{\Xi}$ は $k \times q$ 未知行列であり，$\boldsymbol{\mathcal{E}}$ は $E[\boldsymbol{\mathcal{E}}] = \mathbf{0}_n \mathbf{0}_p'$ で $\mathrm{Var}[\mathrm{vec}(\boldsymbol{\mathcal{E}})] = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$ の $n \times p$ 誤差行列であり，$\mathbf{0}_r$ は全ての要素が 0 の r 次元ベクトル，$\boldsymbol{\Sigma}$ は未知の $p \times p$ 正則行列である．

$\boldsymbol{Y}$ の経時的な変動 (経時変動) を推定する際には，個体内計画行列 $\boldsymbol{X}$ の i 行目として，i 番目の測定時点 t_i $(i = 1, \ldots, p,\ t_1 < \cdots < t_p)$ の関数からなる行列が用いられることを紹介した．特に，本講演では分かりやすさのために，$\boldsymbol{X}$ の i 行目として $(t_i^0, \ldots, t_i^{q-1})$ を用いて未知の $\boldsymbol{\mu}$ や $\boldsymbol{\Xi}$ を推定することで，切片と各次数の係数に対応した部分が推定でき，経時変動を t_i の $(q-1)$ 次多項式で推定できることを例示した．

このモデル (1) において，$\boldsymbol{\mu}$ や $\boldsymbol{\Xi}$ の推定は次のリスクを最小にすることで行われることを紹介した;

$$\mathrm{tr}\{(\boldsymbol{Y} - \mathbf{1}_n \boldsymbol{\mu}' \boldsymbol{X}' - \boldsymbol{A}\boldsymbol{\Xi}\boldsymbol{X}')\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{1}_n \boldsymbol{\mu}' \boldsymbol{X}' - \boldsymbol{A}\boldsymbol{\Xi}\boldsymbol{X}')'\}. \tag{2}$$

このリスクを最小にする $\boldsymbol{\mu}$ や $\boldsymbol{\Xi}$ を求める際に，従来は $\mathrm{rank}(\boldsymbol{A}) = k$ かつ $\mathrm{rank}(\boldsymbol{X}) = q$ を仮定していることを紹介し，本講演ではこれらの仮定が満たされない場合の推定法を提案した．そこでは，永井 (2021) のアイデアを用いて，シンプルに $\boldsymbol{A}$ や $\boldsymbol{X}$ を $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_{d_1})$ や $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{d_2})$ と分割し，$\boldsymbol{\mu}$ や $\boldsymbol{\Xi}$ をそれぞれの大きさに合わせて $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{d_2})$ や $\boldsymbol{\Xi}$ の j 行目を $(\boldsymbol{\Xi}_{j1}, \ldots, \boldsymbol{\Xi}_{jd_2})$ $(j = 1, \ldots, d_1)$ と分割することを考えた．この分割した行列などを用いて，$\boldsymbol{\mu}$ や $\boldsymbol{\Xi}$ を推定する代わりに $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{d_2}$ や $\boldsymbol{\Xi}_{11}, \ldots, \boldsymbol{\Xi}_{1d_2}, \ldots, \boldsymbol{\Xi}_{d_1 1}, \ldots, \boldsymbol{\Xi}_{d_1 d_2}$ を推定することで，$\mathrm{rank}(\boldsymbol{A}) = k$ かつ $\mathrm{rank}(\boldsymbol{X}) = q$ より非常に緩い仮定の下で，罰則を付けずにリスク (2) を最小にする $\boldsymbol{\mu}$ や $\boldsymbol{\Xi}$ の推定法を提案した．さらに，分割した行列を用いてリスクを直接最小にする $\boldsymbol{\Xi}_{ij}$ $(i = 1, \ldots, d_1;\ j = 1, \ldots, d_2)$ などを求めると，それらを得るために求めようとしている部分以外の全ての部分が必要となり，初期値や反復計算が必要となることを紹介した．この問題を解決するため，リスクの展開を工夫した形を構築し，それに基づいた推定法を提案した．

引用文献:

[1] Katayama, S. & Imori, S. (2014) Lasso penalized model selection criteria for high-dimensional multivariate linear regression analysis. *J. Multivariate Anal.*, **132**, 138–150.

[2] 永井 勇 (2021) 高次元小標本における多変量線形回帰モデルでの推定法，2021 年度統計関連学会連合大会 講演.

[3] Potthoff, R. F. & Roy, S. N. (1964) A generalized multivariate analysis of variance model usefule specially for growth curve problems. *Biometrika*, **51**, 313–326.

[4] Srivastava, M. S. (2002) *Methods of Multivariate Statistics*, Wiley-Interscience.

[5] Yanagihara, H., Nagai, I., Fukui, K. & Hijikawa, Y. (2023) Modified Cp criterion in widely applicable models. *Smart Innov. Syst. Tec.*, **352**, Intelligent Decision Technologies 2023: Proceedings of the 15th KES International Conference on Intelligent Decision Technologies KES-IDT-2023 (eds. I. Czarnowski, R. J. Howlett & L. C. Jain), 173-182.

Two step estimations via the Dantzig selector for models of stochastic processes with high-dimensional parameters

Kou Fujimori[1] and Koji Tsukuda[2]

[1]Faculty of Economics and Law, Shinshu University.
[2]Faculty of Mathematics, Kyushu University.

Let $\theta \in \Theta \subset \mathbb{R}^p$ be an unknown parameter of interest and $h \in H$ a possibly infinite-dimensional nuisance parameter where H is a metric space equipped with a metric d_H. Consider the following random maps:

$$\psi_n^{(1)} : \Theta \to \mathbb{R}^p, \quad \tilde{\psi}_n^{(1)} : \Theta \to \mathbb{R}^p,$$

$$\Psi_n : \Theta \times H \to \mathbb{R}^p, \quad \tilde{\Psi}_n : \Theta \times H \to \mathbb{R}^p,$$

where n is a number of observations. The random maps $\psi_n^{(1)}$ and Ψ_n are corresponding to some score functions and $\tilde{\psi}_n^{(1)}$ and $\tilde{\Psi}_n$ are their compensators, respectively. We suppose that the true values $\theta_0 \in \Theta$ and $h_0 \in H$ satisfy that

$$\tilde{\psi}_n^{(1)}(\theta_0) \approx 0, \quad \tilde{\Psi}_n(\theta_0, h_0) \approx 0.$$

We are interested in the estimation problem for $\theta_0 = (\theta_{01}, \dots, \theta_{0p})^\top$ under the following high-dimensional and sparse settings;

We first construct an estimator for T_0 to achieve the dimension reduction. To do this, we consider the following Dantzig selector type estimator $\hat{\theta}_n^{(1)}$ for θ_0, based on $\psi_n^{(1)}$:

$$\hat{\theta}_n^{(1)} := \arg\min_{\theta \in \mathcal{C}_n} \|\theta\|_1, \quad \mathcal{C}_n = \{\theta \in \Theta : \|\psi_n^{(1)}(\theta)\|_\infty \leq \lambda_n\},$$

where λ_n is a tuning parameter. Using $\hat{\theta}_n^{(1)}$, we define $\hat{T}_n$ as follows:

$$\hat{T}_n := \{j : |\hat{\theta}_{nj}| > \tau_n\},$$

where τ_n is a threshold level, which is a tuning parameter. Then, we have the following theorem.

Theorem 1. *Under some regularity conditions, it holds that*

$$P\left(\|\hat{\theta}_n^{(1)} - \theta_0\|_\infty > \frac{2}{\delta} c_n\right) \to 0, \quad n \to \infty, \tag{0.1}$$

where $c_n := \|\psi_n^{(1)}(\hat{\theta}_n^{(1)}) - \psi_n^{(1)}(\theta_0)\|_\infty$. *Especially, it holds that*

$$\|\hat{\theta}_n^{(1)} - \theta_0\|_\infty = O_p(\lambda_n), \quad n \to \infty.$$

Moreover, if the threshold τ_n *satisfies that* $4\lambda_n/\delta < \tau_n < \inf_{j \in T_0} |\theta_{0j}|/2$, *then it holds that*

$$P(\hat{T}_n = T_0) \to 1, \quad n \to \infty. \tag{0.2}$$

For every index set T, we consider the following random map restricted by T:

$$\Psi_{nT} : \Theta_T \times H \to \mathbb{R}^{|T|}, \quad \tilde{\Psi}_{nT} : \Theta_T \times H \to \mathbb{R}^{|T|},$$

where Θ_T is a set of sub-vectors of Θ restricted by T. Let $\hat{h}_n$ be an estimator of $h \in H$ such that $d_H(\hat{h}_n, h_0) = o_p(1)$ as $n \to \infty$. Then, we consider the new estimator $\tilde{\theta}_n$ for the parameter θ of interest, with help of $\hat{h}_n$ and $\hat{T}_n$, as a solution to the following equations:

$$\Psi_{n\hat{T}_n}(\tilde{\theta}_{n\hat{T}_n}, \hat{h}_n) \approx 0, \quad \tilde{\theta}_{n\hat{T}_n^c} = 0.$$

Then, we can derive the asymptotic distribution of $\tilde{\theta}_n$, as well as Theorem 2.1 of Nishiyama (2009). In this talk, we apply the general theory discussed above to ergodic time series models and models of diffusion processes.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 21K13271(K.F.), 21K13836 (K.T.).

References

Bühlmann, P. and van de Geer, S. A. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Candes, E. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351.

Huang, J., Sun, T., Ying, Z., Yu, Y., and Zhang, C.-H. (2013). Oracle inequalities for the LASSO in the Cox model. *Ann. Statist.* **41** 1142–1165.

Nishiyama, Y. (2009). Asymptotic theory of semiparametric Z-estimators for stochastic processes with applications to ergodic diffusions and time series. *Ann. Statist.* **37** 3555–3579.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288.

Statistical mathematics of biodiversity indices for spatial species distribution model

Shimatani, Ichiro Ken (The Institute of Statistical Mathematics)

In order to express biodiversity, many indices were proposed. The number of different species (species richness) has been most widely used, and when we have species abundance data, the following two indices have been most commonly used.

Shannon entropy $H = -\Sigma_s p_s \log p_s$,

Simson index $D = 1 - \Sigma_s p_s^2$,

where p_s is the relative frequency of species s (=the probability that a randomly chosen individual is species s).

It is shown that the three indices are unitedly interpreted in the Hill number defined by

$${}^qD = (\sum_{s=1}^{S} p_s{}^q)^{\frac{1}{1-q}}$$

When q = 0, it is identical to the species richness. When q = 1, it is equivalent to the Shannon entropy;

$${}^1D = e^{-\sum_s p_s \ln p_s},$$

and when q = 2, it corresponds to the Simpson index;

$${}^2D = \frac{1}{\sum_{s=1}^{S} p_s{}^2} \quad .$$

Thus, we can comparatively examine biodiversity of two or more communities not by a single number but by curves parametrized by $0 \le q$ (practically, in many case, ≤ 2 or 3), which indicates a balance (weight) between rare species and dominant ones.

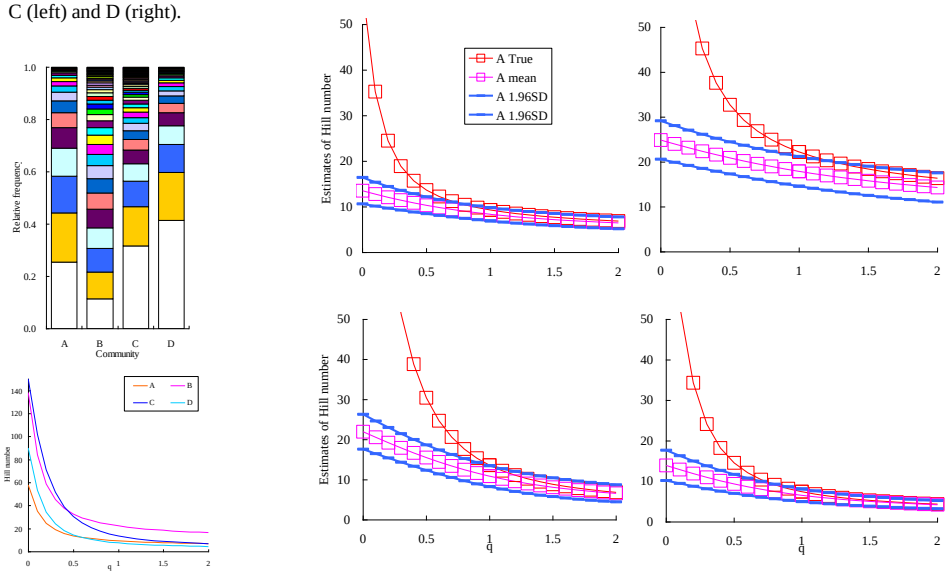
The Hill number, however, does not provide a perfect quantification of biodiversity. The biggest problem is that the Hill number is highly underestimated if observed relative frequencies are inserted into the equation (Fig. 1).

Other indices are also proposed and comparatively examined. An example is the generalized Simpson's entropy defined as;

$$\varsigma_r = \sum_{s=1}^{S} p_s(1-p_s)^r \qquad (\mathrm{r} = 1, 2, \ldots)$$

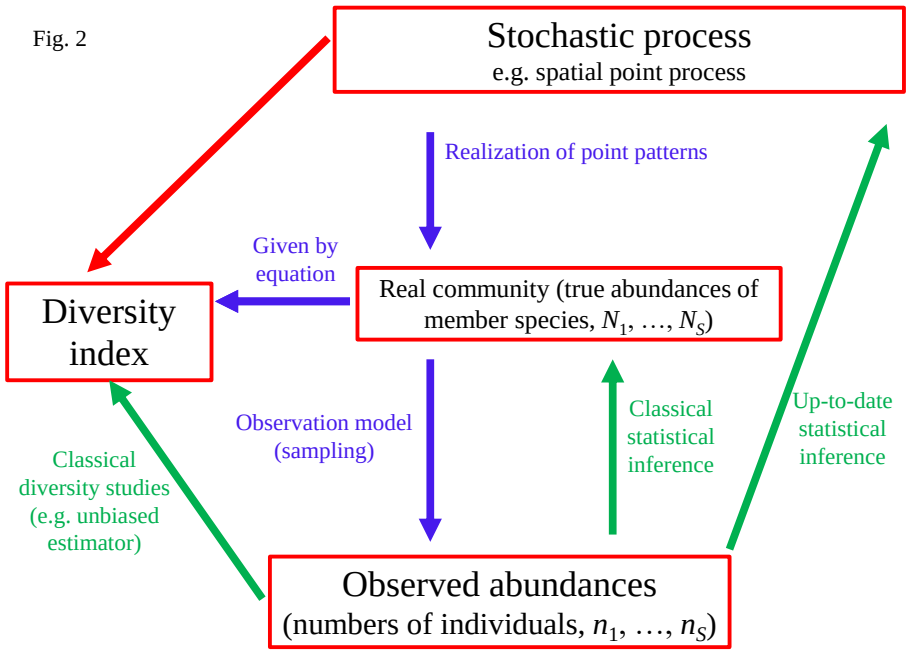
This is interpreted as the probability that (r + 1)st observation will be of a species that has not been observed before, and an unbiased estimator is known.

Fig. 1 Hill number is highly biased if observed relative frequencies are directly used in the equation. Examples using 100 samples. The left panels indicate the given relative frequencies and the Hill numbers for the four ecological communities. The upper right two graphs are for community A (left) and B (right). The lower two are for community C (left) and D (right).



Recently. a goal of community ecology tends to change from finding the "true" community to identifying a hidden stochastic process that controls realized communities and an observed community is nothing more than one realization of the stochastic process (Fig. 2).

Fig. 2



In this regard, the presence of an unbiased estimator may not be so big advantage and a new overall statistical inference is an on-going statistical issue in biodiversity studies..

Simultaneous estimation of Poisson means: recent developments and its applications

Yuan-Tsung Chang (Mejiro University), Shinozaki Nobuo (Keio University)

In estimating $p(\geq 2)$ independent Poisson means, Clevenson and Zidek (1975) have proposed a class of estimators that shrink the unbiased estimator to the origin and dominate the unbiased one under the normalized squared error loss. This class of estimators was subsequently enlarged in several directions. Here, we disscuss

1. the problem and proposes new classes of dominating estimators using prior information pertinently. Dominance is shown by partitioning the sample space into disjoint subsets and averaging the loss difference over each subset. Estimation of several Poisson mean vectors is also discussed. Further, simultaneous estimation of Poisson means under order restriction is treated and estimators which dominate the isotonic regression estimator are proposed for some types of order restrictions.(Chang and Shinozaki (2019)).

2. the shrinkage estimation of Poisson means when observations are given in the form of a two-way contingency table. Assuming a multiplicative Poisson model, estimators which shrink to the specified values or an order statistic in one dimension and in two dimensions are considered and are shown to dominate the maximum likelihood estimator (MLE) under normalized squared error loss. Further, assuming the full model, shrinkage to the multiplicative model is devised to improve upon the unbiased estimator. Shrinkage is made after determining the basic cells so that the observed frequency is not smaller than the estimated frequency for each of the other cells. (Chang and Shinozaki (2022)).

1 Some classes of improved estimators using prior information

Let X_i be distributed as $Po(\lambda_i), i = 1, \ldots, p$, and suppose that $X_1, \ldots, X_p$ are statistically independent. Let the loss function when we estimate $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$ by $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ be the normalized squared error one

$$L(\boldsymbol{\lambda},\ \hat{\boldsymbol{\lambda}}) = \sum_{i=1}^{p} \frac{(\hat{\lambda_i} - \lambda_i)^2}{\lambda_i}. \tag{1.1}$$

Then Clevenson and Zidek (1975) were the first to propose a class of estimators of the form

$$\hat{\lambda}_i^{CZ}(\boldsymbol{X}) = X_i - \frac{\varphi(Z)}{Z + p - 1} X_i, \qquad i = 1, \ldots, p, \tag{1.2}$$

where $Z = \sum_{i=1}^{p} X_i$. Clevenson and Zidek (1975) have shown that when $p \geq 2$ and if $\varphi(\cdot)$ is a non-decreasing function satisfying $0 \leq \varphi(\cdot) \leq 2(p-1)$, then $\hat{\boldsymbol{\lambda}}^{CZ}(\boldsymbol{X}) = (\hat{\lambda}_1^{CZ}(\boldsymbol{X}), \ldots, \hat{\lambda}_p^{CZ}(\boldsymbol{X}))^t$ dominates $\boldsymbol{X}$. Since then broad classes of dominating estimators have been given by many authors, including Tsui and Press (1982), Hwang (1982), Ghosh et al. (1983), and Chou (1991).

1.1 Shrinking to a specified point other than the origin

We first consider a class of estimators which shrink X_i to a_i only when $x_i \geq a_i, i = 1, \ldots, p$, where a_i's are specified nonnegative values which are chosen according to prior information about λ_i's. Let $\mathcal{C} = \{(x_1, \ldots, x_p) \mid x_i \geq a_i, i = 1, \ldots, p\}$ and let $I_{\mathcal{C}}$ be its indicator function. We consider a class of estimators of the form

$$\hat{\lambda}_i^{a}(\boldsymbol{X}) = X_i - \varphi(Z) \frac{X_i - a_i}{Z + d} I_{\mathcal{C}}, \qquad i = 1, \ldots, p,$$

where $Z = \sum_{i=1}^{p} (X_i - a_i)$ and d is a positive constant.

Theorem 1.1. Let $\varphi(\cdot)$ be a non-decreasing function which satisfies $0 \leq \varphi(\cdot) \leq 2(p-1)$ and suppose that $d \geq \sup \varphi(\cdot)/2$. Then $\hat{\boldsymbol{\lambda}}^{a}(\boldsymbol{X})$, whose i-th component is given by (2.1), dominates $\boldsymbol{X}$ under the normalized squared error loss (1.1).

1.2 Shrinking to the order statistics

Here we consider a class of estimators which shrink to the order statistics. We first consider the following shrinkage estimator toward $X_{(1)} = min\{X_1, \ldots, X_p\}$:

$$\hat{\lambda}_i^{(1)}(\boldsymbol{X}) = X_i - \varphi(W) \frac{X_i - X_{(1)}}{W + d}, \qquad i = 1, 2, \ldots, p,$$

where $W=\sum_{i=1}^{p}(X_i-X_{(1)})$. Then we have the following.

Theorem 1.2. Suppose that $p\geq 3$. If $\varphi(\cdot)$ is a non-decreasing function which satisfies $0\leq\varphi(\cdot)\leq 2(p-2)$ and $d\geq \sup\varphi(\cdot)/2$, then $\hat{\boldsymbol{\lambda}}^{(1)}(\boldsymbol{X})$, whose i-th component is given by (2.4), dominates $\boldsymbol{X}$ under the normalized squared error loss (1.1).

2 Shrinkage estimators of multiplicative Poisson means in two-way contingency tables

We consider two-way multiplicative model where $x_{ij},\ i=1,\ldots,I,\ j=1,\ldots,J$, are independent random Poisson random variables with means

$$\lambda_{ij}=\lambda\alpha_i\beta_j,\quad i=1,\ldots,I,\ \ j=1,\ldots,J,$$

where $\alpha_i\geq 0$ and $\beta_j\geq 0$ satisfy $\sum_{i=1}^{I}\alpha_i=1$ and $\sum_{j=1}^{J}\beta_j=1$, respectively. We denote the one-dimensional frequencies and the total frequency by

$$x_{i+}=\sum_{j=1}^{J}x_{ij},\ i=1,\ldots,I,\ x_{+j}=\sum_{i=1}^{I}x_{ij},\ j=1,\ldots,J,\ x_{++}=\sum_{i=1}^{I}\sum_{j=1}^{J}x_{ij}.$$

As discussed in Hara and Takemura (2006) complete sufficient statistics are $\boldsymbol{x}_1=(x_{1+},\ldots,x_{I+})$ and $\boldsymbol{x}_2=(x_{+1},\ldots,x_{+J})$. The MLE of λ_{ij} is

$$\hat{\lambda}_{ij}^{ML}=\begin{cases}\dfrac{x_{i+}x_{+j}}{x_{++}} & \text{if}\quad x_{++}\neq 0\\ 0 & \text{if}\quad x_{++}=0.\end{cases}$$

They have given a class of improved estimators which shrink the MLE toward the origin under the normalized squared error loss. The simple one is

$$\delta_{ij}^{HT}=\frac{x_{i+}x_{+j}}{x_{++}}\left\{1-\frac{d}{x_{++}+d}\right\},\quad i=1,\ldots,I,\ j=1,\ldots,J,$$

2.1 One-dimensional shrinkage to an order statistic or a specified point

Let $x_{(\ell)+}$ be the ℓ-th smallest observation among $x_{1+},\ldots,x_{I+}$. We assume that $I\geq\ell+2$ and consider the following estimator which shrinks x_{i+} toward $x_{(\ell)+}$ when $x_{i+}\geq x_{(\ell)+}$:

$$\delta_{ij}^{(1)}=\frac{x_{+j}}{x_{++}}\left\{x_{i+}-\varphi(W)\frac{(x_{i+}-x_{(\ell)+})^{+}}{W+d}\right\},\quad i=1,\ldots,I,\ j=1,\ldots,J,$$

where $W=\sum_{i=1}^{I}(x_{i+}-x_{(\ell)+})^{+}$, $a^{+}=\max(0,a)$ and d is a positive constant. Then we have the following.

Theorem 2.1. Suppose that $\varphi(W)$ is a non-decreasing function satisfying $0\leq\varphi(W)\leq 2(I-\ell-1)$ and that $d\geq\sup\varphi(W)/2$. Then $\delta_{ij}^{(1)},i=1,\ldots,I$ improves upon the MLE $\lambda_{ij}^{ML},i=1,\ldots,I$ under the loss function $\sum_{i=1}^{I}(\hat{\lambda}_{ij}-\lambda_{ij})^2/\lambda_{ij}$ for any $j=1,\ldots,J$.

2.2 One-dimensional shrinkage to a specified point

Let $b_i\geq 0,i=1,\ldots,I$ be given numbers and we propose the following shrinkage estimator which shrinks x_{i+} to b_i when $x_{i+}\geq b_i$:

$$\delta_{ij}^{(2)}=\frac{x_{+j}}{x_{++}}\left\{x_{i+}-\varphi(N,W)\frac{(x_{i+}-b_i)^{+}}{W+d(N)}\right\},\quad i=1,\ldots,I,\ j=1,\ldots,J,$$

where $W=\sum_{i=1}^{I}(x_{i+}-b_i)^{+}$ and $N=\#\{i|x_{i+}\geq b_i\}$. Then we have the following.

Theorem 2.2, Suppose that $\varphi(N,W)$ is a non-decreasing function of W and satisfies $0\leq\varphi(N,W)\leq 2(N-1)^{+}$ for any $0\leq N\leq I$. Suppose that $d(N)\geq\sup_W\varphi(N,W)/2$. Then $\delta_{ij}^{(2)},i=1,\ldots,I$ improves upon the MLE $\hat{\lambda}_{ij}^{ML},i=1,\ldots,I$ under the loss function $\sum_{i=1}^{I}(\hat{\lambda}_{ij}-\lambda_{ij})^2/\lambda_{ij}$ for any $j=1,\ldots,J$.

Fredholm 行列式に関する Darling の公式とその拡張

田中 勝人（一橋大学名誉教授）

Brown 運動の 2 次汎関数で表現される統計量の特性関数を導出するにあたり，特定の核関数に対しては，従来の複雑な手続きを回避して簡単で効率的な方法が可能であることを証明し，具体例を示した．内容は以下の通りである．

ほぼ定符号の核関数 $K(s,t)$ と Brown 運動 $\{W(t)\}$ に対して，統計量

$$S = \int_0^1 \int_0^1 K(s,t)\, dW(s)\, dW(t)$$

を考える．このとき，次の定理が成立する．

定理 1：Anderson-Darling の定理． 統計量 S の特性関数は，

$$\begin{aligned} \mathrm{E}\left(e^{i\theta S}\right) &= \mathrm{E}\left[\exp\left\{i\theta \int_0^1 \int_0^1 K(s,t)\, dW(s)\, dW(t)\right\}\right] \\ &= \prod_{n=1}^{\infty} \left(1 - \frac{2i\theta}{\lambda_n}\right)^{-1/2} = \Big(D(2i\theta)\Big)^{-1/2} \end{aligned}$$

で与えられる．ここで，$\lambda_1, \lambda_2, , \ldots,$ は $K(s,t)$ の固有値で重複度の分だけ繰り返される．また，$D(\lambda)$ は $K(s,t)$ の Fredholm 行列式 (FD) である．

FD の定義から FD を導出することは一般に困難である．そのために積分方程式を同値な微分方程式（＋境界条件）に変換して，その解から FD の候補を見つけ，実際に FD であることを証明するのが通常の方法である．しかし，核関数が複雑な場合は，導出が困難となる．

ある条件をみたす核関数に限定した上で，微分方程式の議論などを回避することにより，簡単に効率的に FD を計算できる方法を提案した．そのための方法の一つとして Darling (1955) の公式を紹介した．

定理 2：Darling の公式． 核関数

$$K_1(s,t) = \min(s,t) - st - \psi(s)\psi(t), \quad \psi(0) = 0, \quad \psi(1) = 0$$

を定義する．ここで，$\psi(t)$ は連続微分可能である．このとき，$K_1(s,t)$ の FD は次式で与えられる．

$$D_1(\lambda) = \frac{\sin\sqrt{\lambda}}{\sqrt{\lambda}} - 2\int_0^1 \left(\int_0^t \psi'(s)\, \cos\sqrt{\lambda} s\, ds\right) \psi'(t)\, \cos\sqrt{\lambda}(1-t)\, dt. \tag{1}$$

Darling の公式の変形バージョンとして，次の公式を証明した．

定理 3. 核関数

$$K_2(s,t) = 1 - \max(s,t) - \psi(s)\psi(t), \quad \psi(1) = 0$$

を定義する．ここで，$\psi(t)$ は連続微分可能である．このとき，$K_2(s,t)$ の FD は次式で与えられる．

$$D_2(\lambda) = \cos\sqrt{\lambda} + 2\sqrt{\lambda}\int_0^1 \left(\int_0^t \psi'(s)\sin\sqrt{\lambda}s\,ds\right)\psi'(t)\cos\sqrt{\lambda}(1-t)\,dt. \tag{2}$$

さらに，核関数が次の場合の FD に関する定理を証明した．

定理 4. 核関数

$$K_3(s,t) = \min(s,t) - st - \sum_{j=1}^{m}\psi_j(s)\psi_j(t), \quad \psi_j(0) = \psi_j(1) = 0 \quad (j = 1, \ldots, m)$$

の FD は

$$D_3(\lambda) = \frac{\sin\sqrt{\lambda}}{\sqrt{\lambda}}\,|P(\lambda)| \tag{3}$$

で与えられる．ここで，$P(\lambda)$ は $m \times m$ の対称行列で，その (j,k) 要素は次のように定義される．

$$\begin{aligned}
P_{jj}(\lambda) &= 1 - \frac{2\sqrt{\lambda}}{\sin\sqrt{\lambda}}\int_0^1 \left(\int_0^t \psi_j'(s)\cos\sqrt{\lambda}s\,ds\right)\psi_j'(t)\cos\sqrt{\lambda}(1-t)\,dt,\\
P_{jk}(\lambda) &= -\frac{\sqrt{\lambda}}{\sin\sqrt{\lambda}}\int_0^1\int_0^1 \psi_j'(s)\psi_k'(t)L_1(s,t)\,ds\,dt \qquad (j \neq k),\\
L_1(s,t) &= \begin{cases} \cos\sqrt{\lambda}s\,\cos\sqrt{\lambda}(1-t) & (s \le t)\\ \cos\sqrt{\lambda}t\,\cos\sqrt{\lambda}(1-s) & (s \ge t). \end{cases}
\end{aligned}$$

核関数が複雑となったもう一つの場合についても証明した．

定理 5. 核関数

$$K_4(s,t) = 1 - \max(s,t) - \sum_{j=1}^{m}\psi_j(s)\psi_j(t), \quad \psi_j(1) = 0$$

の FD は

$$D_4(\lambda) = \cos\sqrt{\lambda}\,|P(\lambda)| \tag{4}$$

で与えられる．ここで，

$$\begin{aligned}
P_{jj}(\lambda) &= 1 + \frac{2\sqrt{\lambda}}{\cos\sqrt{\lambda}}\int_0^1 \left(\int_0^t \psi_j'(s)\sin\sqrt{\lambda}s\,ds\right)\psi_j'(t)\cos\sqrt{\lambda}(1-t)\,dt,\\
P_{jk}(\lambda) &= \frac{\sqrt{\lambda}}{\cos\sqrt{\lambda}}\int_0^1\int_0^1 \psi_j'(s)\psi_k'(t)L_2(s,t)\,ds\,dt \qquad (j \neq k),\\
L_2(s,t) &= \begin{cases} \sin\sqrt{\lambda}s\,\cos\sqrt{\lambda}(1-t) & (s \le t)\\ \sin\sqrt{\lambda}t\,\cos\sqrt{\lambda}(1-s) & (s \ge t). \end{cases}
\end{aligned}$$

Comprehensive Interval-Valued Time Series Model with Application to the S&P 500 Index and PM2.5 Level Data Analysis

Liang-Ching Lin[a*], Hao Sung[a] and Sangyeol Lee[b]

[a] *Department of Statistics, Institute of Data Science, National Cheng Kung University, Tainan, Taiwan.*

[b] *Department of Statistics, Seoul National University, Seoul, 08826, South Korea.*

1 Introduction

In recent years, modern data architecture has been recorded in short time intervals, but still contains rich information resources. Such data may contain several problems such as missing data, repeated observations, and recording non-equidistant time points. To resolve such issues, researchers have considered reorganizing the data into interval-valued data consisting of daily or weekly maximum and minimum values as it can preserve more information than merely aggregated daily or weekly mean data. For instance, the PM2.5 levels are recorded hourly but there are 5588 missing observations. To handle missing data, Gao and Tsay (2019) aggregated the data into weekly mean observations. However, we arranged the data into weekly maximum and minimum observations. Then, we consider to fit the auto-interval-regressive (AIR) model proposed by Lin et al. (2021) and the auto-interval-regressive moving average (AIRMA) model by Lin et al. (2023). Further, we fit the GHVAIRMA (generalized heteroscedastic volatility AIRMA) model to improve the accuracy of the one-step-ahead prediction of the high and low prices of the S&P 500 index. Overall, our findings strongly confirm the adequacy of the proposed model.

2 Interval-Valued Time Series Model

Let $\mathbf{Y}_t = (Y_{u,t}, Y_{l,t})^T$. Lin et al. (2021) proposed the AIR(p) model as follows:

$$\mathbf{Y}_t = \phi_1 \mathbf{Y}_{t-1} + \cdots + \phi_p \mathbf{Y}_{t-p} + \mathbf{A}_t,$$

where ϕ_i, $i = 1, \ldots, p$, are model parameters, $\mathbf{A}_t = (A_{u,t}, A_{l,t})^T$ are i.i.d. errors with $A_{u,t} = \max\{A_{1,t}, \ldots, A_{n,t}\}$ and $A_{l,t} = \min\{A_{1,t}, \ldots, A_{n,t}\}$; and $A_{i,t}$, $i = 1, \ldots, n$, are i.i.d. normal random variables with a mean of zero and variance of σ^2 for some $n > 1$. Lin et al. (2023) proposed the IVMA(q) if the following equation is satisfied:

$$\mathbf{Y}_t = \mathbf{A}_t - \theta_1 \mathbf{A}_{t-1} - \cdots - \theta_q \mathbf{A}_{t-q}. \tag{2.1}$$

$\{\mathbf{Y}_t\}$ is said to follow an AIRMA($p,\ q$) if the following equation holds:

$$\mathbf{Y}_t = \phi_1 \mathbf{Y}_{t-1} + \cdots + \phi_p \mathbf{Y}_{t-p} + \mathbf{A}_t - \theta_1 \mathbf{A}_{t-1} - \cdots - \theta_q \mathbf{A}_{t-q}. \tag{2.2}$$

If the error terms $A_{i,t}$, $i = 1, \ldots, n$ have a mean of zero and the conditional heteroscedastic variance of σ_t^2. Then, $\mathbf{A}_t$ follows a general heteroscedastic volatility model with lags r and s,

*E-mail address: lclin@ncku.edu.tw

abbreviated as GHV(r, s), if σ_t^2 satisfies equation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \gamma_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \text{ where } \gamma_t^2 = \left(\frac{A_{u,t} - A_{l,t}}{\Phi^{-1}\left(\frac{n-\alpha}{n-2\alpha+1}\right) - \Phi^{-1}\left(\frac{1-\alpha}{n-2\alpha+1}\right)} \right)^2. \quad (2.3)$$

Combined with the AIRMA model from (2.2), we have the GHV(r, s)-AIRMA(p, q), where the interval-valued time series $\{\mathbf{Y}_t\}$ satisfies

$$\mathbf{Y}_t = \phi_1 \mathbf{Y}_{t-1} + \cdots + \phi_p \mathbf{Y}_{t-p} + \mathbf{A}_t - \theta_1 \mathbf{A}_{t-1} - \cdots - \theta_q \mathbf{A}_{t-q}. \quad (2.4)$$

3 Real Data Analysis

In this application, we consider the prediction of one-step high and low prices for the S&P 500 index. The MDEs of the proposed GHV(1,1)-IVMA(3) model outperform the VAR-DCC and HVAIR(1,1) models.

For the PM2.5 level data, we fit the AIR(p_1), IVMA(q_1), and AIRMA(p_2, q_2) models with $p_1 \leq 10$, $q_1 \leq 12$, $p_2 \leq 2$, $q_2 \leq 4$, and the exogenous parameter $n = 24 \times 7 = 168$ to the dataset, and choose the best models as those with the lowest AIC and BIC values. We investigate the relationships among the stations by using the dandelion plot proposed by Zhang and Lin (2022), which establishes the pairwise correlations of the center-to-center and range-to-range of the residuals. The residuals deduced from the proposed models provide more intuitively appealing results. For comparison, we also fit a VAR(1) model to the weekly mean data. All other stations exhibit higher correlations with each other. The results indicate that too much information has been aggregated into a single point. These comparative results indicate that our method preserves more information and provides a more intuitively appealing result.

4 Concluding Remarks

Our models can be relaxed by replacing it with other distributions, and the coefficients of the maximum and minimum values can be allowed to differ from each other as follows:

$$\begin{pmatrix} Y_{u,t} \\ Y_{l,t} \end{pmatrix} = \begin{pmatrix} \phi_1 Y_{u,t-1} \\ \phi_2 Y_{l,t-1} \end{pmatrix} + \begin{pmatrix} A_{u,t} \\ A_{l,t} \end{pmatrix} - \begin{pmatrix} \theta_1 A_{u,t} \\ \theta_2 A_{l,t-1} \end{pmatrix}.$$

All these issues are worth further investigation and should be examined in future studies.

Reference

1. Gao, Z., and Tsay, R. S. (2019). A Structural-Factor Approach to Modeling High-Dimensional Time Series and Space-Time Data. *Journal of Time Series Analysis*, **40(3)**, 343-362.

2. Lin, L.-C., Chien, H.-L., and Lee, S. (2021). Symbolic interval-valued data analysis for time series based on auto-interval-regressive models. *Statistical Methods & Applications*, **30**, 295-315.

3. Lin, L.-C., Hao, S., and Lee, S. (2023). Comprehensive interval-valued time series model with application to the S&P 500 index and PM2.5 level data analysis. *Applied Stochastic Models in Business and Industry*, **39(2)**, 198-218.

4. Zhang, M., and Lin, D. K. (2022). Visualization for Interval Data. *Journal of Computational and Graphical Statistics*, DOI: 10.1080/10618600.2022.2066678.

Rank-R Matrix Autoregressive Models for Modeling Spatio-Temporal Data

Nan-Jung Hsu, Institute of Statistics, National Tsing-Hua University
Hsin-Cheng Huang, Institute of Statistical Science, Academia Sinica
Ruey S. Tsay, Booth School of Business, University of Chicago
Tzu-Chieh Kao, Institute of Statistics, National Tsing-Hua University

September 30, 2023

Abstract

We develop a matrix-variate autoregressive (MAR) model to analyze spatio-temporal data organized on a regular grid in space. The model is an extension of the bilinear MAR spatial model of Hsu, Huang and Tsay (2021) by increasing its flexibility and applicability in empirical applications. Specifically, we propose to model each autoregressive (AR) coefficient matrix of the MAR model by R bilinear terms, thereby establishing a rank-R model. The extension can be interpreted as decomposing the AR dynamics of the data into R bilinear MAR components. We further incorporate a banded neighborhood structure for AR coefficient matrices and utilize a flexible nonstationary low-rank covariance model for the spatial innovation process, leading to a parsimonious model without sacrificing its flexibility. We estimate all parameters of the model by the maximum likelihood method and develop a computationally efficient alternating direction method of multipliers algorithm, involving only closed-form expressions in all steps. Applications to a wind-speed dataset and an employment dataset, as well as two simulation experiments, demonstrate the effectiveness of the proposed method in estimation, model selection, and prediction.

Keywords: Alternating direction method of multipliers, Bayesian information criterion, Kronecker product, Low-rank approximation, Matrix-variate time series, Nonstationary spatial model, Singular value decomposition.

1 Matrix Autoregressive Spatio-Temporal Models

Consider a zero-mean $n \times m$ matrix-variate time series $\boldsymbol{Y}_t$ with T observations, $\{\boldsymbol{Y}_t = (y_{t,i,j})_{n\times m} : t = 1, \ldots, T\}$.

We proposed an MAR(p) model with rank-R bilinear forms focusing on the dynamic information in neighboring columns and rows of $\boldsymbol{Y}_t$'s to achieve dimension reduction (Hsu, Huang and Tsay, 2021; Hsu, Huang, Tsay and Kao, 2023):

$$\boldsymbol{Y}_t = \sum_{j=1}^{p} \sum_{r=1}^{R} \sigma_{j,r} \boldsymbol{A}_{j,r} \boldsymbol{Y}_{t-j} \boldsymbol{B}_{j,r}' + \boldsymbol{\eta}_t, \tag{1}$$

which forms a specific case of an vector AR model:

$$\boldsymbol{y}_t \equiv \operatorname{vec}(\boldsymbol{Y}_t) = \sum_{j=1}^{p} \left(\sum_{r=1}^{R} \sigma_{j,r} \boldsymbol{B}_{j,r} \otimes \boldsymbol{A}_{j,r} \right) \boldsymbol{y}_{t-j} + \operatorname{vec}(\boldsymbol{\eta}_t), \tag{2}$$

with structured AR coefficient matrices satisfying

$$\boldsymbol{\Phi}_j = \sum_{r=1}^{R} \sigma_{j,r} \left(\boldsymbol{B}_{j,r} \otimes \boldsymbol{A}_{j,r} \right), \quad j = 1, \ldots, p. \tag{3}$$

For model identifiability, we set

$$\operatorname{tr}(\boldsymbol{A}_{j,r}' \boldsymbol{A}_{j,s}) = \operatorname{tr}(\boldsymbol{B}_{j,r}' \boldsymbol{B}_{j,s}) = \delta_{rs}, \quad \sigma_{j,1} \geq \cdots \geq \sigma_{j,R} > 0; \quad 1 \leq r, s \leq R,\, j = 1, \ldots, p, \tag{4}$$

where $\delta_{rs} = 1$ if $r = s$; and 0 otherwise. The AR coefficient matrix $\boldsymbol{\Phi}_j$ in (3) has a close connection to the singular value decomposition (SVD) of a re-shaped matrix with rank R.

Regarding the specification of the covariance $\boldsymbol{\Psi}$ for the innovation $\boldsymbol{\eta}_t$, we adopt the fixed-rank covariance model (Cressie and Johannesson, 2008):

$$\boldsymbol{\Psi} \equiv \operatorname{var}(\operatorname{vec}(\boldsymbol{\eta}_t)) = \boldsymbol{F} \boldsymbol{M} \boldsymbol{F}' + \sigma_\eta^2 \boldsymbol{I}, \tag{5}$$

where $\boldsymbol{F}$ is a pre-specified $nm \times K$ matrix of basis functions, $\boldsymbol{M}$ and $\sigma_\eta^2 \geq 0$ are unknown parameters. We recommend using the multi-resolution spline basis functions of Tzeng and Huang (2018) for $\boldsymbol{F}$ which are flexible to characterize features up to a chosen scale.

2 Maximum Likelihood Estimation

Denote $\boldsymbol{\Phi} \equiv [\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p]$, $\boldsymbol{Y}_{t_1:t_2} \equiv [\boldsymbol{y}_{t_1}, \cdots \boldsymbol{y}_{t_2}]$, and let

$$\boldsymbol{Y}^* \equiv \begin{bmatrix} \boldsymbol{y}_p & \boldsymbol{y}_{p+1} & \cdots & \boldsymbol{y}_{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{y}_2 & \boldsymbol{y}_3 & \cdots & \boldsymbol{y}_{T-p+1} \\ \boldsymbol{y}_1 & \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_{T-p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_{p:(T-1)} \\ \boldsymbol{Y}_{(p-1):(T-2)} \\ \vdots \\ \boldsymbol{Y}_{1:(T-p)} \end{bmatrix}.$$

The Gaussian log-likelihood of $(\boldsymbol{\Phi}, \boldsymbol{\Psi})$, conditional on $\boldsymbol{Y}_{1:p}$, can be expressed as

$$\ell(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = -\frac{T-p}{2} \log(\det(\boldsymbol{\Psi})) - \frac{1}{2} \operatorname{tr}\left[\left(\boldsymbol{Y}_{(p+1):T} - \boldsymbol{\Phi} \boldsymbol{Y}^* \right)' \boldsymbol{\Psi}^{-1} \left(\boldsymbol{Y}_{(p+1):T} - \boldsymbol{\Phi} \boldsymbol{Y}^* \right) \right]. \tag{6}$$

Thus, the constrained ML estimators (MLEs) of $(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ (conditional on $\boldsymbol{Y}_{1:p}$) are given by

$$(\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}) \equiv \underset{\boldsymbol{\Phi},\boldsymbol{\Psi}}{\arg\min} \left\{ \log(\det(\boldsymbol{\Psi})) + \frac{1}{T-p} \operatorname{tr}\left[\left(\boldsymbol{Y}_{(p+1):T} - \boldsymbol{\Phi}\boldsymbol{Y}^*\right)' \boldsymbol{\Psi}^{-1} \left(\boldsymbol{Y}_{(p+1):T} - \boldsymbol{\Phi}\boldsymbol{Y}^*\right) \right] \right\}, \tag{7}$$

subject to the constraints (3), (4), (5).

2.1 ADMM Algorithm for Obtaining MLEs

The constrained parameter estimates $\hat{\boldsymbol{\Phi}}$ and $\hat{\boldsymbol{\Psi}}$ do not have closed-form expressions, We develop an alternating direction method of multipliers (ADMM) algorithm (Boyd, *et al.*, 2011) to solve the resulting constrained optimization problem. First, we augment an auxiliary $(nm) \times (pnm)$ parameter matrix, $\boldsymbol{C} \equiv [\boldsymbol{C}_1, \ldots, \boldsymbol{C}_p]$ with

$$\boldsymbol{C}_j = \sum_{r=1}^{R} \sigma_{j,r} \left(\boldsymbol{B}_{j,r} \otimes \boldsymbol{A}_{j,r}\right); \quad j = 1, \ldots, p, \tag{8}$$

to alleviate the direct handling of the rank-R structure in $\boldsymbol{\Phi}$. This transfer of the rank-R constraint from $\boldsymbol{\Phi}$ to $\boldsymbol{C}$ simplifies the overall optimization process. By introducing $\boldsymbol{C}$, the augmented Lagrangian of (7) with the constraint $\boldsymbol{C} = \boldsymbol{\Phi}$ follows

$$\begin{aligned} L_\rho(\boldsymbol{\Phi}, \boldsymbol{C}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}) = \; & \log(\det(\boldsymbol{\Psi})) + \frac{1}{T-p} \operatorname{tr}\left\{ \left(\boldsymbol{Y}_{(p+1):T} - \boldsymbol{\Phi}\boldsymbol{Y}^*\right)' \boldsymbol{\Psi}^{-1} \left(\boldsymbol{Y}_{(p+1):T} - \boldsymbol{\Phi}\boldsymbol{Y}^*\right) \right\} \\ & + \operatorname{tr}\left[\boldsymbol{\Lambda}'(\boldsymbol{\Phi} - \boldsymbol{C})\right] + \frac{\rho}{2} \|\boldsymbol{\Phi} - \boldsymbol{C}\|_F^2, \end{aligned} \tag{9}$$

where $\boldsymbol{\Lambda} \equiv [\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_p]$ is the matrix of Lagrangian multipliers, $\rho > 0$ is the augmented Lagrangian parameter. Starting with some initial estimates $\boldsymbol{C}^{(0)}$, $\boldsymbol{\Psi}^{(0)}$, $\boldsymbol{\Lambda}^{(0)}$ of $\boldsymbol{C}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Lambda}$, the proposed ADMM algorithm iteratively cycles through four steps: $\boldsymbol{\Phi}$-step, $\boldsymbol{C}$-step, $\boldsymbol{\Psi}$-step, and $\boldsymbol{\Lambda}$-step, until convergence. The algorithm is computationally efficient due to the presence of simple closed-form expressions in all four steps, as detailed in Hsu, Huang, Tsay and Kao (2023). The algorithm's pseudo-code is outlined in Algorithm 1.

2.2 Model Selection

The proposed MAR model consists of four structure parameters to select, including the autoregressive order p, the reshaped matrix rank R for $\{\boldsymbol{A}_{j,r}\}$ and $\{\boldsymbol{B}_{j,r}\}$, the dimension K for the spatial covariance matrix $\boldsymbol{M}$, and the bandwidth L of $\{\boldsymbol{A}_{j,r}\}$ and $\{\boldsymbol{B}_{j,r}\}$. Note that, the rank R of the reshaped matrix is upper bounded by $R_{\max} \equiv \min\{n^2, m^2\}$ corresponding to the maximum rank of SVD representation of $\mathcal{R}(\boldsymbol{\Phi}_j)$. We consider Akaike's information criterion (AIC, Akaike, 1973) and Bayesian information criterion (BIC, Schwarz, 1978) to select these structure parameters (p, R, K, L):

$$\begin{aligned} \text{AIC} &= -\frac{2}{T-p} \ell(\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}) + \frac{2}{T-p} \left\{ \text{df}(\boldsymbol{\Phi}) + \text{df}(\boldsymbol{\Psi}) \right\}, \\ \text{BIC} &= -\frac{2}{T-p} \ell(\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}) + \frac{\log(T-p)}{T-p} \left\{ \text{df}(\boldsymbol{\Phi}) + \text{df}(\boldsymbol{\Psi}) \right\}, \end{aligned}$$

Algorithm 1 An ADMM Algorithm for the MLEs

Input:
initial estimates: $\boldsymbol{C}^{(0)}, \boldsymbol{\Psi}^{(0)}, \boldsymbol{\Lambda}^{(0)}$;
learning rate ρ;
Output:
set $k \leftarrow 0$;
repeat
$\boldsymbol{\Phi}$-step: $\boldsymbol{\Phi}^{(k+1)} = \arg\min_{\boldsymbol{\Phi}} L_\rho\left(\boldsymbol{\Phi}, \boldsymbol{C}^{(k)}, \boldsymbol{\Psi}^{(k)}, \boldsymbol{\Lambda}^{(k)}\right)$;
$\boldsymbol{C}$-step: $\boldsymbol{C}^{(k+1)} = \arg\min_{\boldsymbol{C}} L_\rho\left(\boldsymbol{\Phi}^{(k+1)}, \boldsymbol{C}, \boldsymbol{\Psi}^{(k)}, \boldsymbol{\Lambda}^{(k)}\right)$, subject to (4) and (8);
$\boldsymbol{\Psi}$-step: $\boldsymbol{\Psi}^{(k+1)} = \arg\min_{\boldsymbol{\Psi}} L_\rho\left(\boldsymbol{\Phi}^{(k+1)}, \boldsymbol{C}^{(k+1)}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}^{(k)}\right)$, subject to (5);
$\boldsymbol{\Lambda}$-step: $\boldsymbol{\Lambda}^{(k+1)} = \boldsymbol{\Lambda}^{(k)} + \rho\left(\boldsymbol{\Phi}^{(k+1)} - \boldsymbol{C}^{(k+1)}\right)$;
$k \leftarrow k + 1$;
until a stopping criterion is satisfied.

Table 1: Model selection results for the wind-speed data.

p	R	K	L	BIC	PMSE
1	3	150	3	-447.13	1.556
2	3	140	2	-459.33	1.583
3	2	140	2	-462.24	1.560
4	3	140	1	-466.02	1.546
5	2	140	1	-466.54	1.549
6	2	140	1	**-469.08**	1.558
7	2	140	2	-466.86	1.534

where $\mathrm{df}(\boldsymbol{\Phi})$ and $\mathrm{df}(\boldsymbol{\Psi})$ represent the numbers of free parameters in $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. Under the banded constraints on $(\boldsymbol{A}, \boldsymbol{B})$ with a bandwidth L,

$$\mathrm{df}(\boldsymbol{\Phi}) = pR[(2n-1-L)L + (2m-1-L)L + n + m - R],\ 0 \le L \le L_{\max},\ 1 \le R \le R_{\max}.$$

Under the Toeplitz constraints on $(\boldsymbol{A}, \boldsymbol{B})$ with a bandwidth L,

$$\mathrm{df}(\boldsymbol{\Phi}) = pR[4L + 1 - (R-1)], \quad \text{provided} \quad R < 4L + 2.$$

The number of free parameters in $\boldsymbol{\Psi}$ has been shown by Tzeng and Huang (2018) to be:

$$\mathrm{df}(\boldsymbol{\Psi}) = \begin{cases} K(K+1)/2 + 1, & \text{if } K < T, \\ KT + 1 - T(T-1)/2, & \text{if } K \ge T, \end{cases}$$

where $K \le nm$.

3 Applications

We apply the extended MAR models to the wind-speed data, which consist of 6-hour averaged U-wind (i.e., eastward wind component) values at 289 locations (17x17 regular grid), collected from November 1992 to February 1993 over the region within latitudes 14°S-16°N and longitudes 145°E-175°E.

Similarly to previous studies (Hsu, Chang and Huang, 2012; Hsu, Huang and Tsay, 2021), we first removed the temporal average for each of the 289 locations and applied our method to the mean-adjusted data $\{\boldsymbol{Y}_t : t = 1, \ldots, 480\}$ with $\boldsymbol{Y}_t$ being a 17×17 matrix.

For fitting and performance evaluation, we split the data into a training sample $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{400}\}$ and a validation sample $\{\boldsymbol{Y}_{401}, \ldots, \boldsymbol{Y}_{480}\}$. We select the best MAR model via BIC among the candidate set:

$$(p, R, K, L) \in \mathcal{M} \equiv \{1, \ldots, 7\} \times \{1, \ldots, 4\} \times \{10, 20, \ldots, 280\} \times \{1, \ldots, 16\}.$$

Table 1 reports the performance of our best model compared with the popular models in high-dimensional time series. Our MAR model has an overall satisfactory performance among others.

References

Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1) 1–122.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets, *Journal of the Royal Statistical Society, Series B*, **70**, 209–226.

Hsu, N.-J., Huang, H.-C., and Tsay, R. S. (2021). Matrix autoregressive spatio-temporal models, *Journal of Computational and Graphical Statistics*, **30**, 1143–1155.

Hsu, N-J, Huang, H-C, and Tsay, R. S., Kao, T-C (2023+) Rank-R matrix autoregressive models for modeling spatio-temporal data, forthcoming in *Statistics and Its Interface*.

Tzeng, S. and Huang, H.-C. (2018). Resolution adaptive fixed rank kriging, *Technometrics*, **60**, 198–208.

Van Loan, C. and Pitsianis, N. (1993). Approximation with Kronecker products. In *Linear Algebra for Large Scale and Real Time Applications* (M. S. Moonen and G. H. Golub eds.), Kluwer Academic Publishers, Dordrecht, 293–314.

Stable P-value Assignment in High-Dimensional Regression via Data Splitting

Chien-Chung Wang[a], Chih-Hao Chang[b], Hsin-Cheng Huang[c]

[a]*Department of Statistics, Colorado State University, USA*
[b]*Department of Statistics, National Chengchi University, Taiwan*
[c]*Institute of Statistical Science, Academia Sinica, Taiwan*

Abstract

In the rapidly evolving landscape of statistical research, high-dimensional regression models have emerged as a critical tool for tackling complex data structures. These models, which feature a large number of predictors relative to the sample size, present unique challenges that are not adequately addressed by traditional statistical methods. Conventional approaches, such as the ordinary least squares (OLS) estimator, often fall short of providing reliable estimates and inferences in high-dimensional settings. This limitation is particularly evident in the realms of variable selection, confidence interval construction, and hypothesis testing targeting individual regression coefficients.

A plethora of specialized model selection algorithms have been developed to navigate the complexities of high-dimensional data. These include Lasso, SCAD, Adaptive Lasso, SIS, MCP, and OGA, among others. Each of these algorithms brings its own set of advantages and limitations to the table, and they have been widely adopted in various applications of high-dimensional regression analysis. However, a common challenge that persists across these methods is the difficulty in conducting rigorous statistical inference on the selected models. Traditional inferential procedures often fail to account for the intricacies involved in high-dimensional model selection, leading to biased or inconsistent results.

In this work, we introduce a series of pivotal modifications to the existing multiple-splitting paradigm, aiming to improve the robustness and reliability of variable selection in high-dimensional regression models. Our modifications are designed to mitigate the conser-

vatism that is inherent in the original multiple-splitting approach. We redefine the p-value assignment mechanism for individual data splits, incorporating a more nuanced statistical framework that better captures the underlying data distribution. Furthermore, we integrate the stability-selection approach into our methodology, which allows for a more rigorous control of the false discovery rate.

To further enhance the overall performance of the variable selection process, we introduce the concept of multiple dependent p-value aggregation. This innovative framework capitalizes on the underlying correlation structure among the predictors, thereby offering a more comprehensive and accurate variable selection strategy. Unlike existing methods, our approach does not focus on computational efficiency but prioritizes the robustness and reliability of the variable selection process.

To validate our methodology, we conduct extensive simulation studies encompassing various scenarios, including varying sample sizes, predictor correlations, and signal-to-noise ratios. These simulations confirm the robustness of our approach in different high-dimensional settings. While we have not yet applied our methodology to real-world data sets in fields such as finance and social sciences, the empirical results suggest that it holds promise for such applications.

In summary, our work offers an advancement in p-value calculations for high-dimensional regression models. By redefining the p-value assignment mechanism and introducing multiple dependent p-value aggregation, we provide a more rigorous and reliable framework for statistical inference. This focus on p-values ensures that our methodology not only excels in variable selection but also provides more accurate and robust hypothesis tests. We believe these advancements will be a cornerstone for future research in high-dimensional statistical inference.

Keywords: Exchangeable p-values; Family-wise error rate; False discovery rate; Multiple testing; Post-selection inference; Selection consistency; Variable selection

Modeling Joint Cylindrical Distributions and Related Markov Processes

Toshihiro Abe[1], Tomoaki Imoto[2],
Takayuki Shiohama[3], Yoichi Miyata[4]

[1]Hosei University, [2]University of Shizuoka,
[3]Nanzan University [4]Takasaki City University of Economics

Abstract

Joint cylindrical distributions are key probability distributions that make possible a multivariate regression analysis as well as Markov models on a cylinder. In this study, some joint cylindrical distributions are proposed, and their statistical properties together with algorithms for random number generation are investigated. For proposed joint cylindrical distributions, the marginal distributions for various combinations of linear and circular variables are obtained. These marginal distributions are applied to the Markov process on a cylinder. The maximum likelihood estimation for unknown model parameters is investigated. To illustrate the applicability of the proposed models, time series analysis using wind speed and direction data is investigated.
keywords : circular statistics; cylindrical data; Markov process; maximum likelihood estimation.

1 Introduction

Bivariate data which has a form of the combination of angular and linear variables are called cylindrical data. Examples of cylindrical data can be found in various fields in the environment and natural science, engineering, information science, and so forth. Because a cylindrical data analysis may provide clear insights into the probabilistic structures and relationships between the angular and linear variables, many common and conventional statistical tools are applied to the data analysis.

Fitting cylindrical data using cylindrical distribution has a long history. Example of early works includes the distribution proposed by [5], where they considered the exponential distribution on a linear variable and von Mises distribution on a circular variable. The normal distribution on a linear part and von Mises distribution on a circular part are proposed by [7], where the mean and variance of the normal distributions are the functions of circular variables. The drawbacks of this model were found to be difficult to express the marginal distribution of a linear variable. In [2], a Weibull von Mises distribution on a cylinder was considered and illustrated the model's tractability for real data analysis. This distribution was extended to describe heavy tails on linear variables by [4].

Time series analysis for cylindrical data is of special importance in the prediction of wind speed and directions in wind firm management with accurately predicting wind energy. Hidden Markov Models based on the Weibull von Mises distribution proposed by [2] are considered by [6], and they found that their models can explain parsimonious accommodations of circular-linear correlation, multimodality, skewness, and temporal autocorrelation of the cylindrical data.

Joint cylindrical distributions are beneficial as they provide a regression analysis for both response and explanatory variables that take values on a cylinder. For time series modeling, the joint cylindrical distribution is needed to formulate a Markov model on a cylinder. Despite the necessity for such needs in statistical modeling, there is little known about the joint distribution of a cylinder. Recently, joint cylindrical distribution is proposed by [1]. However, the proposed distribution has several shortcomings to fit the models in real data applications. The idea behind this study is a natural extension of the modeling that is used in the study [3] and [1]. The possible way of constructing a joint cylindrical distribution is to apply an approach adopted by [8] and [5]. However, this modeling needs to handle distribution functions for each marginal density which makes it difficult to formulate various combinations of joint marginal distributions. Similar difficulties arise when we apply copula-based multivariate modeling.

The joint cylindrical distributions proposed in this study and their applications to regression analysis and time series analysis are classified into a distributional regression approach. These statistical methods may provide more flexible modeling than nonparametric regression models as they provide whole probabilistic information on the conditional distribution of the response variable given explanatory variable, including quantiles and median of the predicted response variable.

Because of its mathematical tractability and a good fit to real-world data, the Markov property of a conditional cylindrical distribution is a basic concept in state-space modeling and is widely applied in various fields. In this study, we elucidate how to construct joint cylindrical distributions that are tractable for data analysis and deduce conditional distribution which is applicable for Markov transition density for time series analysis. We show that the resulting Markov processes are found to have a non-linear first-order autoregressive structure. In addition, conditional circular autocorrelation functions are obtained for the proposed processes.

References

[1] Abe, T., Imoto, T., Shiohama, T. & Miyata, Y. : On some flexible models for circular, toroidal, and cylindrical data. *Directional Statistics for Innovative Applications*, Springer, pp. 229–243. (2022)

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

[3] Imoto, T. & Abe, T. : The simple construction of a toroidal distribution from independent circular distributions. *Journal of Multivariate Analysis*, **186**, 104799 (2021)

[4] Imoto, T., Shimizu, K., & Abe, T. : A cylindrical distribution with heavy-tailed linear part. *Japanese Journal of Statistics and Data Science*, **2** (1), pp. 129–154. (2019)

[5] Johnson, R. A. & Wehrly, T. E. : Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, **73** (363), pp. 602–606. (1978)

[6] Lagona, F., Picone, M. & Maruotti, A. : A hidden Markov model for the analysis of cylindrical time series. *Environmetrics*, **26** (8), pp. 534–544. (2015)

[7] Mardia, K. V. and Sutton, T. W. : A model for cylindrical variables with applications. *Journal of the Royal Statistical Society: Series B*, **40** (2), pp. 229–233. (1978)

[8] Wehrly, T. E. & Johnson, R. A. : Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, **67** (1), pp. 255–256. (1980)

Test for the existence of the residual spectrum with application to brain functional connectivity detection

Yuichi Goto (Kyushu University)[*1]
Xuze Zhang (University of Maryland)
Benjamin Kedem (University of Maryland)
Shuo Chen (University of Maryland)

本講演では, 非線形構造を考慮した残差スペクトルという時系列解析の周波数領域における指標を導入し, 残差スペクトルが存在するかどうかの検定問題に対する検定手法を提案した. コヒーレンスは, 2つの時系列間の類似度の指標であり, ピアソンの相関係数の時系列への拡張となっている. しかし, コヒーレンスでは時系列間の線形関係しか捉えることができないため, 非線形関係を捉えられる指標が必要とされている. 応用例として, 脳障害と深く関連している脳の機能的結合の検出が挙げられ, コヒーレンスでは検出できなかった機能的結合が残差スペクトルでは検出が可能であることを示した. この講演は, Goto et al. (2023) に基づいたものである.

1. モデルと主定理

本講演では, 次のモデルを考えた;

$$X_0(t) := \zeta + \sum_{i=1}^{K} \sum_{k_i=-\infty}^{\infty} b_i(k_i) X_i(t-k_i) + \epsilon(t). \tag{1}$$

ここで, $X_0(t)$ は応答変数, ζ は切片, $\boldsymbol{X}(t) := (X_1(t), \ldots, X_K(t))^\top$ は共変量であり, 自己共分散行列 $\Gamma_{\boldsymbol{X}}(u) := \mathrm{E}\boldsymbol{X}(t)\boldsymbol{X}(t-u)$ を持つ. ただし, $X_i(t) := X_i' - \mathrm{E}X_i'$ であり, $(X_1'(t), \ldots, X_K'(t))^\top$ は, 任意の $s \in \mathbb{N}$ に対して, s 次定常過程. また, $\epsilon(t)$ は任意のモーメントを持つ平均ゼロである i.i.d. 誤差過程で $\{\boldsymbol{X}; t \in \mathbb{Z}\}$ と独立, 係数 $b_i(k)$ は $\sum_{k=-\infty}^{\infty} k^2 |b_i(k)| < \infty$ を満たす.

次の条件を仮定する:

仮定 1. 任意の2以上の整数 ℓ と $(i_1, \ldots, i_\ell) \in \{1, \ldots, K\}^\ell$ に対して,

$$\sum_{s_2, \ldots, s_\ell = -\infty}^{\infty} \left(1 + \sum_{j=2}^{\ell} |s_j|^2\right) |\mathrm{Cum}\{X_{i_1}(0), X_{i_2}(s_2), \ldots, X_{i_\ell}(s_\ell)\}| < \infty$$

を満たす.

講演では, 適当なモーメント条件を満たす geometrically α-mixing s次定常過程は, 仮定1を満たすということを示した.

仮定1の下で, $(X_0(t), X_1(t), \ldots, X_K(t))$ に対するスペクトル行列 $\boldsymbol{f}(\lambda) := (f_{ij}(\lambda))_{i,j=0,\ldots,K}$ が存在し, 二回連続微分可能である. モデル (1) の直交分解の結果, 次で定義される残差

This research was supported by JSPS Grant-in-Aid for Early-Career Scientists JP23K16851 (Y.G.) and NIH grant DP1 DA048968 (S.C.). We thank Brin Mathematics Research Center for supporting the first author's (Y.G.) research stay at the University of Maryland. Part of this research was conducted while the first author (Y.G.) was visiting the Brin Mathematics Research Center at the University of Maryland.

[*1] e-mail: yuichi.goto@math.kyushu-u.ac.jp

スペクトルが自然に導出されることを明らかにした:

$$f_{G_1G_1}(\lambda) = \frac{|f_{10}(\lambda)|^2}{f_{11}(\lambda)} \quad \text{and } f_{G_jG_j}(\lambda) = \frac{\left|A_{jj}\left(e^{i\lambda}\right)\right|^2 \det\left(\boldsymbol{f}_j(\lambda)\right)}{\det\left(\boldsymbol{f}_{j-1}(\lambda)\right)} \quad \text{for } j = 2, \ldots, K.$$

ここで, $\boldsymbol{f}_j(\lambda) := (f_{ij}(\lambda))_{i,j=1,\ldots,j}$ と $\boldsymbol{f}^{\dagger}_{a,b} := (f_{1b}(\lambda), \ldots, f_{ab}(\lambda))^{\top}$ に対して,

$$A_{jj}\left(e^{i\lambda}\right) := \frac{-\sum_{i=1}^{j-1} \det\left(\overline{\boldsymbol{f}^{\ddagger}_{i,j}(\lambda)}\right) f_{i0}(\lambda) + \det\left(\boldsymbol{f}_{j-1}(\lambda)\right) f_{j0}(\lambda)}{\det\left(\boldsymbol{f}_j(\lambda)\right)},$$

$$\boldsymbol{f}^{\ddagger}_{i,j}(\lambda) := \left(\boldsymbol{f}^{\dagger}_{j-1,1}(\lambda), \ldots, \boldsymbol{f}^{\dagger}_{j-1,i-1}(\lambda), \boldsymbol{f}^{\dagger}_{j-1,j}(\lambda), \boldsymbol{f}^{\dagger}_{j-1,i+1}(\lambda), \ldots, \boldsymbol{f}^{\dagger}_{j-1,j-1}(\lambda)\right).$$

上で定義した $A_{KK}\left(e^{i\lambda}\right)$ の分子を $\Phi_K\left(\boldsymbol{f}(\lambda)\right)$ と書くとする. 次の仮説検定問題を考える:

$$H_0 : \int_{-\pi}^{\pi} \left|\Phi_K\left(\boldsymbol{f}(\lambda)\right)\right|^2 = 0, \text{ v.s. } K_0\text{:}H_0\text{が成立しない.}$$

これは, モデル (1) に $\{X_K(t)\}$ が含まれているかどうかに対応する. 検定統計量を

$$T_n := \frac{n}{\sqrt{M_n}} \int_{-\pi}^{\pi} \left|\Phi_K\left(\hat{\boldsymbol{f}}(\lambda)\right)\right|^2 \mathrm{d}\lambda - \hat{\mu}_{n,K}$$

と定義する. ここで, $\hat{\boldsymbol{f}}(\lambda)$ は $\boldsymbol{f}(\lambda)$ のカーネル推定量, $\hat{\mu}_{n,K}$ はバイアス調整項で

$$\hat{\mu}_{n,K} := \sqrt{M_n}\eta_{\omega,2} \int_{-\pi}^{\pi} \mathrm{tr}\left(\left.\frac{\partial \Phi_K\left(\hat{\boldsymbol{f}}(\lambda)\right)}{\partial \boldsymbol{Z}^{\top}}\right|_{\boldsymbol{Z}=\hat{\boldsymbol{f}}(\lambda)} \hat{\boldsymbol{f}}(\lambda) \overline{\left.\frac{\partial \Phi_K\left(\hat{\boldsymbol{f}}(\lambda)\right)}{\partial \boldsymbol{Z}}\right|_{\boldsymbol{Z}=\hat{\boldsymbol{f}}(\lambda)}} \hat{\boldsymbol{f}}(\lambda) \right) \mathrm{d}\lambda,$$

$\eta_{\omega,2} := \int_{-\infty}^{\infty} \omega^2(x)\mathrm{d}x < \infty$, $\omega(x)$ はカーネル推定量に含まれるラグウィンドウ. 仮定 1 とその他の正則条件, 帰無仮説 H_0 の下で, T_n は平均ゼロで分散 σ_K^2 の正規分布に分布収束する. ここで, $\eta_{\omega,4} := \int_{-\infty}^{\infty} \omega^4(x)\mathrm{d}x < \infty$,

$$\sigma_K^2 := 4\pi\eta_{\omega,4} \int_{-\pi}^{\pi} \left| \mathrm{tr}\left(\left.\frac{\partial \Phi_K\left(\boldsymbol{f}(\lambda)\right)}{\partial \boldsymbol{Z}^{\top}}\right|_{\boldsymbol{Z}=\boldsymbol{f}(\lambda)} \boldsymbol{f}(\lambda) \overline{\left.\frac{\partial \Phi_K\left(\boldsymbol{f}(\lambda)\right)}{\partial \boldsymbol{Z}}\right|_{\boldsymbol{Z}=\boldsymbol{f}(\lambda)}} \boldsymbol{f}(\lambda) \right) \right|^2$$
$$+ \left| \mathrm{tr}\left(\left.\frac{\partial \Phi_K\left(\boldsymbol{f}(\lambda)\right)}{\partial \boldsymbol{Z}^{\top}}\right|_{\boldsymbol{Z}=\boldsymbol{f}(\lambda)} \boldsymbol{f}(\lambda) \left.\frac{\partial \Phi_K\left(\boldsymbol{f}(\lambda)\right)}{\partial \boldsymbol{Z}^{\top}}\right|_{\boldsymbol{Z}=\boldsymbol{f}(\lambda)} \boldsymbol{f}(\lambda) \right) \right|^2 \mathrm{d}\lambda.$$

したがって, 有意水準 α に対して, $T_n/\hat{\sigma}_K \geq z_\alpha$ を満たすとき帰無仮説 H_0 を棄却する検定は漸近サイズ α である. ここで, z_α は標準正規分布の上側 α%点, $\hat{\sigma}_K^2$ は σ_K^2 の一致推定量.

さらに講演では提案検定手法が 一致性を持つことも証明した. 講演では, 数値実験の結果と脳の機能的結合への応用例について紹介した.

References

Goto, Y., Zhang, X., Kedem, B., and Chen, S. (2023). Residual Spectrum Applied in Brain Functional Connectivity. ArXiv E-prints, Available at arXiv:2305.19461.

Prediction-based statistical inference for multiple time series

Yan Liu [1]

Abstract

A new minimum contrast estimator is proposed in this work for multivariate time series in the frequency domain. It is based on the prediction problem of time series. The properties of the proposed contrast estimation function are explained in details. We also derive the asymptotic normality of the proposed estimator and compare the asymptotic variance with the existing results. The asymptotic efficiency of the proposed minimum contrast estimation is also considered. The theoretical results are illustrated by some numerical simulations.

keywords: Minimum contrast estimation, Prediction problem, Spectral analysis, Multiple time series, Spectral density matrix, Asymptotic efficiency

1. Prediction-based contrast function

Let $\{\boldsymbol{z}(t)\}$ be a second-order stationary process

$$\boldsymbol{z}(t) = \sum_{j=0}^{\infty} G(j)\boldsymbol{e}(t-j), \qquad t \in \mathbb{Z}. \tag{1.1}$$

Assuming $\sum_{j=0}^{\infty} \operatorname{tr} G(j)KG(j)^{\top} < \infty$, the process has a spectral density matrix

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi}\boldsymbol{k}(\omega)K\boldsymbol{k}(\omega)^*, \qquad \omega \in \Lambda := (-\pi, \pi],$$

where $\boldsymbol{k}(\omega) = \sum_{j=0}^{\infty} G(j)\mathrm{e}^{\mathrm{i}j\omega}$.

We propose a new contrast function $\mathscr{D}$ for multivariate time series analysis as follows:

$$\mathscr{D}(\boldsymbol{f}_{\boldsymbol{\theta}}, \boldsymbol{g}) = \int_{-\pi}^{\pi} a(\boldsymbol{\theta}) \operatorname{tr}[|\boldsymbol{f}_{\boldsymbol{\theta}}(\omega)|^p \boldsymbol{g}(\omega)] \, \mathrm{d}\omega, \tag{1.2}$$

where $|\boldsymbol{f}_{\boldsymbol{\theta}}(\omega)| := \sqrt{\boldsymbol{f}_{\boldsymbol{\theta}}(\omega)^* \boldsymbol{f}_{\theta}(\omega)}$ is the square root of the matrix $\boldsymbol{f}_{\boldsymbol{\theta}}(\omega)^* \boldsymbol{f}_{\boldsymbol{\theta}}(\omega)$, and

$$a(\boldsymbol{\theta}) = \begin{cases} \left(\int_{-\pi}^{\pi} \operatorname{tr}[|\boldsymbol{f}_{\boldsymbol{\theta}}(\omega)|^{p+1}] \, \mathrm{d}\omega \right)^{-p/(p+1)}, & p \neq -1, \\ 1, & p = -1. \end{cases}$$

Lemma 1.1. *Let $p < 0$. Suppose the matrices $\boldsymbol{f}_{\boldsymbol{\theta}}$ and $\boldsymbol{g}$ are both Hermitian and semi-definite in any $\omega \in \Lambda$. The following inequality holds:*

$$\mathscr{D}(\boldsymbol{f}_{\boldsymbol{\theta}}, \boldsymbol{g}) \geq \left(\int_{-\pi}^{\pi} \operatorname{tr}[|\boldsymbol{g}(\omega)|^{p+1}] \, \mathrm{d}\omega \right)^{1/(p+1)}. \tag{1.3}$$

[1]This work was supported by JSPS Grant-in-Aid for Scientific Research (C) 23K11018.

2. Statistical inference

In this section, we discuss the parameter estimation problem based on the new contrast function (1.2). The following assumptions are imposed.

Assumption 2.1.

(i) The parameter space Θ is a compact subset of $\mathbb{R}^d$.
(ii) If $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $\boldsymbol{f}_{\boldsymbol{\theta}_1} \neq \boldsymbol{f}_{\boldsymbol{\theta}_2}$ on a set of positive Lebesgue measure.
(iii) The spectral density matrix $\boldsymbol{f}_{\boldsymbol{\theta}}(\omega)$ is three times continuously differentiable with respect to $\boldsymbol{\theta}$ and the second derivative $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\boldsymbol{f}_{\boldsymbol{\theta}}(\omega)$ is continuous in ω.

Lemma 2.2.
Under Assumption 2.1, when $p < 0$, $\boldsymbol{\theta}_0$ minimizes the disparity $\mathscr{D}(\boldsymbol{f}_{\boldsymbol{\theta}}, \boldsymbol{f}_{\boldsymbol{\theta}_0})$ if $\boldsymbol{\theta}_0 \in \Theta$.

Theorem 2.3. *Under regularity conditions, if $\boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}$, then the following holds:*

(a) $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$.
(b) $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, H(\boldsymbol{\theta}_0)^{-1}V(\boldsymbol{\theta}_0)H(\boldsymbol{\theta}_0)^{-1})$.

Assumption 2.4.

- The Gaussian-Fisher information matrix $\mathscr{F}(\theta_0)$ is invertible.
- The eigenvectors of $\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)$ do not depend on the parameter $\boldsymbol{\theta}$.

Theoretically, it can be shown under assumptions that

$$\mathscr{F}(\theta_0)^{-1} \leq H(\theta_0)^{-1}\tilde{V}(\theta_0)H(\theta_0)^{-1},$$

where the equality holds then $\alpha = -1$, or the spectral density matrices does not depend on ω.

3. Numerical simulations

In this section, we present the numerical performance of our minimum contrast estimator for multivariate time series. The data are generated from the following models:

Model (i)

$$\boldsymbol{X}_t = \begin{pmatrix} 0.2 & \theta_1 \\ 0 & 0.4 \end{pmatrix} \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{2\times 2});$$

Model (ii)

$$\boldsymbol{X}_t = \begin{pmatrix} 0.2 & \theta_2 \\ 0 & 0.4 \end{pmatrix} \boldsymbol{X}_{t-1} + \begin{pmatrix} 0 & 0 \\ \theta_2/3 & 0 \end{pmatrix} \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{2\times 2}),$$

where $\boldsymbol{I}_{2\times 2}$ denotes the identity matrix of size 2.

The finite sample performance of the new minimum contrast estimators is shown by the Q-Q plots, and a thorough investigation based on different parameters $\theta_1 = 0, 0.2, 0.4, 0.6, 0.8, 1$ and $\theta_2 = 0, 0.2, 0.4, 0.6, 0.8, 1$. The size of the data is $n = 100$. We conclude that the estimators when $p = -2, -3, -4$ concentrate around 0 more than the Whittle estimator (when $p = -1$).

References

Liu, Y. (2023). A Minimum Contrast Estimation for Spectral Densities of Multivariate Time Series. In *Research Papers in Statistical Inference for Time Series and Related Models* 325–342. Springer.

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

Junichi Hirukawa and Kou Fujimori
Niigata University and Shinshu University

ABSTRACT

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

1 Introduction

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

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

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

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

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

1.1 The fundamental martingale M

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

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

where

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

and B is the beta funtion

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

Then, the centered Gaussian process

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

has independent increments and variance function

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

where

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

In particular, M is a martingale.

2 Weak convergence of $I(d)$ process

Now, we obtain the following functional central limit result for $I(d)$ process, if $d > 0$,

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

References

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

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

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

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