科研費シンポジウム

「多様な分野における統計科学に関する諸問題」
（Problems related to statistical science in various fields）

日時：2019年9月14日（土）〜9月16日（月）
(Date: Saturday, 14, – Monday, 16, September, 2019)

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

科学研究費・基盤研究（A)（課題番号：15H01678）
「大規模複雑データの理論と方法論の総合的研究」
研究代表者：青嶋 誠（筑波大学）
Makoto Aoshima (University of Tsukuba)
開催責任者：蛭川 潤一（新潟大学）
Junichi Hirukawa (Niigata Universiry)

Program

Saturday, 14, September

Reception 13:00-13:25

Opening 13:25-13:30 **Junichi HIRUKAWA** (Niigata University)

Afternoon Session I (in Japanese) 13:30-15:00
Chair **Kou Fujimori** (Waseda University)
1. 13:30-14:15 **黒木　裕鷹 (KUROKI, Yutaka)**
東京理科大学大学院工学研究科 (Graduate School of Engineering, Tokyo University of Science)
「レーティング手法を用いたネットワーク中心性の構築とその統計的性質」
(Constructing Network Centrality Measure based on Rating Methods and its Statistical Properties)
2. 14:15-15:00 **KOTSUBO, Takuto**
Graduate School of Engineering, Tokyo University of Science
Hidden Markov models for cylindrical data and its application for animal movement analysis

Coffee Break 15:00-15:15

Afternoon Session II (in Japanese) 15:15-16:45
Chair **Takayuki Shiohama** (Tokyo University of Science)
3. 15:15-16:00 **Nakayama Yugo**
Graduate School of Pure and Applied Sciences, University of Tsukuba
Support vector machine and optimal choice of parameters for high-dimensional imbalanced data
4. 16:00-16:45 **新村秀一 (Shinmura Shuichi)**
成蹊大学名誉教授 (Emeritus Professor of Seikei Univ.)
「高次元遺伝子データ解析の呪いからの解放」
(Release of Curse of High-dimensional Data Analysis)

Sunday, 15, September

Morning Session (in English) 9:15-11:30
Chair **Muneya Matsui** (Nanzan University)
5. 9:15-10:00 **Yuichi Goto**, Marc Hallin, Masanobu Taniguchi
Waseda University, Université libre de Bruxelles, Waseda University
Kolmogorov-Smirnov Tests for Laplace Spectral Density Kernels
6. 10:00-10:45 **Kou Fujimori**, Sota Sakamoto and Yasutaka Shimizu
Waseda University
Generalized maximum composite likelihood estimator for determinantal point processes
7. 10:45-11:30 **Junichi Hirukawa** and Kou Fujimori
Niigata University and Waseda University
Weak convergence of the partial sum of $I(d)$ process to a fractional Brownian motion in finite interval representation

Lunch 11:30-13:15

Afternoon Session I (in English) 13:15-15:30
Chair **Nobuaki, HOSHINO** (Kanazawa University)
8. 13:15-14:00 **Toshihiro Abe**
Nanzan University
A closed form EM algorithm for a multivariate skew-normal model
9. 14:00-14:45 **Muneya Matsui**
Nanzan University
Asymptotics of maximum likelihood estimation for stable law with continuous parameterization
10. 14:45-15:30 **張　元宗 (Chang Yuan Tsung)**, 篠崎　信雄 (Nobuo Shinozaki), William, E. Strawderman
目白大学・社会学部・社会情報学科（Mejiro University, Department of Social Information, Faculty of Studies on Contemporary Society),
慶應大学・理工学部 (Faculty of Science and Technology, Keio University),
Department of Statistics and Biostatistics, Rutgers University
Pitman Closeness Domination in Predictive Density Estimation for Two Ordered Normal Means Under α-Divergence Loss

Coffee Break 15:30-15:45

Afternoon Session II (Guest speakers session) 15:45-16:45

Chair **Junichi Hirukawa** (Niigata University)

11. 15:45-16:45 **Konstantinos Fokianos**

Lancaster University

Auto-Distance Covariance Function for Time Series Analysis

Monday, 16, September

Morning Session (in Japanese) 9:15-11:30
Chair **Toshihiro Abe** (Nanzan University)
12. 9:15-10:00 **入江　薫 (Kaoru Irie)**
東京大学経済学部 (Faculty of Economics, The Univeristy of Tokyo)
「縮小事前分布と状態空間モデル」
(Shrinkage priors and state space models)
13. 10:00-10:45 **永井 勇 (Isamu Nagai)**
中京大学 国際教養学部（School of International Liberal Studies, Chukyo University）
「バランス型経時測定データにおける Extended GMANOVA モデルの解釈と新たな推定法」
(Interpretation for the extended GMANOVA model in the balance type longitudinal data, and new estimation method)
14. 10:45-11:30 **星野 伸明 (Nobuaki, HOSHINO)**
金沢大学経済学類 (School of Economics, Kanazawa University)
「分散可変な一般化多項分布について」
(Generalized multinomial distributions with scalable variance)

Closing 11:30-11:35 **Junichi HIRUKAWA** (Niigata University)

レーティング手法を用いたネットワーク中心性の構築とその統計的性質

黒木裕鷹
東京理科大学大学院工学研究科経営工学専攻

塩濱敬之
東京理科大学工学部情報工学科

概 要

ソーシャル・ネットワーキング・サービス (SNS) の普及と定着により, 人間関係におけるコメントや意見のやり取りを記録したネットワークデータが大量に蓄積されている. ネットワーク分析において, 中心的な役割を果たす個人 (ノード) を特定することが重要な課題であり, これまで様々な中心性指標が提案されてきた. 代表的な中心性指標のほとんどはネットワークの構造に由来するが, 正規方程式に基づいたレーティングアルゴリズムである Massey のレーティングは, エッジの重みを考慮した中心性として解釈することができる. 本研究では, Massey のレーティングを中心性指標として再解釈し, その統計的性質を検討する.

1 はじめに

近年, 価値や金銭のやりとり, 交友関係に着目した社会的ネットワークの分析が, グラフ理論や統計的手法を用いて盛んに研究されている. 特に, 情報技術の発展により現実世界に存在する巨大で複雑なネットワークの性質を明らかにする複雑ネットワーク分析が注目されている. 複雑ネットワークの分析を通じて, ネットワークを構成するコミュニティの理解や位置付け, 将来時点のネットワーク特性の予測が可能となることにより, ソーシャルネットワーキングサービス (SNS) やブロックチェーン技術を利用した情報サービスの新たな価値創造が期待されている. 社会ネットワークを構成する個人や組織をノード (頂点), ノード間の関係性をエッジ (辺) といい, ノードとエッジで構成されるグラフはネットワークの社会的相互作用を記述する. 相互関係のあるデータにおいて重要な課題は, ネットワーク内で重要な役割を果たすノードを把握することにある. このような重要性の指標は中心性 (centrality) と呼ばれ, 主に社会科学の分野で様々な中心性指標が提案されてきた. たとえば, Bonacich (1987) や Katz (1953) を参考にすること.

代表的な中心性指標には, 次数中心性, 媒介中心性, 近接中心性, 固有ベクトル中心性などが知られている. これらの中心指標はネットワークの構造から直接求めることができるがノードの特性やエッジの重みは考慮されない. たとえば, SNS における「賛成」「反対」の絶対数や, 投げ銭システムに代表されるようなトークンの付与回数などを考慮した分析において, このような中心性指標を用いてノードの重要性を評価することはできない. 関係性のセマンティクスを考慮した, 非対称な有向グラフを用いたネットワーク指標を構成するには, レーティングシステムやランキングアルゴリズムを利用した中心性の指標の構築が必要となる. このような研究例として, スポーツやゲームの試合結果を利用するレーティングシステムや Page Rank に代表されるランキングシステムがあり, 検索エンジンの設計や推薦システム, 影響力のある拡散者の特定など, 広範囲に応用されている.

スポーツやゲームのデータ解析では,「強さ」の指標であるレーティングの開発が盛んに行われており, 算出されたレートは結果の予測やトーナメントの出場プレーヤーの選出など, 様々な場面で利用される. また, Page Rank のアルゴリズム (Brin and Page, 1998) は, Google 検索エンジンにおいて Web ページをランク付けするために考案された. スポーツやゲームにおけるレーティングは, 一対比較モデルを応用して勝敗確率を予測する Bradley-Terry モデル (Bradley and Terry, 1952), チェスに端を発する Elo レーティング (Elo, 1978), ラプラスの継起則を応用した Colley のレーティング (Colley, 2002) などが存在する. 全米大学体育協会 (NCAA) の主催するバスケットボールトーナメントの出場チームの選出には, アルゴリズムは非公開ではあるものの, シーズン中のデータに基づいた指標である Rating Percentage Index (RPI) が利用されている. 動的なレーティングの更新が可能な Elo レーティングはもともとチェスの実力を表すために考案されたものだが, 世界ラグビー協会は修正した手法を採用している. また, 正規方程式に基づいたレーティング手法として, Massey のレーティング手法 (Massey, 1997) が知られている. Massey の手法は, 代表的なネットワーク中心性である Katz の中心性 (Katz, 1953) と関連する.

本研究では, Massey のレーティング手法をネットワーク中心性として再解釈し, その統計的性質を検証する. 本稿の構成は以下の通りである. 第 2 章では, 代表的な中心性指標をいくつか紹介し, 各指標の特徴を述べる. 第 3 章では Massey のレーティング手法について説明し, 中心性指標としての解釈, 特に Katz の中心性との関連を考察する. また, レーティング手法の試合結果に対する感度に関する先行研究も取り上げる. 第 4 章ではデータ解析の例として, 2018 年テニス ATP ワールドツアーの試合結果を解析し, その統計的性質を明らかにする. 最後に, 今後の課題や展望について第 5 章で述べる.

参考文献

Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, 92(5):1170–1182.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117.

Colley, W. N. (2002). Colley ’s bias free college football ranking method: The colley matrix explained. *Princeton University, Princeton*.

Elo, A. E. (1978). *The rating of chessplayers, past and present*. Arco Pub.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43.

Massey, K. (1997). Statistical models applied to the rating of sports teams. *Bluefield College*.

Hidden Markov models for cylindrical data and its application for animal movement analysis

Takuto Kotsubo
Graduate School of Engineering, Tokyo University of Science
Takayuki Shiohama
Department of Information and Computer Technology

Abstract

Analyzing animal tracking is a growing field in ecology and various models have been proposed in literature. State space model is often used to fit the animal tracking data, however the limitations of the linear and Gaussian assumption are sometimes reported. An alternative approach for modeling animal movement data is Hidden Markov Models (HMM) with cylindrical distribution, where the cylindrical data consists of circular and linear variables. For linear part of the cylindrical distribution, we adopt the generalized Palate-type distribution for heavy tailed observations in linear part, which includes the Weibull-von Mises distributions on cylinder. Estimation for the model parameters are done by using Expectation and Maximization (EM) algorithms, and we propose a modified M-step procedure and a Maximum at posteriori (MAP) estimation for model parameters to avoid local maximum solutions in EM algorithms.

Keywords: Animal movement, cylindrical data, EM-algorithms, maximum likelihood estimation.

1 Introduction

A cylindrical data refers to a bivariate data which consists of circular and linear variables. Circular data usually deal with direction and has periodic properties with frequency 2π. Linear variables arise in cylindrical data are usually defined on a positive real line. Cylindrical data arise often in various fields in natural sciences, for example, wind directions and speeds in meteorology (Breckling (1989); Ailliot et al. (2015)), wave directions and heights in oceanology (Ris et al. (1999)), ozone concentrations and wind directions in environmental science (Camalier et al. (2007); Yi and Prybutok (1996)) and the turning angels and step lengths of the animal movement in ecology (Jonsen (2016); Adam et al. (2019)). These data consist of time series nature, however time series modeling of cylindrical data are not fully investigated in literature. The reason behind this is that there exists a few distributions on cylinder.

The well known cylindrical distribution is called the Johnson and Wehrly distribution (Johnson and Wehrly (1978)). The drawbacks of the Johnson and Wehrly distribution are reported in Abe and Ley (2017), where they pointed out that the mode of the linear variables always defined at 0, which can not be applied for many actual data analysis. Recently, Abe and Ley (2017) proposed cylindrical distribution with combining sine-skewed von Mises and Weibull distribution to implement the skewness structures in directional variables and to have the mode at some point on the domain of the probability distribution function. Imoto et al. (2019) extends the Abe-Ley distributions to have more heavy tail distributions on real part. There exists another class of the cylindrical distributions which includes the distributions of Mardia and Sutton (1978) and Kato and Shimizu (2008). The normalizing constants of these distributions are somewhat complex and that is not expressed in an analytic form, hence in this study we focus on applying the distributions of Abe and Ley (2017) and Imoto et al. (2019).

The rest of paper is organized as follows. Section 2 introduces our HMM models. Several useful distributional properties of cylindrical distributions are explained. Section 3 provides the maximum likelihood estimation procedures which utilizing EM algorithms. Some technical improvements are also discussed to avoid the local maximum solution in optimization together with to avoid unboundedness in likelihood functions. In Section 4, some Monte Carlo simulations are performed to assess the performance of our proposed procedures. Section 5 provides real data analysis of animal movement trajectories. Section 6 concludes our paper.

References

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

Adam, T., C. A. Griffiths, V. Leos-Barajas, E. N. Meese, C. G. Lowe, P. G. Blackwell, D. Righton, and R. Langrock (2019). Joint modelling of multi-scale animal movement data using hierarchical hidden markov models. to appear in *Methods in Ecology and Evolution*.

Ailliot, P., J. Bessac, V. Monbet, and F. Pene (2015). Non-homogeneous hidden markov-switching models for wind time series. *Journal of Statistical Planning and Inference 160*, 75–88.

Breckling, J. (1989). *The analysis of directional time series: applications to wind speed and direction*, Volume 61. Springer Science & Business Media.

Camalier, L., W. Cox, and P. Dolwick (2007). The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmospheric Environment 41*(33), 7127–7137.

Imoto, T., K. Shimizu, and T. Abe (2019). A cylindrical distribution with heavy-tailed linear part. to appear in *Japanese Journal of Statistics and Data Science*.

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

Jonsen, I. (2016). Joint estimation over multiple individuals improves behavioural state inference from animal movement data. *Scientific Reports 6*, 20625.

Kato, S. and K. Shimizu (2008). Dependent models for observations which include angular ones. *Journal of Statistical Planning and Inference 138*(11), 3538–3549.

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

Ris, R., L. Holthuijsen, and N. Booij (1999). A third-generation wave model for coastal regions: 2. verification. *Journal of Geophysical Research: Oceans 104*(C4), 7667–7681.

Yi, J. and V. R. Prybutok (1996). A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environmental pollution 92*(3), 349–357.

Support vector machine and optimal choice of parameters for high-dimensional imbalanced data

Yugo Nakayama
Graduate School of Pure and Applied Sciences, University of Tsukuba

1 Introduction

In this talk, we considered the classification for high-dimensional data. Suppose we have independent and d-variate two populations, $\pi_i,\ i=1,2$, having an unknown mean vector $\boldsymbol{\mu}_i$ and unknown covariance matrix $\boldsymbol{\Sigma}_i\ (\geq \boldsymbol{O})$. Let $\Delta=\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|^2$, where $\|\cdot\|$ denotes the Euclidean norm. We have independent and identically distributed (i.i.d.) observations, $\boldsymbol{x}_{i1},\dots,\boldsymbol{x}_{in_i}$, from each π_i. We assume $n_i\geq 2,\ i=1,2$ and $n_1\leq n_2$. Let $\boldsymbol{x}_0$ be an observation vector of an individual belonging to one of the two populations. We assume $\boldsymbol{x}_0$ and $\boldsymbol{x}_{ij}$s are independent. Let $n=n_1+n_2$.

In recent years, the margin-based classification methods such as support vector machine (SVM) are studied for high-dimensional data. It is well known that many classifiers, including SVM, overfit the data in the high-dimensional setting. Nakayama et al. (2019) claimed that SVM gives inconsistent properties for imbalanced data:

$$e(1)=1+o_P(1)\ \text{ and }\ e(2)=o_P(1)\ \text{ as } d\to\infty;\quad \text{or} \tag{1}$$

$$e(1)=o_P(1)\ \text{ and }\ e(2)=1+o_P(1)\ \text{ as } d\to\infty, \tag{2}$$

where $e(i)$ denote the error rate of misclassifying an individual from π_i into the other class for $i=1,2$. The cause of inconsistency is due to the imbalance of sample sizes and a huge bias of high-dimensional space. In overcome this difficulty, Nakayama (2019) proposed linear robust SVM (RSVM) to imbalanced data. In this paper, we investigate the asymptotic properties of RSVM more deeply in the framework of $n/d\to 0$ as $m\to\infty$, where $m=\min\{d,n_1,n_2\}$.

2 Asymptotic properties of robust SVM

In this section, we compare SVM and RSVM for high-dimensional data. First, we introduce a formulation of SVM. Let $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)=(\boldsymbol{x}_{11},\dots,\boldsymbol{x}_{1n_1},\boldsymbol{x}_{21},\dots,\boldsymbol{x}_{2n_2})$. $\#S$ denotes the number of elements in a set S. The discriminant function of SVM is given by

$$y(\boldsymbol{x})=\sum_{j\in\hat{S}}\hat{\alpha}_j t_j\boldsymbol{x}_j^T\boldsymbol{x}+\frac{1}{n_{\hat{S}}}\sum_{j\in\hat{S}}\Big(t_j-\sum_{j'\in\hat{S}}\hat{\alpha}_{j'}t_{j'}\boldsymbol{x}_j^T\boldsymbol{x}_{j'}\Big), \tag{3}$$

where $\hat{S}=\{j|\hat{\alpha}_j\neq 0,\ j=1,\dots,n\}$ and $n_S=\#S$. $\hat{\boldsymbol{\alpha}}=(\hat{\alpha}_1,\dots,\hat{\alpha}_n)$ is obtained by solving the following maximization

$$\hat{\boldsymbol{\alpha}}=\underset{\boldsymbol{\alpha}}{\text{argmax}}\left(\sum_{j=1}^n\alpha_j-\frac{1}{2}\sum_{j=1}^n\sum_{k=1}^n\alpha_j\alpha_{j'}t_jt_{j'}\boldsymbol{x}_j^T\boldsymbol{x}_{j'}\right)$$

subject to

$$0\leq\alpha_j\leq C,\ j=1,\dots,n,\ \text{ and }\ \sum_{j=1}^n\alpha_jt_j=0, \tag{4}$$

where $C > 0$ is a regulariation parameter. Note that (3) becomes a hard-margin type when $C \to \infty$. There exist some $\boldsymbol{x}_j$s satisfying that $t_j y(\boldsymbol{x}_j) = 1$ (i.e., $\hat{\alpha}_j \neq 0$). Such $\boldsymbol{x}_j$s are called the support vector. SVM classifies $\boldsymbol{x}_0$ into π_1 if $y(\boldsymbol{x}_0) < 0$ and into π_2 otherwise.

The regularization parameter C influences the properties of SVM. Let $\Delta_* = \Delta + \mathrm{tr}(\boldsymbol{\Sigma}_1)/n_1 + \mathrm{tr}(\boldsymbol{\Sigma}_2)/n_2$. For C, we consider the following conditions:

$$\limsup_{m\to\infty} \frac{2}{n_1 \Delta_* C} \le 1; \textit{ or} \tag{5}$$

$$\liminf_{m\to\infty} \frac{2}{n_1 \Delta_* C} > 1. \tag{6}$$

Let $\delta = \mathrm{tr}(\boldsymbol{\Sigma}_1)/n_1 - \mathrm{tr}(\boldsymbol{\Sigma}_2)/n_2$. Then we have the following result.

Theorem 1. *Assume some regularity conditions and*

$$\liminf_{m\to\infty} |\delta|/\Delta < 1. \tag{7}$$

Then, under (5), (3) holds (8).

$$e(i) = 1 + o_P(1) \quad \textit{as } m \to \infty \textit{ for } i = 1, 2. \tag{8}$$

Under (6), we have the following result.

Theorem 2. *Assume some regularity conditions and*

$$\liminf_{m\to\infty} \frac{n_1^2 C \Delta}{n} > 0; \ \textit{ and} \tag{9}$$

$$\limsup_{m\to\infty} \Big(\frac{n_{i'} - n_i}{n_1 n_i \Delta C} + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_i) - \mathrm{tr}(\boldsymbol{\Sigma}_{i'})}{n_i \Delta} \Big) < 1, \quad i \neq i'. \tag{10}$$

Then, under (6), (3) holds (8).

SVM gives (1) and (2) for high-dimensional imbalanced data. We improved SVM by using a robust intercept to imbalanced data. Let $S_i = \{j | \hat{\alpha}_{ij} \neq 0,\ j = 1, \dots, n_i\}$ for $i = 1, 2$. Then we define robust SVM:

$$y_r(\boldsymbol{x}_0; \boldsymbol{w}) = \sum_{j\in S} \hat{\alpha}_j t_j \boldsymbol{x}_j^T \boldsymbol{x}_0 + \sum_{j\neq j'(j,j'\in S_1)} \frac{\hat{\alpha}_{1j} \boldsymbol{x}_{1j}^T \boldsymbol{x}_{1j'}}{2(n_{S_1} - 1)} - \sum_{j\neq j'(j,j'\in S_2)} \frac{\hat{\alpha}_{2j} \boldsymbol{x}_{2j}^T \boldsymbol{x}_{2j'}}{2(n_{S_2} - 1)} \tag{11}$$

Note that $\sum_{j=1}^{n_1} \hat{\alpha}_{1j} = \sum_{j=1}^{n_2} \hat{\alpha}_{2j} (= \widehat{A},$ say) from (4). We have the following theorem.

Theorem 3. *Under some regularity conditions, (11) holds (8).*

RSVM gives (8) for imbalanced data and any C satisfying (9) without assuming (7) and (10).

In this talk, we generalized RSVM for kernel functions and discussed the choice of tuning parameters. Finally, we checked the performance of RSVM and the validity of the tuning parameter by numerical simulations and actual data analysis.

References

[1] Nakayama, Y. (2019). Robust support vector machine for high-dimensional imbalanced data. *Communications in Statistics - Simulation and Computation, in press (doi: 10.1080/03610918.2019.1586922).*

[2] Nakayama, Y., Yata, K., and Aoshima, M. (2019). Bias-corrected support vector machine with gaussian kernel in high-dimension, low-sample-size settings. *Annals of Institute of Mathematical Statistics, in press (doi: 10.1007/s10463-019-00727-1).*

報告書：多様な分野における統計科学に関する諸問題

成蹊大学名誉教授
新村秀一

「高次元 Microarray データを用いて癌遺伝子の特定と癌の亜種を見つける研究」が 1970 年から行われてきた。これらの研究で用いられたデータが公開されているので、統計に限らず工学系の機械学習（AI）、パターン認識、Bio 工学の新テーマとして研究されてきた。本研究に関する医学以外の多くの研究が「3 つの困難」を指摘していることが示す通り、いずれも成功していない。

しかし、仮に症例数が n=100 で遺伝子数が p=10,000 の発現量データとすれば、2 群判別が最も適した手法である。筆者は 2015 年 10 月 28 日から 12 月 25 日の僅か 54 日間で、簡単にこの問題を解決した。用いたデータは、1999 年から 2004 年の間に米国の 6 医学研究プロジェクトが Science や The New England Journal of Medicine に論文を発表し、研究に用いた Microarray の公開データである。これらは Alon と Singh は癌と健常の 2 クラス、Golub らの 4 種は異なった 2 種の癌のクラスである。異なった癌種の判別であるが、ほぼ共通の結果が得られた。

結果は非常に単純である。6 種のデータは線形分離可能なデータ（Linearly Separable Data, LSD）である（判別分析の新事実 Fact3）。この重要な事実であり信号が、これまでの研究では指摘していない。唯一、青嶋、矢田、石井らが、これ等の 6 種を含む 10 種以上の Microarray データで 2 群が 2 つの異なった球体上に布置していることを高次元 PCA 等の研究で指摘している。さらに、筆者が開発した Matryoshka Feature SelectioNMethod（Method2）で簡単に線形分離可能な含まれる遺伝子数が n 個以下の遺伝子の k 組の部分空間(SMall Matryoshka, SM)と最小誤分類数(Minimum Number of Misclassifications, MNM)が 1 以上の雑音の遺伝子の部分空間に分割できた（判別分析の新事実 Fact4）。各 SM は統計分析が得意とする小標本であるが、ロジスティック回帰以外、線形分離可能な事実が示されなかった。そこで、MNM 基準による改定 IP-OLDF（RIP）の判別スコア（RIP Discriminant Score, RipDS）を変数とし、n*k 次元の信号データ（k<=n）を作成した。これを一元配置の分散分析、 t 検定、相関分析、クラスター分析、PCA で癌の遺伝子診断と Malignancy Indexes を世界で初めて提案でき、今年 5 月に Springer から 2 冊目の”High-dimensional Microarray Data Analysis- Cancer Gene Diagnosis and Malignancy Indexes by Microarray”を出版した。

以上の研究が簡単にできたのは、大学卒業以来行ってきた判別分析の新理論(New Theory of Discriminant Analysis After R. Fisher, Springer 2016)が 2015 年に完成し、新理論がその応用問題として Microarray を用いた癌の遺伝子解析（判別分析の問題 5）を簡単に解決できたためである。本来であれば癌の遺伝子研究の専門家でない筆者が「癌の遺伝子診断」までを行うことは適していないことは十分理解している。しかし、癌は遺伝子の病気であり、高次元の Microarray 空間で 2 群が完全に分かれていて、かつ MNM＝0 である k 個の SM に分割できる。そして信号データを作ることで、統計手法で線形分離可能であり有効と考えられる結果が数多く発見できた。しかし、これ等のどの結果が医学的に役に立つか否かは医学専門家の検証が必要である。残念ながら Golub らの研究後に、「NIH が乳がん以外の癌に関してこの種研究は成果が出ないと判断し、医学研究が終わったようである」。今、いかに医学専門家の

検証につなげるかを模索している。しかし NIH の報告を知らずに研究を続けている工学研究者の在り方は、一般的に問題であろう。また、データが LSD であるのに、そのデータを学習標本に用いる AI 研究が LSD の事実を指摘しない点だけが、まだ説明できていない。

大学卒業以来の研究テーマである判別の新理論を確立し、その応用として「高次元 Microarray の癌の遺伝子解析と診断」に成功した。そこでこれまでの研究を見直した結果、LSD である高次元データは、ケース数 n 個以下の遺伝子の k 組の小標本に必ず分割できるという事実が統計にとって一番重要と考えた。すなわち、我々は「高次元データの呪いから数理計画法(MP)の LP と IP で解放される。そして、分割された SM を統計分析することでこれまで見えてこなかった新しい世界が広がる（QP で定式化した SVM ではできない）」ことになる。以上の重要なテーマを以下の 4 回のシンポジュームで報告したい。

1. 九州大学：「高次元遺伝子解析の呪いからの解放 1　-統計が 1970 年からこの問題を解決できなかった理由-」
2. 新潟大学：「高次元遺伝子解析の呪いからの解放 2　-癌の遺伝子診断-」
3. 東京工業大学：「高次元遺伝子解析の呪いからの解放 3　-機械学習などの工学研究の問題点-」
4. 秋田大学：「高次元遺伝子解析の呪いからの解放 4　-高次元データの分割法の最新結果-」

本発表では、統計的判別関数がなぜ役に立たなかったかの理由を報告する。正規分布を仮定した分散共分散行列に基づく判別関数は、得点で試験の合否判定を行うと数学ⅠaやⅡbで誤分類確率が 2 割を超えるものがあることを既に報告した。そして Microarray では 17%も誤判別する。すなわち LSD 判別には全く役に立たないばかりか（判別分析の Problem2)、判別分析の基本統計の誤分類数(NM)も問題が多い（判別分析の Problem1)。このため、SM という小標本の判別分析でも分散共分散行列に基づく判別関数の Fisher の LDF，QDF，正則化判別関数（RDA)、LASSO は全く役に立たない。これに対して、最尤推定法で求めたロジスティック回帰は、Microarray では役に立たないが、全ての SM を NM＝0 で判別した。LASSO 研究者は、癌の遺伝子解析への応用を考えているが未だ成功していない。彼らは回帰係数や判別係数を 0 にした非ゼロのモデルを用いれば、変数選択の NP-Hard を克服でき問題が解決できると単純に考えている。MNM 基準による RIP では、筆者の学位論文に示す通り、LSD でない Fisher のアイリスデータで自然に多くの判別係数が 0 になることを示している。重要なのは LSD であるかどうかであり、次に判別係数が 0 になるかどうかという階層構造の優先度が全く理解されていない。ハードマージン最大化 SVM(H-SVM)と RIP だけが理論的に LSD を判別できる。筆者の研究では、筆者の開発した 3 種の最適判別関数(OLDF)とソフトマージン SVM(Ｓ-SVM)も LSD であることを確認した。医学研究でＳＶＭを用いた研究もあるが Fact3 を報告したものがないことを考察する。そして残念なことに、RIP と Revised LP-OLDF だけが、k 組の SM と雑音に簡単に分割できるが、2 次計画法(QP)を用いた SVM ができない理由を数理計画法の入門的な知識で説明する。一方、RIP が分割できる理由を筆者の定義した IP と LP モデルの実行可能領域と、線形代数の初歩的知識である連立方程式の解という初歩的知識の組み合わせで明快に説明する。すなわち 1970 年から医学研究者の努力にかかわらず、また多くの統計や工学研究者の研究でも成果が出なかった理由を考察する。そして何故 MP の IP と LP モデルだけが高次元データの呪いから我々を解放できたかを説明する。これによって、今後高次元の他のデータであっても、容易に統計分析の研究対象になることを説明する。研究は質が高く、2 群が LSD であるという検証しやすいデータを用いたことによって、はじめて役に立つ研究を退官後に完成できたことは医学データを研究対象としたことが幸運であった。

Kolmogorov-Smirnov Tests for Laplace Spectral Density Kernels

Yuichi Goto (Waseda University)
Marc Hallin (Université libre de Bruxelles)
Masanobu Taniguchi (Waseda University)

1. Introduction

A new type of spectral density, called the Laplace spectral density kernel, was proposed by [1]. Laplace spectral density kernels characterize the collection of all marginal bivariate distributions of a given stationary stochastic process, without requiring any moment assumptions. In this presentation, we proposed a Kolmogorov-Smirnov (KS) test for Laplace spectral density kernels.

2. Setting and Main Results

Let $\{X_t : t \in \mathbb{Z}\}$ be a strictly stationary process with continuous and strictly increasing marginal distribution function $F(\cdot)$. We assume $\{X_t\}$ satisfies the following assumption.

Assumption 1. There exists constants $\rho \in (0,1)$ and $K \in \mathbb{R}$ such that, for arbitrary intervals $A_1, \ldots, A_\ell \subset \mathbb{R}$ and arbitrary $t_1, \ldots, t_\ell \in \mathbb{Z}$,

$$\left|\mathrm{Cum}\left(\mathbb{I}_{\{Z_{t_1} \in A_1\}}, \ldots, \mathbb{I}_{\{Z_{t_p} \in A_\ell\}}\right)\right| \leq K\mathrm{P}(Z_t \in A_1) \ldots \mathrm{P}(Z_t \in A_\ell)\rho^{\max_{i,j\in\{1,\ldots,\ell\}}|t_i - t_j|}.$$

The Laplace cross-covariance kernel of $\{X_t\}$ is defined, for all $k \in \mathbb{Z}$ and $(x_1, x_2) \in \mathbb{R}^2$, as

$$c_{x_1,x_2}(k) := c_{F(x_1),F(x_2)}(k) := \mathrm{Cov}\left(\mathbb{I}_{\{X_{t-k} \leq x_1\}}, \mathbb{I}_{\{X_t \leq x_2\}}\right) = \mathrm{Cov}\left(\mathbb{I}_{\{U_{t-k} \leq F(x_1)\}}, \mathbb{I}_{\{U_t \leq F(x_2)\}}\right),$$

where $U_t := F(Z_t)$. Associated with this cross-covariance kernel, the *Laplace spectral density kernel* is defined, for $(x_1, x_2) \in \mathbb{R}^2$ and $\lambda \in [-\pi, \pi]$, as

$$f_{x_1,x_2}(\lambda) := f_{F(x_1),F(x_2)}(\lambda) := \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} c_{x_1,x_2}(k) \exp\{-ik\lambda\}.$$

In order to estimate the Laplace spectrum of $\{X_t\}$, the *Laplace periodogram* is defined, for $(x_1, x_2) \in \mathbb{R}^2$ and $\lambda \in [-\pi, \pi]$, as

$$I_n^{x_1,x_2}(\lambda) := I_{n,U}^{F(x_1),F(x_2)}(\lambda) := \frac{1}{2\pi n} d_n^{x_1}(\lambda) d_n^{x_2}(-\lambda),$$

where

$$d_n^x(\lambda) := \sum_{s=0}^{n-1} \mathbb{I}_{\{X_s \leq x\}} \exp\{-i\lambda s\} = \sum_{s=0}^{n-1} \mathbb{I}_{\{U_s \leq F(x)\}} \exp\{-i\lambda s\}.$$

This research supported by Grant-in-Aid for JSPS Research Fellow Grant Number JP201920060 (Y. Goto), and the Research Institute for Science & Engineering of Waseda University and JSPS Grant-in-Aid for Scientific Research (S) Grant Number JP18H05290 (M. Taniguchi).

Generalized maximum composite likelihood estimators for determinantal point processes

Kou Fujimori*, Sota Sakamoto and Yasutaka Shimizu
Waseda University

In this talk, we dealt with the estimation problem for determinantal point processes (DPPs). DPPs are the classes of spatial point processes with repulsive properties for each pair of realized points and studied intensively in terms of statistical physics to capture the behavior of fermions.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For the kernel function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, the point process X called determinantal point process with kernel K if the measure;

$$\mathbb{E}\left[\sum_{(x_1,\dots,x_p)\in X^p}^{\neq} 1_{\{x_1\in A_1,\dots,x_p\in A_p\}}\right], \quad p = 2, 3, \dots$$

where $A_j,\ j = 1, \dots, p$ are bounded Borel sets on $\mathbb{R}^d$ and X^p is p-direct product of the point process X, has the following density function (p-th order joint intensity function)

$$\rho^{(p)}(x_1, \dots, x_p) = \det[K](x_1, \dots, x_p),$$

with

$$[K](x_1, \dots, x_p) := (K(x_i, x_j))_{1\le i,j\le p} \in \mathbb{R}^{p\times p}.$$

We consider the case when the kernel K satisfies the following form:

$$K(x, y) = \sqrt{\rho(x, \lambda)\rho(y, \lambda)}C_\alpha(x - y),$$

where the function $\rho(\cdot, \lambda)$ is the intensity function with an unknown parameter $\lambda > 0$ and C_α is the pair correlation function which satisfies $C_\alpha(0) = 1$ with an unknown parameter $\alpha \in \mathbb{R}^q$. We denote this kernel $K = K_\theta$, where $\theta = (\lambda, \alpha)$.

Our goal was to construct estimators for unknown parameter θ. Since well-known point processes such as the Gibbs point process or Cox process only have the exact form of joint intensities of second order, we often use the second order estimation function such as composite likelihood functions. On the other hand, joint intensities of DPPs are given by the determinant of positive definite kernels, which allows us to compute the joint intensities of general order. Therefore, we can introduce the two step composite likelihood approach by using p-th order joint intensity function $\rho^{(p)}$ based on observations from fixed window $D_n \subset \mathbb{R}^d$ which is the generalized version of 2nd order estimation function. For the intensity parameter λ, we considered the following normalized quasi-likelihood function.

$$\mathbb{H}_{n1}(\lambda) = \int_{D_n} \log \rho(x, \lambda)N(dx) - \int_{D_n} \rho(u, \lambda)du,$$

where $D_n \subset \mathbb{R}^d$ is the observation window centered 0 which satisfies the following condition:

$$|D_n| \asymp n^d, \quad \mu_{d-1}(\partial D_n) \asymp n^{d-1}, \quad n \to \infty$$

with $|\cdot|$ and $\mu_{d-1}(\cdot)$ is the d and $d-1$-dimensional Lebesgue measure, respectively. We defined the following normalized p-th order composite likelihood function for every integer $p \geq 2$ to estimate the parameter α:

$$\begin{aligned} \mathbb{H}_{n2}^{(p)}(\lambda, \alpha) &:= \int_{D_n^p} \left\{ \log[\rho_\theta^{(p)}(x_1, \ldots, x_p)] - \log[K_{w,p}(r : \theta)] \right\} \\ &\quad \times w_r(x_1, \ldots, x_p) N^{(p)}(dx_1 \cdots dx_p), \end{aligned} \tag{1}$$

where $r > 0$ is a tuning parameter, $\rho_\theta^{(p)}$ and $K_{w,p}$ are respectively the joint intensity of p-th order of DPP(C_θ) and the modified K-function of p-th order:

$$\begin{aligned} \rho_\theta^{(p)}(x_1, \ldots, x_p) &= \det[K_\theta](x_1, \ldots, x_p), \\ &= \prod_{i=1}^{p} \rho(x_i, \lambda) \det[C_\theta](x_1, \ldots, x_p) \end{aligned}$$

and

$$K_{w,p}(r : \theta) = \int_{D_n^p} \rho_\theta^{(p)}(x_1, \ldots, x_p) w_r(x_1, \ldots, x_p) dx_1 \cdots x_p,$$

where w_r is a bounded weight function whose support is given by

$$S_r^p := \{(x_1, \ldots, x_p) : |x_1 - x_j| \leq r, \ 1 \leq j \leq p\}.$$

For example, a simple choice of the weight function w_r is given by

$$w_r(x_1, \ldots, x_p) = 1_{S_r^p}(x_1, \ldots, x_p).$$

Note that $N^{(p)}$ is a counting measure of p-th order induced by the point process X, i.e.,

$$N^{(p)}\left(\prod_{j=1}^{p} A_j\right) = \sum_{(x_1,\ldots,x_p)\in X^p}^{\neq} 1_{\{x_1 \in A_1, \ldots, x_p \in A_p\}},$$

where $A_j, \ j = 1, \ldots, p$ are bounded Borel sets on $\mathbb{R}^d$. Using the estimating functions $\mathbb{H}_{n1}(\lambda)$ and $\mathbb{H}_{n2}(\lambda, \alpha)$, we defined the following two-step estimator for $\theta = (\lambda, \alpha)$.

Definition 1. *The estimator* $\hat{\theta}_n^{(p)} = (\hat{\lambda}_n, \hat{\alpha}_n^{(p)})$ *is called generalized maximum composite likelihood estimator if*

$$\hat{\lambda}_n := \arg \sup_{\lambda \in \Theta_\lambda} \mathbb{H}_{n1}(\lambda), \tag{2}$$

$$\hat{\alpha}_n^{(p)} := \arg \sup_{\alpha \in \Theta_\alpha} \mathbb{H}_{n2}^{(p)}(\hat{\lambda}_n, \alpha). \tag{3}$$

In particular for stationary case, *i.e.*, the case when the kernel function satisfies the following condition;

$$K_\theta(x, y) = \lambda C_\alpha(x - y), \quad x, y \in \mathbb{R}^d,$$

we proved the consistency and the moment convergence of the estimators as the volume of the observation window $|D_n| \to \infty$ under some suitable conditions. Moreover, we presented the information criterion based on the proposed estimating method in this talk.

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

Junichi Hirukawa and Kou Fujimori
Niigata University and Waseda University

ABSTRACT

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

1 Introduction

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

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

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

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

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

1.1 The fundamental martingale M

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

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

where

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

and B is the beta funtion

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

Then, the centered Gaussian process

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

has independent increments and variance function

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

where

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

In particular, M is a martingale.

2 Weak convergence of $I(d)$ process

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

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

References

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

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

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

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

A closed form EM algorithm for a multivariate skew-normal model

Toshihiro Abe (Nanzan University)
Hironori Fujisawa (The Institute of Statistical Mathematics, RIKEN)
Takayuki Kawashima (Tokyo Institute of Technology)

The most famous skew distribution is a skew-normal distribution, which was proposed by Azzalini (1985, 1986). Various extensions to multivariate skew distributions have been proposed. Azzalini & Dalla Valle (1996) introduced a two-dimensional skew distribution whose marginal distribution is a univariate skew-normal distribution through additional parameters. Azzalini & Capitanio (1999) proposed a multivariate skew-normal distribution and investigated many properties of multivariate skew-normal distributions. Arellano-Valle & Genton (2005) presented some expressions of skew distribution and discussed basic classes of skew distribution, including skew-symmetric, skew-elliptical, and skew-spherical distributions. It is known that the maximum likelihood estimation for skew-normal distributions presents a problematic aspect in the neighborhood of base distribution.

In this talk, we consider the maximum likelihood estimation via an EM algorithm (Dempster et al., 1977) for a multivariate skew-normal distribution. Some stochastic representations were proposed to obtain a multivariate skew-normal distribution. A stochastic representation often enables us to make an EM algorithm in a convenient manner (McLachlan & Krishnan, 2007). Lin et al. (2007) constructed an EM algorithm for a mixture of univariate skew-normal distributions, but the EM algorithm obtained there demands a solution of complicated estimating equation for a skew parameter. The same problem happens even in a usual univariate skew-normal distribution, so that it is hard to extend the idea to a multivariate case. Lin (2009) considered an EM algorithm for a mixture of multivariate skew-normal distributions, although the multivariate skew-normal distribution treated there was based on Sahu et al. (2003) and a different type from Azzalini & Capitanio (1999). The EM algorithm demands multi-dimensional numerical integrations, which are hard to compute in a high-dimensional case. So far, there has no EM algorithm in a closed form.

Here we review the stochastic representation which implies the multivariate skew-normal distribution (Azzalini & Capitanio, 1999). Suppose

$$\begin{pmatrix} \boldsymbol{Y} \\ Y_0 \end{pmatrix} \sim N_{p+1}(\mathbf{0}, \Omega^*), \quad \Omega^* = \begin{pmatrix} \Omega & \boldsymbol{\delta} \\ \boldsymbol{\delta}^T & 1 \end{pmatrix}. \tag{1}$$

Then, $\boldsymbol{U} = \mathrm{sgn}(Y_0)\boldsymbol{Y}$ has the following density function:

$$f(\boldsymbol{u}) = 2\Phi(\boldsymbol{\alpha}^T \boldsymbol{u})\phi(\boldsymbol{u}; \mathbf{0}, \Omega), \tag{2}$$

where $\boldsymbol{\alpha} = \Omega^{-1}\boldsymbol{\delta}/(1 - \boldsymbol{\delta}^T\Omega^{-1}\boldsymbol{\delta})^{1/2}$, $\Phi(z)$ is the cumulative density function of the standard normal distribution and $\phi(\boldsymbol{z}; \mathbf{0}, \Omega)$ is the normal density function with mean vector $\mathbf{0}$ and covariance matrix Ω. The random variable $\boldsymbol{U}$ with the density (2) is said to have the multivariate skew-normal distribution (with mean zero). The distribution of $\boldsymbol{U} + \boldsymbol{\mu}$ is expressed as $\mathrm{SN}_p(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha})$.

In this talk, we construct an EM algorithm in a closed form, using the stochastic representation (1) with a slight modification. We change the covariance structure of $(\boldsymbol{Y}, Y_0)$ from Ω^* to

$$\Sigma = \begin{pmatrix} \Omega & \tau\Omega^{1/2}\boldsymbol{\delta} \\ \tau\boldsymbol{\delta}^T\Omega^{1/2} & \tau^2 \end{pmatrix}, \tag{3}$$

and then we consider the random variable $\boldsymbol{X} = \mathrm{sgn}(Y_0)\boldsymbol{Y}$. Here we give two devices. The first one is essential when we construct an EM algorithm. We add the parameter τ in (3). We can consider the covariance matrix Σ on the whole space of positive definite matrix, although Ω^* has the restriction $\tau = 1$ in (1). This device makes it easy to optimize the objective function in the M-step of the EM algorithm. The second one is a slightly different expression of parameter. We replace $\boldsymbol{\delta}$ in Ω^* by $\Omega^{1/2}\boldsymbol{\delta}$ in Σ, which changes the positive definite condition of the covariance matrix from $\boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta} < 1$ to the simple condition $\boldsymbol{\delta}^\top\boldsymbol{\delta} < 1$. We will show that $\boldsymbol{X} = \mathrm{sgn}(Y_0)\boldsymbol{Y}$ has a multivariate skew-normal distribution $\mathrm{SN}_p(\boldsymbol{0}, \Omega, \boldsymbol{\alpha})$ with $\boldsymbol{\alpha} = \Omega^{-1/2}\boldsymbol{\delta}/\sqrt{1-\boldsymbol{\delta}^\top\boldsymbol{\delta}}$, which is the same as the original multivariate skew-normal distribution if $\Omega^{1/2}\boldsymbol{\delta}$ is placed back to $\boldsymbol{\delta}$. A remarkable point is that the distribution of $\boldsymbol{X}$ does not depend on the parameter τ, since the joint distribution of $(\boldsymbol{Y}, Y_0)$ includes the parameter τ. In this sense, the parameter τ can be called the overparameter. The overparameter τ is not necessary when we consider a multivariate skew-normal distribution, but it enables us to construct an EM algorithm in a closed form. The R package `snem` has been developed to obtain the maximum likelihood estimate via the EM algorithm.

References

Arellano-Valle, R.B. & Genton, M.G. (2005). On fundamental skew distributions. Journal of Multivariate Analysis **96** (1), 93–116.

Azzalini, A. (1985). A class of distributions which includes the normal ones. Scandinavian Journal of Statistics **12** (2), 171–178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. Statistica **46** (2), 199–208.

Azzalini, A. & Capitanio, A. (1999). Statististial applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **61** (3), 579–602.

Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. Biometrika **83** (4), 715–726.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological) **39**, 1–38.

Lin, T.I. (2009). Maximum likelihood estimation for multivariate skew normal mixture models Journal of Multivariate Analysis **100** (2), 257–265.

Lin, T.I., Lee, J.C. & Yen, S.Y. (2007). Finite mixture modelling using the skew normal distribution. Statistica Sinica **17**, 909–927.

McLachlan, G. & Krishnan, T. (2007). The EM Algorithm and Extensions. John Wiley & Sons.

Sahu, S.K., Dey, D.K. & Branco, M.D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. Canadian Journal of Statistics **31** (2), 129–150.

Asymptotics of maximum likelihood estimation for stable law with continuous parameterization

Nanzan University Muneya Matsui

Abstract Asymptotics of maximum likelihood estimation for α-stable law are analytically investigated with a continuous parameterization. The consistency and asymptotic normality are shown on the interior of the whole parameter space. Although these asymptotics have been provided with Zolotarev's (B) parameterization, there are several gaps between. Especially in the latter, the density, so that scores and their derivatives are discontinuous at $\alpha = 1$ for $\beta \neq 0$ and usual asymptotics are impossible. This is considerable inconvenience for applications. By showing that these quantities are smooth in the continuous form, we fill gaps between and provide a convenient theory. We numerically approximate the Fisher information matrix around the Cauchy law $(\alpha, \beta) = (1, 0)$. The results exhibit continuity at $\alpha = 1$, $\beta \neq 0$ and this secures the accuracy of our calculations.

Main contents We use the following definitions and notations throughout. Denote ch.f. of stable law in continuous form by

$$\varphi(t) = \begin{cases} \exp\Big(-|\sigma t|^\alpha \big\{1 + i\beta \,\mathrm{sign} t \tan \frac{\pi\alpha}{2} (|\sigma t|^{1-\alpha} - 1)\big\} + i\mu t\Big) & \text{if } \alpha \neq 1 \\ \exp\Big(-|\sigma t| - i\sigma t \,(2\beta/\pi) \log |\sigma t| + i\mu t\Big) & \text{if } \alpha = 1, \end{cases}$$

where $\mu \in \mathbb{R}, \sigma \in \mathbb{R}_+, \alpha \in (0,2], \beta \in [-1,1]$ with $\mathbb{R}_+ = (0,\infty)$. We denote this parameter space by Θ_M and its interior by Θ_M°. A parameter vector is denoted by $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)' = (\mu, \sigma, \alpha, \beta)'$. As usual f', f'' mean the first and the second derivatives with respect to (w.r.t.) x and $f_\theta = (f_{\theta_1}, f_{\theta_2}, f_{\theta_3}, f_{\theta_4})'$ denotes a vector of partial derivatives of f w.r.t. θ. The second order partial derivatives w.r.t. x and θ are denoted by

$$f'_{\theta_i} = \frac{\partial^2 f}{\partial x \partial \theta_i} = \frac{\partial^2 f}{\partial \theta_i \partial x}, \quad f_{\theta_i\theta_j} = \frac{\partial^2 f}{\partial \theta_i \partial \theta_j} = \frac{\partial^2 f}{\partial \theta_j \partial \theta_i}, \quad i, j = 1, \ldots, 4,$$

i.e. all derivatives will be shown to be interchangeable in our case. Moreover, denote the log-likelihood function of f and its scores respectively by

$$\ell(x;\theta) = \log f(x;\theta) \quad \text{and} \quad \ell_{\theta_i}(x;\theta) = \frac{\partial \ell(x;\theta)}{\partial \theta_i} = \frac{f_{\theta_i}(x;\theta)}{f(x;\theta)}, \quad 1 \le i \le 4,$$

where for convenience we sometimes write $\ell(x)$ and $\ell_{\theta_i}(x)$ for these quantities. The second order derivatives of score functions w.r.t. θ and x, denoted by

$$\ell'_{\theta_i}(x) = \frac{\partial \ell_{\theta_i}(x;\theta)}{\partial x} = \frac{1}{f^2(x;\theta)} \big(f'_{\theta_i}(x;\theta) f(x;\theta) - f_{\theta_i}(x;\theta) f'(x;\theta)\big),$$

$$\ell_{\theta_i\theta_j}(x) = \frac{\partial \ell_{\theta_i}(x;\theta)}{\partial \theta_j} = \frac{1}{f^2(x;\theta)} \big(f_{\theta_i\theta_j}(x;\theta) f(x;\theta) - f_{\theta_i}(x;\theta) f_{\theta_j}(x;\theta)\big), \quad 1 \le i, j \le 4,$$

are also investigated. Again orders of partial derivatives w.r.t. (x, θ) are all exchangeable.

Based on the following technical Lemma 0.1 and Proposition 0.2, we establish asymptotics of maximum likelihood estimation as Theorem 0.3.

Lemma 0.1. *For every $x \in \mathbb{R}$, $f(x;\theta) : \theta \in \Theta_M^\circ$ is twice continuously differentiable w.r.t. θ, and f_{θ_i}, $i = 1, \ldots, 4$ is continuously differentiable w.r.t. x. Moreover f_{θ_i}, f'_{θ_i}, $f_{\theta_i\theta_j}$ $i, j = 1, \ldots, 4$ are jointly continuous in (x, θ) on $\mathbb{R} \times \Theta_M^\circ$.*

The tails of f and its derivatives for sufficiently large $|x|$ satisfy

$$\begin{array}{lll} f = O(|x|^{-(1+\alpha)}), & f'_\mu = -f_{\mu\mu} = O(|x|^{-(3+\alpha)}), & f_{\sigma\alpha} = O(|x|^{-(1+\alpha)} \log|x|), \\ f_\mu = -f' = O(|x|^{-(2+\alpha)}), & f'_\sigma = -f_{\mu\sigma} = O(|x|^{-(2+\alpha)}), & f_{\sigma\beta} = O(|x|^{-(1+\alpha)}), \\ f_\sigma = O(|x|^{-(1+\alpha)}), & f'_\alpha = -f_{\mu\alpha} = O(|x|^{-(2+\alpha)} \log|x|), & f_{\alpha\alpha} = O(|x|^{-(1+\alpha)} \log^2|x|), \\ f_\alpha = O(|x|^{-(1+\alpha)} \log|x|), & f'_\beta = -f_{\mu\beta} = O(|x|^{-(2+\alpha)}), & f_{\alpha\beta} = O(|x|^{-(1+\alpha)} \log|x|), \\ f_\beta = O(|x|^{-(1+\alpha)}), & f_{\sigma\sigma} = O(|x|^{-(1+\alpha)}), & f_{\beta\beta} = O(|x|^{-(1+\alpha)}). \end{array}$$

Furthermore for $\alpha = 1$, $\beta \in (-1, 1)$, we have

$$f_{\sigma\sigma} = O(|x|^{-3} \log|x|), \quad f_{\beta\beta} = O(|x|^{-3} \log|x|).$$

Proposition 0.2. *Let $\theta \in \Theta_M^\circ$. For every $x \in \mathbb{R}$,*

$$\ell_{\theta_i}(x), \ell'_{\theta_i}(x) \quad \text{and} \quad \ell_{\theta_i\theta_j}(x), \quad i, j = 1, \ldots, 4,$$

are well-defined and continuous in θ, and they are jointly continuous in (x, θ) on $\mathbb{R} \times \Theta_M^\circ$. Concerning tail behaviors, we have for sufficiently large $|x|$, $x \in \mathbb{R}$,

$$\begin{array}{ll} \ell_\mu(x) = O(|x|^{-1}), & \ell_{\mu\mu}(x) = -\ell'_\mu(x) = O(|x|^{-2}), \\ \ell_\sigma(x) = O(1), & \ell_{\mu\sigma}(x) = -\ell'_\sigma(x) = O(|x|^{-1}), \\ \ell_\alpha(x) = O(\log|x|), & \ell_{\mu\alpha}(x) = -\ell'_\alpha(x) = O(|x|^{-1} \log|x|), \\ \ell_\beta(x) = O(1), & \ell_{\mu\beta}(x) = -\ell'_\beta(x) = O(|x|^{-1}), \end{array}$$

and moreover,

$$\begin{array}{lll} \ell_{\sigma\sigma}(x) = O(1), & \ell_{\sigma\alpha}(x) = O(\log|x|), & \ell_{\sigma\beta}(x) = O(1), \\ \ell_{\alpha\alpha}(x) = O(\log^2|x|), & \ell_{\alpha\beta}(x) = O(\log|x|), & \ell_{\beta\beta}(x) = O(1). \end{array}$$

Theorem 0.3. *Let $\hat{\theta}_n$ be the maximum likelihood estimator based on i.i.d. n observations from stable law $(P_\theta : \theta \in \Theta_M)$. Assume that the true parameter θ_0 is in the interior $\theta_0 \in \Theta_M^\circ$ and prepare an arbitrary compact set $C \subset \Theta_M^\circ$ such that $\theta_0 \in C$. Then MLE $\hat{\theta}_n$ restricted on C is consistent and has asymptotic normality. In particular we have an expression*

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = I_{\theta_0}^{-1} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \ell_{\theta_0}(X_k) + o_{P_{\theta_0}}(1),$$

where $\sqrt{n}\,(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I_{\theta_0}^{-1})$ as $n \to \infty$, and I_{θ_0} is the Fisher information matrix.

Other remaining results including numerical works are given in the reference. By applying our present work, we are examining Quasi-maximum likelihood estimation for stable law.

References

[1] Matsui, M. (2019) Asymptotics of maximum likelihood estimation for stable law with continuous parameterization. arXiv:1901.09303.

Pitman Closeness Domination in Predictive Density Estimation for Two Ordered Normal Means Under α-Divergence Loss

Yuan-Tsung Chang (Mejiro University) Nobuo Shinozaki (Keio University)
William, E. Strawderman (Rutgers University)

We consider Pitman closeness domination in predictive density estimation problems when the underlying loss metric is α-divergence $\{D(\alpha)\}$, a loss introduced by Csiszàr (1967) given by

$$D_\alpha\{\hat{p}(\tilde{y}|y), p(\tilde{y}|\psi)\} = \int f_\alpha\left(\frac{\hat{p}(\tilde{y}|y)}{p(\tilde{y}|\psi)}\right) p(\tilde{y}|\psi) d\tilde{y}, \tag{1}$$

where, for $-1 \le \alpha \le 1$

$$f_\alpha(z) = \begin{cases} \frac{4}{1-\alpha^2}(1 - z^{(1+\alpha)/2}), & |\alpha| < 1 \\ z \log z, & \alpha = 1 \\ -\log z, & \alpha = -1. \end{cases} \tag{2}$$

Here KL loss corresponds to $\alpha = -1$. The case $\alpha = 1$ is sometimes referred to as reverse KL loss.
We consider the normal distributions case and generalized Bayesian predictive densities case.

1) **Case of normal distributions**
If the true density function of Y is $N(\mu, \sigma^2)$ and the estimated predictive density of Y is $N(\hat{\mu}, \hat{\sigma}^2)$, Chang and Strawderman (2014) have derived the general form of D_α loss and have shown that it is a concave monotone function of quadratic loss and is also a function of the variances (observed, predicand, and plug-in). The general form is given as following:
a) for $-1 < \alpha < 1$,

$$D_\alpha(N(\tilde{y}|\hat{\mu}, \hat{\sigma}^2), N(\tilde{y}|\mu, \sigma^2)) = \frac{4}{1-\alpha^2}\left(1 - d(\sigma^2, \hat{\sigma}^2) e^{-A(\sigma^2,\hat{\sigma}^2)\frac{(\hat{\mu}-\mu)^2}{2}}\right), \tag{3}$$

where

$$d(\sigma^2, \hat{\sigma}^2) = \frac{\sigma^{(\alpha-1)/2}\tau}{\hat{\sigma}^{(\alpha+1)/2}}, \quad A(\sigma^2, \hat{\sigma}^2) = \left(\frac{1-\alpha}{2\sigma^2}\right)\left(1 - \frac{(1-\alpha)\tau^2}{2\sigma^2}\right) > 0, \quad \frac{1}{\tau^2} = \left(\frac{1+\alpha}{2\hat{\sigma}^2} + \frac{1-\alpha}{2\sigma^2}\right).$$

Further, $d(\sigma^2, \hat{\sigma}^2) < 1$ and $A(\sigma^2, \hat{\sigma}^2) > 0$.

b) for $\alpha = +1$, $D_{+1}(N(\tilde{y}|\hat{\mu}, \hat{\sigma}^2), N(\tilde{y}|\mu, \sigma^2)) = \frac{1}{2}\left[\left(\frac{\hat{\sigma}^2}{\sigma^2} - \log\frac{\hat{\sigma}^2}{\sigma^2} - 1\right) + \frac{(\hat{\mu}-\mu)^2}{\sigma^2}\right]$. (4)

c) for $\alpha = -1$, $D_{-1}(N(\tilde{y}|\hat{\mu}, \hat{\sigma}^2), N(\tilde{y}|\mu, \sigma^2)) = \frac{1}{2}\left[\left(\frac{\sigma^2}{\hat{\sigma}^2} - \log\frac{\sigma^2}{\hat{\sigma}^2} - 1\right) + \frac{(\hat{\mu}-\mu)^2}{\hat{\sigma}^2}\right]$. (5)

The underlying distributions considered are normal, including the distribution of the observables, the distribution of the variable whose density is to be predicted, and the estimated predictive density which will be taken to be of the plug-in type. We demonstrate $\{D(\alpha)\}$ Pitman closeness domination of certain plug-in predictive densities over others for the entire class of metrics simultaneously when related Pitman's closeness domination holds in the problem of estimating the mean.

Examples of Pitman closeness domination presented relate to the problem of estimating the predictive density of the variable with the larger mean. More precisely, let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ be two independent random normal variables, where $\mu_1 \le \mu_2$. Under the above restriction we wish to predict a normal population with mean equal to the larger mean, μ_2, and variance equal to σ^2, $\tilde{Y} \sim N(\mu_2, \sigma^2)$. We consider different versions of this problem, depending on whether the $\sigma_i^2, i = 1, 2$ are known or are unknown but satisfy the additional order restriction, $\sigma_1^2 \le \sigma_2^2$. The case of two ordered normal means with known covariance matrix is also considered.

We also consider $\{D(\alpha)\}$ Pitman domination of certain generalized Bayesian (best invariant) procedures suggested by Corcuera and Giummole (1999) in next section.

2) **Case of generalized Bayesian predictive densities**
In this section we discuss improving the generalized Bayesian predictive densities suggested by Corcuera and Giummole (1999) under $D(\alpha)$ loss.

Based on the data

$$X_{ij} \sim N(\mu_i, \sigma_i^2), i = 1, 2, \ j = 1, \cdots, n_i,$$

we predict the density $\tilde{Y} \sim N(\mu_i, \sigma_i^2), i = 1, 2$. We denote its density function by $p(\tilde{y}; \mu_i, \sigma_i)$, where μ_i and σ_i^2 are unknown.

When $-1 \leq \alpha < 1$, Corcuera and Giummole (1999) have established that the best invariant predictive density of $p(\tilde{y}; \mu_i, \sigma_i)$ based solely on $x_{i1}, \cdots x_{in_2}$ is

$$\hat{p}_\alpha(\tilde{y}; \bar{x}_i, \tilde{\sigma}_i) \propto \left[1 + \frac{1-\alpha}{2n_i + 1 - \alpha}\left(\frac{y - \bar{x}_i}{\tilde{\sigma}_i}\right)^2\right]^{-(2n_i - 1 - \alpha)/2(1-\alpha)}, \tag{6}$$

where $\bar{x}_i$ is the sample mean and $\tilde{\sigma}_i^2 = ((n_i - 1)/n_i)s_i^2$ is the sample variance. Corcuera and Giummole (1999) have also shown that $\hat{p}_\alpha(\tilde{y}; \bar{x}_i, \tilde{\sigma}_i)$ is the generalized Bayesian predictive density for the prior density $f(\mu_i, \sigma_i) \propto 1/\sigma_i, 0 < \sigma_i < \infty$. It is to be noted that $\hat{p}_\alpha(\tilde{y}; \bar{x}_i, \tilde{\sigma}_i)$ is not a normal distribution, although the plug-in density $N(\bar{x}_i, s_i^2)$ is the generalized Bayes rule when $\alpha = 1$.

We consider the following two cases separately where order restrictions on μ_i and/or σ_i^2 are present, *i)* when $\mu_1 \leq \mu_2$ case and *ii)* when $\mu_1 \leq \mu_2$ and $\sigma_1^2 \leq \sigma_2^2$.

We consider to improve $\hat{p}_\alpha(\tilde{y}; \bar{x}_i, \tilde{\sigma}_i)$ or $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i^{OS}, \tilde{\sigma}_i)$ by replacing $\bar{x}_i$ with $\hat{\mu}_i^{OS}$ or $\hat{\mu}_i^{OS}$ with $\hat{\mu}_i^{CS}$, respectively, where

$$\hat{\mu}_1^{OS} = \min\left\{\bar{X}_1, \hat{\mu}^{GD} \equiv \frac{n_1 s_2^2}{n_1 s_2^2 + n_2 s_1^2}\bar{X}_1 + \frac{n_2 s_1^2}{n_1 s_2^2 + n_2 s_1^2}\bar{X}_2\right\}, \quad \hat{\mu}_2^{OS} = \max\{\bar{X}_2, \hat{\mu}^{GD}\},$$

$$\hat{\mu}_1^{CS} = \begin{cases} \hat{\mu}_1^{OS}, & if\ s_1^2 \leq s_2^2 \\ \min\left\{\bar{X}_1, \frac{n_1}{n_1+n_2}\bar{X}_1 + \frac{n_2}{n_1+n_2}\bar{X}_2\right\}, & if\ s_1^2 > s_2^2 \end{cases}$$

$$\hat{\mu}_2^{CS} = \begin{cases} \hat{\mu}_2^{OS}, & if\ s_1^2 \leq s_2^2 \\ \max\left\{\bar{X}_2, \frac{n_1}{n_1+n_2}\bar{X}_1 + \frac{n_2}{n_1+n_2}\bar{X}_2\right\}, & if\ s_1^2 > s_2^2. \end{cases}$$

The next lemma is usefully for improving the generalized Bayesian predictive densities (6).

Lemma Let $f(\cdot)$ be the probability density function of $X \sim N(0, \tau^2)$. Assume that $g(t) \geq 0$ is symmetric about the origin and is a strictly decreasing function of $|t|$ such that $\int_{-\infty}^{\infty} g(x)f(x)dx < \infty$. Then $\int_{-\infty}^{\infty} g(y-x)f(y-\mu)dy$ is a strictly decreasing function of $|x - \mu|$.

Let $\hat{\mu}_i$ denote an estimator of $\mu_i, i = 1, 2$ in general. Now we show that for any $1 \leq \alpha < 1$, $D_\alpha(\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i))$ is a strictly increasing function of $|\hat{\mu}_i - \mu_i|$.

From Lemma, we see that for $|\alpha| < 1$,

$$D_\alpha(\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i)) \propto 1 - \int_{\infty}^{\infty} g(\tilde{y} - \hat{\mu}_i)f(\tilde{y} - \mu_i)d\tilde{y}$$

is a strictly increasing function of $|\hat{\mu}_i - \mu_i|$, where

$$g(y - x) = \left[1 + \frac{1-\alpha}{2n_i + 1 - \alpha}\left(\frac{y - x}{\hat{\sigma}_i}\right)^2\right]^{-(2n_i - 1 - \alpha)(1+\alpha)/4(1-\alpha)}$$

and

$$f(y - \mu) \propto \exp\left\{-\frac{(1-\alpha)(y-\mu)^2}{4\sigma^2}\right\}.$$

For $\alpha = -1$,

$$D_{-1}(\hat{p}_{-1}(\tilde{y}; \hat{\mu}_i, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i)) = -E_{\tilde{y}}\left\{\log\left[\frac{\hat{p}_{-1}(\tilde{y}; \hat{\mu}_i, \hat{\sigma}_i)}{p(\tilde{y}; \mu_i, \sigma_i)}\right]\right\}$$

is a strictly increasing function of $|\hat{\mu}_i - \mu_i|$ from Lemma.

i) $\mu_1 \leq \mu_2$ **case**

Theorem 1. The predictive density estimate $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i^{OS}, \hat{\sigma}_i), i = 1, 2$ is closer to the predictive density $p(\tilde{y}; \mu_i, \sigma_i)$ than $\hat{p}_\alpha(\tilde{y}; \bar{x}_i, \hat{\sigma}_i)$, respectively, under the $\{D(\alpha)\}$ metric for all $-1 \leq \alpha < 1$ and for every estimator $\hat{\sigma}_i$ if and only if $\hat{\mu}^{GD}$ is Pitman closer to μ than $\bar{X}_i$ for all σ_1^2 and σ_2^2 when $\mu_1 = \mu_2 = \mu$.

ii) $\mu_1 \leq \mu_2$ **and** $\sigma_1^2 \leq \sigma_2^2$ **case**

Theorem 2. The predictive density estimate $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_2^{CS}, \hat{\sigma}_2)$ is closer to the predictive density $p(\tilde{y}; \mu_2, \sigma_2)$ than $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_2^{OS}, \hat{\sigma}_2)$ under the $\{D(\alpha)\}$ metric for all $-1 \leq \alpha < 1$ and for every estimator $\hat{\sigma}_2^2$.

Theorem 3. The predictive density estimate $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_1^{CS}, \hat{\sigma}_1)$ is not Pitman closer to $p(\tilde{y}; \mu_1, s_1)$ than $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_1^{OS}, \hat{\sigma}_1)$ when $\mu_2 - \mu_1$ is sufficiently large, under the $\{D(\alpha)\}$ metric for all $-1 \leq \alpha < 1$ and for any estimator $\hat{\sigma}_1^2$.

Auto-Distance Covariance Function for Time Series Analysis

Konstantinos Fokianos
k.fokianos@lancaster.ac.uk
Department of Mathematics & Statistics, Lancaster University

There has been a considerable recent interest in measuring dependence by employing the concept of distance covariance function, a new measure of dependence for random variables, introduced by Székely et al. (2007). This tool has been recently defined to the context of multivariate time series by Zhou (2012), but without exploring the interrelationships between the various time series components. In this work, we extend the notion of distance covariance to multivariate time series by defining its matrix version.

We denote by $\{\mathbf{X}_t : t = 0, \pm 1, \pm 2, \dots\}$ a d-dimensional time series process, with components $X_{t;r}$, $r = 1, \dots, d$. Suppose we have available a sample of size n, that is $\{\mathbf{X}_t, t = 1, \dots, n\}$. We define the pairwise auto-distance covariance function as a function of the joint and marginal characteristic functions of the pair $(X_{t;r}, X_{t+j;m})$, for $r, m = 1, \dots, d$. Denote by $\phi_j^{(r,m)}(u, v)$ the joint characteristic function of $X_{t;r}$ and $X_{t+j;m}$; that is

$$\phi_j^{(r,m)}(u, v) = E\left[\exp\left(i(uX_{t;r} + vX_{t+j;m})\right)\right], \quad j \in \mathbb{Z},$$

and the marginal characteristic functions of $X_{t;r}$ and $X_{t+j;m}$ as $\phi^{(r)}(u) := \phi_j^{(r,m)}(u, 0)$ and $\phi^{(m)}(v) := \phi_j^{(r,m)}(0, v)$ respectively, where $(u, v) \in \mathbb{R}^2$, and $i^2 = -1$. The *pairwise auto-distance covariance function (ADCV)* between $X_{t;r}$ and $X_{t+j;m}$, $V_{rm}(j)$, is defined as the positive square root of

$$V_{rm}^2(j) = \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{\left|\phi_j^{(r,m)}(u, v) - \phi^{(r)}(u)\phi^{(m)}(v)\right|^2}{|u|^2 \, |v|^2} du dv, \quad j \in \mathbb{Z}.$$

The auto-distance covariance matrix, $V(j)$, is then defined by

$$V(j) = \left[V_{rm}(j)\right]_{r,m=1}^d, \quad j \in \mathbb{Z}.$$

The *pairwise auto-distance correlation function (ADCF)* between $X_{t;r}$ and $X_{t+j;m}$, $R_{rm}(j)$, is a coefficient that lies in the interval $[0, 1]$ and also measures dependence and is defined as the positive square root of

$$R_{rm}^2(j) = \frac{V_{rm}^2(j)}{\sqrt{V_{rr}^2(0)}\sqrt{V_{mm}^2(0)}},$$

for $V_{rr}(0)V_{mm}(0) \neq 0$ and zero otherwise. The auto-distance correlation matrix of $\mathbf{X}_t$, is then defined as

$$R(j) = \left[R_{rm}(j)\right]_{r,m=1}^d, \quad j \in \mathbb{Z}.$$

When $j \neq 0$, $V_{rm}(j)$ measures the dependence of $X_{t;r}$ on $X_{t+j;m}$. In general, $V_{rm}(j) \neq V_{mr}(j)$ for $r \neq m$, since they measure different dependence structure between the series $\{X_{t;r}\}$ and $\{X_{t;m}\}$ for all $r, m = 1, 2, \dots, d$.

Thus, $V(j)$ and $R(j)$ are non-symmetric matrices, but $V(-j) = V'(j)$ and $R(-j) = R'(j)$. More properties can be found in Fokianos and Pitsillou (2018). The empirical pairwise ADCV, $\widehat{V}_{rm}(j)$, for $j \geq 0$, is the non-negative square root of

$$\widehat{V}^2_{rm}(j) = \frac{1}{(n-j)^2} \sum_{t,s=1}^{n-j} A^r_{ts} B^m_{ts},$$

where $A^r = A_{ts}$ and $B^m = B_{ts}$ are Euclidean distance matrices given by

$$A^r_{ts} = a^r_{ts} - \bar{a}^r_{t.} - \bar{a}^r_{.s} + \bar{a}^r_{..},$$

with $a^r_{ts} = |X_{t;r} - X_{s;r}|$, $\bar{a}^r_{t.} = \left(\sum_{s=1}^{n-j} a^r_{ts}\right)/(n-j)$, $\bar{a}^r_{.s} = \left(\sum_{t=1}^{n-j} a^r_{ts}\right)/(n-j)$, $\bar{a}^r_{..} = \left(\sum_{t,s=1}^{n-j} a^r_{ts}\right)/(n-j)^2$. B^m_{ts} is defined analogously in terms of $b^m_{ts} = |X_{t+j;m} - X_{s+j;m}|$. Fokianos and Pitsillou (2018) have shown that for a d-dimensional strictly stationary and ergodic process $\{\mathbf{X}_t\}$ with $E\,|X_{t;r}|^2 < \infty$, for $r = 1, \ldots, d$, then for all $j \in \mathbb{Z}$,

$$\widehat{V}(j) \to V(j),$$

almost surely, as $n \to \infty$. In addition, under pairwise independence it holds that

$$n\widehat{V}^2_{rm}(j) \to Z := \sum_k \lambda_k Z^2_k,$$

in distribution, as $n \to \infty$, where $\{Z_k\}$ is an i.i.d sequence of $N(0,1)$ random variables, and (λ_k) is a sequence of nonzero eigenvalues.

Based on these results, we will be considering the following two problems:

- Testing the iid hypothesis by employing the generalized spectral density approach as developed by (Hong, 1999)

- Developing consistent testing procedures for the detection of multiple change-points in a given stream of data.

References

Fokianos, K. and M. Pitsillou (2018). Testing independence for multivariate time series via the auto-distance correlation matrix. *Biometrika 105*, 337–352.

Hong, Y. (1999). Hypothesis testing in time series via the emprical characteristic function: A generalized spectral density approach. *J. Am. Stat. Assoc.* **94**, 1201–1220.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007). Measuring and testing dependence by correlation of distances. *Ann. Stat.* **35**, 2769–2794.

Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time. Ser. Anal.* **33**, 438–457.

報告書

入江　薫
（東京大学経済学部・講師）

科研費シンポジウム「多様な分野における統計科学に関する諸問題」において、三日目にあたる九月十六日の午前のセッションにて、「縮小事前分布と状態空間モデル」という演題で研究成果を発表した。本発表は Irie, K., “Bayesian Dynamic Fused LASSO”, arXiv preprint をもとにしており、当日もこのタイトルを表示した。
発表は予稿の内容に沿って行い、上記の公表済みの研究について発表した。講演時間が十分にあることを踏まえ、背景となる先行研究を詳細にレビューしたうえで、統計的計算手法の導出（の概略）も含めて当研究を解説した。縮小事前分布の研究は時系列データのベイジアンモデリングの分野においてもさかんに行われているものの、先行研究のほとんどが動的パラメータの経時変化に対する縮小のみを考慮しており、通常の意味での縮小効果、すなわち動的回帰係数パラメータをゼロに近づけるという意味での縮小効果をモデルに含めていない、という点を指摘した。一方で、複数の罰則項を用いて異なる縮小効果を同時に実現する Fused LASSO の方法は上記の問題を解決するように思えるが、動的パラメータの時間発展を記述するものではないため、予測分析に使用することができず実用上は問題の多いアプローチであることを説明した。このような状況で本研究は、Fused LASSO のモデリングを用いたマルコフ過程を動的パラメータの事前分布として使用することを提案する。著しい特徴として、そのような事前分布を用いた状態空間モデルは潜在パラメータを導入することで条件付き線形ガウス型の状態空間モデルとなり、事後分布の計算が簡易なギブス・サンプラーで実行可能となるため、計算の段階においては効率性・スケーラビリティ・実行の簡便さにおいて多くの利点を持つ。以上のようなモデルの特性はシミュレーションデータを用いた分析で例証され、本講演では多数の図表を示しながら分析結果を紹介した。
当日の質疑応答では、シンポジウムの主旨にのっとり、本研究の分野であるベイジアン・モデリングの最近の研究課題にまでさかのぼって、非専門家の方に向けて解説を行う機会が得られた。非ベイズ的な l1 罰則項が二重指数事前分布に対応することはよく知られているが、事前分布のモデリングを発展されることでよりスパースな状況を想定する分布を構成する研究がなされており、特に Carvalho, Polson and Scott (2009, 2010)以降さかんに研究・利用されている馬蹄(horseshoe)事前分布の解説を行った。聴衆から多くの質問があって、ベイズ統計学においてはスタンダードになりつつあるこの事前分布のクラスについて詳細な解説を行うことができ、本シンポジウムの主旨に多少なりとも貢献することができたと考える。
また、当日は縮小推定に詳しい/専門とする先生が多く出席しており、研究の核となるモデ

リングの点について有意義な指摘・提案をしていただいた。たとえば、対立する二種類の縮小効果を調整する重みパラメータは時間に関して一定の定数であるが、この設定が動的なスパース性に対しては制約的であり、重みパラメータの動的なモデリングが望ましいとの指摘があった。この点については従来から認識していたが、もっぱら技術的な問題から無視していたため、今回の指摘は問題の重要さを実感する意味で非常に有意義であった。また、モデリングに関する様々な提案がなされ、階層化、grouped LASSO、二階差分に対する fused LASSO など、魅力的な今後の研究課題として取り組ませていただくことになった。

発表した研究は現在査読を受けている。今回得られたコメントと同様の指摘を査読者から受けることも十分考えられる。本シンポジウムでのやりとりを活かして論文を改善し、早くに出版することを目指したい。

バランス型経時測定データにおける Extended GMANOVA モデルの解釈と新たな推定法

中京大学 国際教養学部 永井 勇

各個体に対して経時的に測定することで得られるデータは経時測定データと呼ばれ, 様々な分野で収集・分析がされている. この経時測定データの分析においては, データに潜む経時的な変動 (経時変動) を上手く捉えることが一つの目的である. 本講演では, 全ての個体で測定時点が揃っている, つまり, n 個全ての個体に対して時点 $t_1, \ldots, t_p$ で測定した経時測定データを考えた. このような経時測定データはバランス型経時測定データと呼ばれる.

本講演では, このバランス型経時測定データにおける経時変動を $t_1, \ldots, t_p$ の多項式などで捉えることを考えた場合, どのようにモデル化されるかを報告し, このモデルが Potthoff and Roy (1964) により提案された次の GMANOVA (一般化多変量分散分析) モデルと対応している事をまず報告した;

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

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, $\boldsymbol{1}_n$ は全ての成分が 1 の n 次元ベクトル, $\boldsymbol{m}$ は q 次元未知ベクトル, $\boldsymbol{X}$ は $p \times q$ 既知行列 (詳しくは後述), $\boldsymbol{A}$ は各行が各個体の性別などからなる $n \times k$ 既知 (説明変数) 行列, Ξ は $p \times q$ 未知行列, $\boldsymbol{\mathcal{E}}$ は $n \times p$ 誤差行列である. モデル化の話より, $(\boldsymbol{1}_n \boldsymbol{m}' \boldsymbol{X}' + \boldsymbol{A} \Xi \boldsymbol{X}')$ が $\boldsymbol{Y}$ の経時変動を表しており, 例えば $\boldsymbol{X}$ の j 行目を $(1, t_j, \ldots, t_j^{q-1})$ とすることが, 経時変動を測定時点 $t_1, \ldots, t_p$ の $(q-1)$ 次多項式で推定することに対応している.

ただし, このモデルには, 全ての説明変数に対して**同一の次数の多項式**など**しか**用いることができないという問題がある. 一方で, $\boldsymbol{1}_n \boldsymbol{m}' \boldsymbol{X}'$ の部分を除いた形のモデルに対しては, Kollo and von Rosen (2005) の Def. 4.1.3 で拡張が考えられている. そこで本講演では, 上記の GMANOVA モデルに Kolloe ら (2005) の拡張したモデルを用いた次のモデル (拡張 GMANOVA モデル) を提案した;

$$\boldsymbol{Y} = \boldsymbol{1}_n \boldsymbol{\mu}' \boldsymbol{X}_0' + \sum_{i=1}^{r} \boldsymbol{A}_i \Xi_i \boldsymbol{X}_i' + \boldsymbol{\mathcal{E}},$$

ここで, $\boldsymbol{\mu}$ は q_0 次元未知ベクトル, $\boldsymbol{X}_i$ は $p \times q_i$ 既知行列 (詳しくは後述; $i = 0, 1, \ldots, r$), $\boldsymbol{A}_i$ は各行が各個体のいくつかの説明変数のグループからなる $n \times k_i$ 既知 (説明変数) 行列 $(i = 1, \ldots, r)$, Ξ_i は $p \times q_i$ 未知行列 $(i = 1, \ldots, r)$ である (他の記号は GMANOVA モデルで用いている記号と同様の意味である). ここで本講演では, $n - \sum_{i=1}^{r} k_i - 1 > 0$ で各 $\boldsymbol{A}_i$ が中心化 (つまり, $\boldsymbol{1}_n' \boldsymbol{A}_i = \boldsymbol{0}_{k_i}$, $\boldsymbol{0}_{k_i}$ は k_i 次元ゼロベクトル) とし, $\mathrm{rank}(\boldsymbol{X}_i) = q_i$, $\mathrm{rank}(\boldsymbol{A}_i) = k_i$ とした. さらに, $\boldsymbol{\mathcal{E}}$ の各行は独立で $E[\boldsymbol{\mathcal{E}}] = \boldsymbol{0}_n \boldsymbol{0}_p'$, $\mathrm{Cov}[\mathrm{vec}(\boldsymbol{\mathcal{E}})] = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$, $\boldsymbol{\Sigma}$ は $\mathrm{rank}(\boldsymbol{\Sigma}) = p$ の $p \times p$ 未知行列とした.

本講演では, GMANOVA モデルと経時変動の対応を踏まえ, 経時測定データに対する分析において提案したモデルの解釈について, 以下のように報告した. GMANOVA モデルと同様に, このモデルにおいても $\left(\boldsymbol{1}_n \boldsymbol{\mu}' \boldsymbol{X}_0' + \sum_{i=1}^{r} \boldsymbol{A}_i \Xi_i \boldsymbol{X}_i' \right)$ が $\boldsymbol{Y}$ の経時変動を表している. さらに, 平均的な経時変動を表す $\boldsymbol{1}_n \boldsymbol{\mu}' \boldsymbol{X}_0'$ と各説明変数のグループ $\boldsymbol{A}_i$ に対する経時変動を表す $\boldsymbol{A}_i \Xi_i \boldsymbol{X}_i'$ $(i = 1, \ldots, r)$ と分割して見ると, それぞれが異なる $\boldsymbol{X}_i$ $(i = 0, \ldots, r)$ を用いていることが分かる. ここで, GMANOVA モデルで

の経時変動の推定と同様に考えると, 例えば $\boldsymbol{X}_i$ の j 行目を $(1, t_j, \ldots, t_j^{q_i-1})$ とすることが, 経時変動を測定時点 $t_1, \ldots, t_p$ の $(q_i - 1)$ 次多項式で推定することに対応していることを報告した. これらをまとめると, 平均部分や**各説明変数のグループごと**に **異なる次数の多項式**など を用いた経時変動の推定ができることが分かる. これにより, 説明変数のあるグループについての知見などを生かした経時変動が推定できる.

しかし, 従来の Kollo ら (2005) で提案されているモデルでの推定の際には, $\boldsymbol{X}_i$ や $\boldsymbol{A}_i$ について非常に厳しい仮定を置き, $\boldsymbol{\mathcal{E}}$ の各行が従う確率分布が正規分布と仮定し, さらに反復計算も必要となる形となっている. 経時測定データの分析を考えた際には, 特に $\boldsymbol{X}_i$ や $\boldsymbol{A}_i$ に対する仮定をそのまま用いる事は, 説明変数に関する知見を生かした経時変動の推定などができなくなり, さらに用いる多項式にも非常に厳しい制約を課すこととなる. そこで本講演では, これらの仮定を置かずに以下の残差平方和 (RSS) を最小にすることで得られる最小二乗推定量 (LSE) がどのような形になるかを報告した;

$$\begin{aligned}&\mathrm{RSS}(\boldsymbol{\mu}, \Xi_1, \ldots, \Xi_r|\boldsymbol{\Sigma})\\ &\overset{\text{def.}}{=} \mathrm{tr}\left\{\left(\boldsymbol{Y} - \boldsymbol{1}_n\boldsymbol{\mu}'\boldsymbol{X}_0' - \sum_{i=1}^{r}\boldsymbol{A}_i\Xi_i\boldsymbol{X}_i'\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \boldsymbol{1}_n\boldsymbol{\mu}'\boldsymbol{X}_0' - \sum_{i=1}^{r}\boldsymbol{A}_i\Xi_i\boldsymbol{X}_i'\right)'\right\}.\end{aligned}$$

つまり, $\boldsymbol{\mu}$ や Ξ_ℓ の LSE $\hat{\boldsymbol{\mu}}$, $\hat{\Xi}_\ell$ が次のように得られることを報告した;

$$\begin{cases}\hat{\boldsymbol{\mu}} = (\boldsymbol{X}_0'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'\boldsymbol{1}_n/n,\\ \hat{\Xi}_\ell = (\boldsymbol{A}_\ell'\boldsymbol{A}_\ell)^{-1}\boldsymbol{A}_\ell'\left(\boldsymbol{Y} - \sum_{i\neq\ell}\boldsymbol{A}_i\hat{\Xi}_i\boldsymbol{X}_i'\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_\ell(\boldsymbol{X}_\ell'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_\ell)^{-1}\ (\ell = 1, \ldots, r).\end{cases}$$

しかしながら, $\hat{\Xi}_\ell$ においては, 他の $\hat{\Xi}_i$ $(i \neq \ell)$ が必要なため, $\hat{\Xi}_i$ 全ての初期値や更新のための反復計算が必要である. そこで本講演では, Lütkepohl (1996) などにある vec 作用素と $\otimes$ (クロネッカー積) の関係, 直行射影行列の性質などを用いて, 次のように反復計算が不要な形での推定量を報告した;

$$\mathrm{vec}(\hat{\Xi}_\ell) = (\boldsymbol{Q}_\ell'\boldsymbol{Q}_\ell)^{-1}\boldsymbol{Q}_\ell'(\boldsymbol{\Sigma}^{-1/2}\otimes\boldsymbol{I}_n)\mathrm{vec}\left(\boldsymbol{Y} - \sum_{i>\ell}^{r}\boldsymbol{A}_i\hat{\Xi}_i\boldsymbol{X}_i'\right)\ (\ell = 1, \ldots, r),$$

ここで, $\sum_{i>r}^{r}\boldsymbol{A}_i\hat{\Xi}_i\boldsymbol{X}_i' = \boldsymbol{O}$ (適当な大きさのゼロ行列) としている. さらに $\boldsymbol{Q}_1 = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}_1 \otimes \boldsymbol{A}_1, \boldsymbol{Q}_j = \left(\boldsymbol{I}_{np} - \sum_{i=1}^{j-1}\boldsymbol{P}_{\boldsymbol{Q}_i}\right)(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}_j \otimes \boldsymbol{A}_j)$ $(j > 2)$, $\boldsymbol{P}_{\boldsymbol{Q}_i} = \boldsymbol{Q}_i(\boldsymbol{Q}_i'\boldsymbol{Q}_i)^{-1}\boldsymbol{Q}_i$ である (ここで, $(\boldsymbol{Q}_i'\boldsymbol{Q}_i)^{-1}$ の存在は仮定する). この推定量では, まず $\mathrm{vec}(\hat{\Xi}_r)$ が得られ, これを用いる事で $\mathrm{vec}(\hat{\Xi}_{r-1})$ が得られる. これを繰り返す事で, r 個全ての推定量が反復計算をせずに得られる. この推定量を用いる事で, Kollo ら (2005) のモデルでの問題点 ($\boldsymbol{X}_i$ や $\boldsymbol{A}_i$ への厳しい仮定, $\boldsymbol{\mathcal{E}}$ の各行への正規性の仮定, $\hat{\Xi}_i$ を得るために反復計算が必要) を全て避けた新たな推定量が得られた.

さらに本講演では, 推定量の不偏性や $\boldsymbol{\Sigma}$ の不偏推定量についても報告した.

引用文献:

[1] Kollo, T. & von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*, Springer.
[2] Lütkepohl, H. (1996). *Handbook of Matrices*, John Wiley & Sons.
[3] Potthoff, R. F. & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.

Generalized Multinomial Distributions with Scalable Variance

Nobuaki Hoshino[*]

September 2019

Abstract

In this talk we generalize the multinomial distribution to keep closure under the collapse of cells, motivated by a practice of statistical disclosure control. The resulting family of distributions is characterized by Bell polynomials, and its parameter spaces are derived. This family is shown to have the same mariginal first moment as that of the multinomial distribution, but its second moment can be overdispersed. Also shown is that an infinite dimensionl distribution of this family exists by Kolmogorov's extension theorem. In this limit the marginal distribution of positive frequencies is reduced to a simple random partitioning distribution of a positve integer, which is useful to deal with a sparse contingency table.

Keywords: Discrete multivariate distribution, Overdispersion, Sparsity

[*]School of Economics, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan. E-mail: hoshino@kenroku.kanazawa-u.ac.jp