

令和元年度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)) によるシンポジウム

統計的推測および確率解析に関する総合的研究

開催責任者: 宇野力 (秋田大学教育文化学部)

場所: カレッジプラザ 講堂

住所: 秋田市中通 2 丁目 1-51 明德館ビル 2 階 (JR 秋田駅近く)

日時: 令和元年 12 月 11 日 (水) 14:00 ~ 12 月 13 日 (金) 12:20

プログラム

12月11日 (水)

14:00~14:10 開会

座長 種市信裕 (北海道教育大学札幌校)

14:10~14:50 「Non-stationary Time Series Classification using Topological Data Analysis」

黒木裕鷹 (東京理科大学大学院工学研究科), 塩濱敬之 (東京理科大学工学部)

14:50~15:30 「Maximum Likelihood Estimation for Hidden Markov Models with Cylindrical Distributions」

小坪琢人 (東京理科大学大学院工学研究科), 塩濱敬之 (東京理科大学工学部)

座長 塩濱敬之 (東京理科大学工学部)

15:45~16:25 「Jensen's alpha in the CAPM market model from the viewpoint of asymptotic accuracy of estimators of β and of asymmetry of distributions of residuals」

三浦良造 (東北大学大学院経済学研究科, 一橋大学)

16:25~17:05 「正方分割表における部分対称性に関する種々のモデルとその尺度」

桃崎智隆 (東京理科大学理工学研究科), 福本伸樹 (東京理科大学理工学研究科),

中川智之 (東京理科大学理工学部), 石井晶 (東京理科大学理工学部),

三枝祐輔 (横浜市立大学医学部), 富澤貞男 (東京理科大学理工学部)

17:05~17:45 「多次元分割表における多要因間独立性と多次元完全独立性の関係」

種市信裕 (北海道教育大学札幌校), 関谷祐里 (北海道教育大学釧路校),

外山淳 (数学利用研究所)

12月12日 (木)

座長 柿沢佳秀 (北海道大学経済学研究院)

10:15~10:55 「単一強スパイク固有値モデルに対する高次元平均ベクトルの 2 標本検定」

石井晶 (東京理科大学理工学部), 矢田和善 (筑波大学数理物質系数学域),

青嶋誠 (筑波大学数理物質系数学域)

10:55~11:35 「13 の癌腫の 78 個の Microarray による LSD の確認と Method1 による SM と BGS 分割の評価 –Big データ時代における LSD の重要な 2 つの特徴–」

新村秀一 (成蹊大学)

11:35~12:15 「ベイズモデルの母数：事前分布と estimand」

柳本武美 (統計数理研究所)

12:15~13:30 昼休み

座長 矢田和善 (筑波大学数理物質系数域)

13:30~14:10 「Gaussian Semiparametric Estimation for Two Dimensional Intrinsic Stationary Random Fields」

矢島美寛 (東北大学大学院経済学研究科)

14:10~14:50 「再生核ヒルベルト空間における射影のモーメントによる二標本検定」

牧草夏実 (千葉大学大学院融合理工学府)

14:50~15:30 「非負データに対する再帰的な密度推定について」

柿沢佳秀 (北海道大学経済学研究院)

座長 島谷健一郎 (統計数理研究所)

15:45~16:25 「ノンパラメトリックな極値統計解析法の漸近理論」

吉田拓真 (鹿児島大学大学院理工学研究科)

16:25~17:05 「ノンパラメトリック重回帰分析における Lasso 型推定の改良」

松島佑樹 (島根大学大学院自然科学研究科), 内藤貫太 (千葉大学大学院理学研究院)

17:05~17:45 「Semiparametric regression with local divergence」

内藤貫太 (千葉大学大学院理学研究院)

12月13日 (金)

座長 内藤貫太 (千葉大学大学院理学研究院)

10:15~10:55 「市場創造時点の特定に向けた文献調査：KIP を前提として」

加藤淳一 (久留米大学商学部)

10:55~11:35 「Estimation, Diagnostics, and Extensions of Nonparametric Hawkes Processes with Kernel Functions」

庄建倉 (統計数理研究所)

11:35~12:15 「非定常クラスター過程のパラメータ推定法：Palm 尤度法の拡張」

島谷健一郎 (統計数理研究所)

12:15~12:20 閉会

Non-stationary Time Series Classification using Topological Data Analysis

Yutaka Kuroki

Graduate School of Engineering, Tokyo University of Science

Takayuki Shiohama

Department of Information and Computer Technology

Abstract

Time series classification (TSC) is an important and challenging problem in data mining. There are hundreds of algorithms for TSC available with the increase of time series data availability. For more details, we refer recent work of Bagnall et al. (2017) and review paper of Fawaz et al. (2019). Some of the machine learning algorithms are based on the bug of patterns, and learning patterns of similarity are key feature extraction for time series data. In this study, we employ feature extraction of time series using Topological Data Analysis (TDA). TDA refers to a collection of methods for finding topological structure in data. Until recently, topological inference relied on deterministic approaches, and it is well known that these inference is easily affected by outliers and/or noisy datasets. Moreover, the high computational costs are required for computing persistence homology with complex datasets in time and space. To overcome these circumstances, two approaches are proposed in literature. The first approach is the bootstrap estimation for persistence diagrams and landscapes of Chazal et al. (2014). The second refers the methods of subsampling of Chazal et al. (2015). In this study, we introduce an unsupervised classification learning for several non-stationary time series using topological features. The bootstrap and subsampling methods are implemented to compare the performance of classifications.

Keywords: Bootstrap estimation, time series analysis, topological data analysis, subsampling method.

References

- Bagnall, A., J. Lines, A. Bostrom, J. Large, and E. Keogh (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery* 31(3), 606–660.
- Chazal, F., B. Fasy, F. Lecci, B. Michel, A. Rinaldo, and L. Wasserman (2015). Subsampling methods for persistent homology. In *International Conference on Machine Learning*, pp. 2143–2151.
- Chazal, F., B. T. Fasy, F. Lecci, A. Rinaldo, A. Singh, and L. Wasserman (2014). On the bootstrap for persistence diagrams and landscapes. *arXiv preprint arXiv:1311.0376*.
- Fawaz, H. I., G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33(4), 917–963.

Maximum Likelihood Estimation for Hidden Markov Models with Cylindrical Distributions

Takuto Kotsubo¹ and Takayuki Shiohama¹

¹ Graduate School of Engineering, Tokyo University of Science

² Department of Information and Computer Technology, Tokyo University of Science

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)) and the turning angles 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).

Hidden Markov models assume a sequence of random variable to be conditionally independent given a sequence of state variables which forms a Markov chain. To provide the consistency and asymptotic normality of maximum likelihood estimators (MLE), we need to consider identifiability of the model parameters. Recently, Miyata et al. (2019a) provide a strong consistency results for the mixtures of asymmetric circular distribution. Miyata et al. (2019b) investigate the identifiability of circular or cylindrical distribution. In this study, we consider the asymptotic properties for the maximum likelihood estimation for the HMM with cylindrical distribution.

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.
- Miyata, Y., T. Shiohama, and T. Abe (2019a). Estimation of finite mixture models of skew-symmetric circular distributions. To appear in *Metrika*.
- Miyata, Y., T. Shiohama, and T. Abe (2019b). Identifiability of asymmetric circular and cylindrical distributions. *arXiv:1908.09114*.
- 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.

発表者

三浦良造 一橋大学名誉教授 東北大学経済学研究科特任(客員)教授

発表題目

Jensen's alpha in the CAPM market model from the viewpoint of asymptotic accuracy of estimators of β and of asymmetry of distributions of residuals.

(This is a joint work with Mr. Navruzбек Karamatov, Ph.D student at Tohoku University)

: 1. 研究の動機

単純線形回帰モデルの傾き β の推定精度とモデルの偶然誤差項の分布の対称性・非対称性を軸にして、回帰モデルの切片の推定値にどの程度の差異が現れるか、二通りの推定値のどちらを採用すべきかの問題を扱う。

金融のアカデミアと実務においては傾きの推定に最小二乗推定(LSE)を使っている。これは BLUE(最小分散不偏線形推定)と呼ばれる。しかし、順位統計量に基づく推定量は非線形推定量であり、外れ値に対してロバストであることが知られている。偶然誤差項の確率分布が裾が重い場合は LSE よりも順位推定を使う方が推定精度がよいという意味で適切だろう。

この立場に立って、二つの異なる回帰残差から求められる切片 α の推定値にどの程度の差があるかを調べ、それが残差分布の対称性・非対称性とどのような関係を持つかを調べた。

: 2. 発表概要

偶然誤差項の確率分布を一般化された Lehmann 対立仮説モデル(セミパラメトリックなモデル)であらわし、残差に基づいて、このモデルが含む、位置母数と非対称度を示すパラメーターの推定量を構成し、これらの推定量の漸近正規性の数学的証明を報告した。

さらに、セミパラメトリックモデルに含まれる関数形が未知である対称な確率分布 F を推定する推定量を構成し、その漸近正規性の証明を報告した。

これらの数理的結果を支えとして、実際の株式市場でこれらのパラメータがとる数値を算出した。米国、英国、日本の3市場の日次データを用いた。

β の推定を二つの方法で行うので、残差が二通り得られ、それぞれがパラメーター α , θ , μ の推定に使われる。ここから得られる膨大な数値結果をまとめる軸として、(1) α を偶然誤差項分布の位置と非対称効果(非対称性がもたらす期待値の大きさへの貢献量)の和として表す α の分解、さ

らに(2)算出された α の二通りの α の値の差が残差の非対称度、偶然誤差項分布の裾の重さとどのように関係するかをグラフに表した。

α は株式ポートフォリオのパフォーマンス分析に用いられるが、LSE による推定値が順位統計量による推定値に比べて過大推定、あるいは過小推定である様子を示す以下のグラフに沿って報告した。そこでは、過大・過少の基準は β の推定精度が高いほうを正しい推定とみなして、正しいほうをもたらす残差から得た α を正しい推定値とした。

▶ F is closer to Normal.

e tail is as heavy as Logistic.

$\theta = 1$ means G is symmetric.

$\theta < 1$ right skewed: right tail is long/heavy

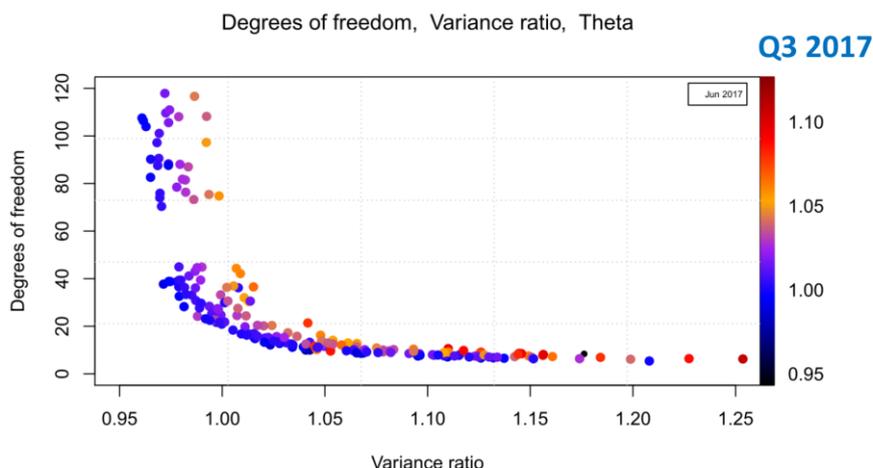
$\theta > 1$ is left skewed

ue, purple and orange dots along with θ values.

pts are located on the right.

Fig. 60

ta makes variance ratio larger for the same df dots.



: References.

Jaekel, L.A. (1972). Estimation of regression coefficients by minimizing the dispersion of the residuals. *The Annals of Mathematical Statistics*. Pp.1449-1458.

Jureckova, J.(1971). Nonparametric estimate of regression coefficients. *The Annals of Mathematical Statistics*. Pp.1328-1338.

Miura, R. and Tsukahara, H.(1993). One-sample estimation for generalized Lehmann's alternative models. *Statistica Sinica*. Pp.83-101.

Miura, R.(2014). A simple linear regression model with Generalized Lehmann's Alternative Model and R-estimators. *Shogaku Ronnkyu*. Vol.61. No.3. pp.89-108 (in Japanese) (Journal of Business Administration Kwansei Gakuin University)

正方分割表における 部分対称性に関する種々のモデルとその尺度

東京理科大学大学院理工学研究科	桃崎 智隆
東京理科大学大学院理工学研究科	福本 伸樹
東京理科大学理工学部	中川 智之
東京理科大学理工学部	石井 晶
横浜市立大学医学部	三枝 祐輔
東京理科大学理工学部	富澤 貞男

正方 $r \times r$ 分割表を考え, (i, j) セル確率を p_{ij} ($i = 1, \dots, r; j = 1, \dots, r$) とする. 本稿では, $\{p_{ij} + p_{ji} > 0\}$ を仮定する.

第1部 正方分割表における条件付き部分対称性とその尺度

行変数を X , 列変数を Y とする. McCullagh (1978) によって提案された条件付き対称 (CS) モデルは,

$$P(X = i, Y = j | X < Y) = P(X = j, Y = i | X > Y) \quad (\text{すべての } i < j)$$

により定義される.

一般に, CS モデルが成り立たないとき, CS モデルとの隔たりの程度を測ることに関心がある. Tomizawa and Saitoh (1999) は CS モデルからの隔たりを測る尺度 $\Phi^{(\lambda)}$ を提案した:

$$\Phi^{(\lambda)} = \frac{1}{2} \sum_{i=1}^{r-1} \sum_{j=i+1}^r (p_{ij}^U + p_{ji}^L) \Psi_{ij}^{(\lambda)} \quad (\lambda > -1),$$

ただし,

$$\Psi_{ij}^{(\lambda)} = 1 - \frac{\lambda 2^\lambda}{2^\lambda - 1} G_{ij}^{(\lambda)}, \quad G_{ij}^{(\lambda)} = \frac{1}{\lambda} \left[1 - (q_{ij})^{\lambda+1} - (q_{ji})^{\lambda+1} \right],$$

$$q_{ij} = \frac{p_{ij}^U}{p_{ij}^U + p_{ji}^L}, \quad q_{ji} = \frac{p_{ji}^L}{p_{ij}^U + p_{ji}^L}, \quad p_{ij}^U = \frac{p_{ij}}{\delta_U}, \quad p_{ji}^L = \frac{p_{ji}}{\delta_L} \quad (i < j),$$

$$\delta_U = \sum_{s < t} \sum p_{st}, \quad \delta_L = \sum_{s > t} \sum p_{st},$$

CS モデルが成り立たないとき, CS モデルより制約を緩めたモデルとその尺度に関心がある. そこで, CS モデルより制約を緩めたモデルである条件付き部分対称 (CPS) モデルと, CPS モデルからの隔たりを測る尺度 $\tau^{(\lambda)}$ を提案した.

CPS モデルを以下のように定義する:

$$P(X = i, Y = j | X < Y) = P(X = j, Y = i | X > Y) \quad (\text{少なくとも 1 組の } (i, j), i < j).$$

また, CPS モデルからの隔たりを測る尺度を次のように提案した:

$$\tau^{(\lambda)} = \prod_{i=1}^{r-1} \prod_{j=i+1}^r \left[\Psi_{ij}^{(\lambda)} \right]^{(p_{ij}^U + p_{ji}^L)/2} \quad (\lambda > -1)$$

$\tau^{(\lambda)}$ は次の性質を持つ.

(i) $0 \leq \tau^{(\lambda)} \leq 1$.

(ii) $\tau^{(\lambda)} = 0 \Leftrightarrow$ 少なくとも1組の (k, l) , $k < l$ に対して,
 $P(X = k, Y = l | X < Y) = P(X = l, Y = k | X > Y)$ が成立.

(iii) $\tau^{(\lambda)} = 1 \Leftrightarrow$ すべての $i \neq j$ に対して, $p_{ij} = 0$ (このとき, $p_{ji} \neq 0$) が成立.

推定尺度 $\hat{\tau}^{(\lambda)}$ は近似的に $N(\tau^{(\lambda)}, n^{-1}\sigma^2[\tau^{(\lambda)}])$ に従う ($n = \sum_{i,j} n_{ij}$, n_{ij} は (i, j) セル度数). 講演では, $\tau^{(\lambda)}$ に関するシミュレーション結果や信頼区間を用いた実データ解析結果について示した.

第2部 正方分割表における部分対称と部分非対称からの隔たりを測るベクトル型尺度

本講演では, 部分的な対称セルで非対称構造を持つモデルとして部分完全非対称 (PAS) モデルを以下で定義した:

$$p_{ij} = 0 \quad (p_{ji} \neq 0) \quad (\text{少なくとも1組の } i \neq j)$$

また, $p_{ij}^c = p_{ij}/(p_{ij} + p_{ji})$, $p_{ij}^* = p_{ij}/\delta$, $\delta = \sum \sum_{i \neq j} p_{ij}$ とする. PAS モデルからの隔たりを測る尺度を次のように提案した:

$$\kappa^{(\lambda)} = \prod_{i=1}^{r-1} \prod_{j=i+1}^r [1 - \phi_{ij}^{(\lambda)}]^{(p_{ij}^* + p_{ji}^*)} \quad (\lambda > -1)$$

ただし,

$$\phi_{ij}^{(\lambda)} = 1 - \frac{\lambda 2^\lambda}{2^\lambda - 1} H_{ij}^{(\lambda)}, \quad H_{ij}^{(\lambda)} = \frac{1}{\lambda} [1 - (p_{ij}^c)^{\lambda+1} - (p_{ji}^c)^{\lambda+1}].$$

部分対称 (PS) モデルからの隔たりを測る尺度を $\Upsilon^{(\lambda)}$ とする. PS モデルと PAS モデルからの隔たりを同時に測るベクトル型尺度を次のように提案した:

$$\mathbf{\Lambda}^{(\lambda)} = (\Upsilon^{(\lambda)}, \kappa^{(\lambda)})^\top \quad (\lambda > -1)$$

$\mathbf{\Lambda}^{(\lambda)}$ は次の性質を持つ.

(i) $\mathbf{\Lambda}^{(\lambda)} = (0, 0)^\top \Leftrightarrow$ PS モデルと PAS モデルが同時に成立.

(ii) $\mathbf{\Lambda}^{(\lambda)} = (0, 1)^\top \Leftrightarrow$ 対称モデル (すべての $i \neq j$ に対して $p_{ij} = p_{ji}$) が成立.

(iii) $\mathbf{\Lambda}^{(\lambda)} = (1, 0)^\top \Leftrightarrow$ 完全非対称モデル (すべての $i \neq j$ に対して $p_{ij}^c = 1$ または $p_{ji}^c = 1$) が成立.

推定尺度 $\hat{\mathbf{\Lambda}}^{(\lambda)}$ は近似的に $N(\mathbf{\Lambda}^{(\lambda)}, n^{-1}\mathbf{\Sigma})$ に従う. 講演では, $\mathbf{\Lambda}^{(\lambda)}$ に関するシミュレーション結果や, 信頼楕円を用いた実データ解析について示した.

参考文献

McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, 65:413–418.

Tomizawa, S. and Saitoh, K. (1999). Measure of departure from conditional symmetry for square contingency tables with ordered categories. *Journal of the Japan Statistical Society*, 29:65–78.

多次元分割表における多要因間独立性と多次元完全独立性の関係

種市 信裕 (北海道教育大・札幌)
関谷 祐里 (北海道教育大・釧路)
外山 淳 (数学利用研究所)

1. 2次元分割表の独立性モデル

2次元の $J \times K$ 分割表において多項分布モデルを考える。 (j, k) セルの観測度数を表す確率変数を X_{jk} とする。ただし, X_{jk} ($j = 1, \dots, J, k = 1, \dots, K$) は, $\sum_{j=1}^J \sum_{k=1}^K X_{jk} = n$ を満たす非負整数の値を取り, 自然数 n は定数とする。また, (j, k) セルのセル確率を p_{jk} とする。ただし, p_{jk} ($j = 1, \dots, J, k = 1, \dots, K$) は, $0 < p_{jk} < 1$ ($j = 1, \dots, J, k = 1, \dots, K$), $\sum_{j=1}^J \sum_{k=1}^K p_{jk} = 1$ を満たすものとする。このとき, $p = (p_{11}, \dots, p_{JK})^\top$ とおき, 確率変数ベクトル $X = (X_{11}, \dots, X_{JK})^\top$ が多項分布 $\mathcal{M}_{JK}(n, p)$ に従う場合について考える。

周辺確率及び周辺度数を,

$$p_{j\cdot} = \sum_{k=1}^K p_{jk}, p_{\cdot k} = \sum_{j=1}^J p_{jk}, X_{j\cdot} = \sum_{k=1}^K X_{jk}, X_{\cdot k} = \sum_{j=1}^J X_{jk} \quad (j = 1, \dots, J, k = 1, \dots, K)$$

と定義するとき, $J \times K$ 分割表における独立性の仮説

$$(1) \quad H_0^{(1)} : p_{jk} = p_{j\cdot} p_{\cdot k} \quad (j = 1, \dots, J, k = 1, \dots, K)$$

を検定するための ϕ -divergence に基づく統計量は,

$$C_\phi^{(1)} = 2n \sum_{j=1}^J \sum_{k=1}^K \frac{X_{j\cdot} X_{\cdot k}}{n^2} \phi \left(\frac{n X_{jk}}{X_{j\cdot} X_{\cdot k}} \right)$$

で定義される。ここで, $\phi(t)$ は, 区間 $(0, \infty)$ における凸関数で, $\phi(1) = \phi'(1) = 0$ と $\phi''(1) = 1$ を満たすものとする。帰無仮説 $H_0^{(1)}$ のもとで, $n \rightarrow \infty$ としたときの検定統計量 $C_\phi^{(1)}$ の極限分布は, すべての関数 ϕ に対して, 自由度 $(J-1)(K-1)$ のカイ二乗分布である。

2. 多次元分割表の独立性モデル

3次元の $J \times K \times L$ 分割表において多項分布モデルを考える。 (j, k, ℓ) セルの観測度数を表す確率変数を X_{jkl} , セル確率を p_{jkl} とする。このとき, 確率変数ベクトル $X = (X_{111}, \dots, X_{JKL})^\top$ が多項分布 $\mathcal{M}_{JKL}(n, p)$ に従う場合について考える。ここで, $p = (p_{111}, \dots, p_{JKL})^\top$ である。

周辺確率と周辺度数に対しては, 前節で用いた記法と同様に, 和を取った添え字を \cdot で表すことによって, 記号 $p_{j\cdot\cdot}, p_{\cdot k\cdot}, p_{\cdot\cdot\ell}, p_{jk\cdot}, p_{j\cdot\ell}, p_{\cdot k\ell}$ 等を定義することとする。このとき, $J \times K \times L$ 分割表のセル確率に関する以下の仮説を考える。

$$(2) \quad H_0^{(2)} : p_{jkl} = p_{j\cdot\cdot} p_{\cdot k\cdot} p_{\cdot\cdot\ell} \quad (j = 1, \dots, J, k = 1, \dots, K, \ell = 1, \dots, L)$$

$$(3) \quad H_0^{(3)} : p_{jkl} = p_{j\cdot\cdot} p_{\cdot k\ell} \quad (j = 1, \dots, J, k = 1, \dots, K, \ell = 1, \dots, L)$$

(2) は完全独立性の仮説, (3) は 1 要因対 2 要因の独立性の仮説である。帰無仮説 $H_0^{(2)}, H_0^{(3)}$ を検定するための ϕ -divergence に基づく統計量はそれぞれ,

$$C_\phi^{(2)} = 2n \sum_{j=1}^J \sum_{k=1}^K \sum_{\ell=1}^L \frac{X_{j..} X_{.k.} X_{..\ell}}{n^3} \phi \left(\frac{n^2 X_{jkl}}{X_{j..} X_{.k.} X_{..\ell}} \right)$$

$$C_\phi^{(3)} = 2n \sum_{j=1}^J \sum_{k=1}^K \sum_{\ell=1}^L \frac{X_{j..} X_{.k\ell}}{n^2} \phi \left(\frac{n X_{jkl}}{X_{j..} X_{.k\ell}} \right)$$

で定義される。帰無仮説のもとで, 検定統計量 $C_\phi^{(2)}, C_\phi^{(3)}$ の極限分布はそれぞれ, 自由度 $JKL - J - K - L + 2$, 自由度 $(J - 1)(KL - 1)$ のカイ二乗分布である。

4 次元の $J \times K \times L \times M$ 分割表の多項分布モデルにおいて, 以下の仮説を考える。

- (4) $H_0^{(4)} : p_{jklm} = p_{j...} p_{.k..} p_{..\ell} p_{...m}$ ($j = 1, \dots, J, k = 1, \dots, K, \ell = 1, \dots, L, m = 1, \dots, M$)
- (5) $H_0^{(5)} : p_{jklm} = p_{j..} p_{.k.} p_{..\ell m}$ ($j = 1, \dots, J, k = 1, \dots, K, \ell = 1, \dots, L, m = 1, \dots, M$)
- (6) $H_0^{(6)} : p_{jklm} = p_{j.} p_{.k\ell m}$ ($j = 1, \dots, J, k = 1, \dots, K, \ell = 1, \dots, L, m = 1, \dots, M$)
- (7) $H_0^{(7)} : p_{jklm} = p_{jk.} p_{..\ell m}$ ($j = 1, \dots, J, k = 1, \dots, K, \ell = 1, \dots, L, m = 1, \dots, M$)

(4) は完全独立性の仮説, (5) は 1 要因対 1 要因対 2 要因の独立性の仮説, (6) は 1 要因対 3 要因の独立性の仮説, (7) は 2 要因対 2 要因の独立性の仮説である。 $H_0^{(4)}, H_0^{(5)}, H_0^{(6)}, H_0^{(7)}$ の各々の帰無仮説を検定するための検定統計量の, $n \rightarrow \infty$ としたときの極限分布はそれぞれ, 自由度 $JKLM - J - K - L - M + 3$, $JKLM - J - K - LM + 2$, $(J - 1)(KLM - 1)$, $(JK - 1)(LM - 1)$ をもつカイ二乗分布である。

5 次元以上の分割表においても同様に, 完全独立性の仮説, および, 種々の多要因間独立性の仮説が考えられる。

3. 分割表の独立性モデルの関係

3 次元分割表のセル確率 p_{jkl} ($j = 1, \dots, J, k = 1, \dots, K, \ell = 1, \dots, L$) に対して, $K \times L$ の長方形に並んでいるセル確率をすべて 1 列に並べ直し, 添え字を付け直すことによって得られる新たなセル確率 p_{ji} ($j = 1, \dots, J, i = 1, \dots, KL$) を考えると, 3 次元の $J \times K \times L$ 分割表における 1 要因対 2 要因の独立性モデル (3) は, 2 次元の $J \times KL$ 分割表における独立性モデルに帰着する。これを一般化すると, 多次元分割表における r 個の要因グループ間の独立性モデルは, r 次元分割表の完全独立性モデルに帰着する。

この関係に着目し, Kobe et al. [1] は, 2 次元分割表における独立性検定についての Taneichi and Sekiya [2] の結果を基に, 3 次元分割表の 2 要因対 1 要因の独立性検定の改良をおこなった。同様に, 多次元分割表における多要因間独立性検定は, より低次元の分割表における完全独立性検定の改良理論を基に, 容易に改良することができる。

参考文献

- [1] T. Kobe, N. Taneichi and Y. Sekiya, Improved transformed statistics for the test of one factor independence from the other two in an $r \times s \times t$ contingency table. *J. Japan Statist. Soc.*, **45**(1), (2015), 77–98.
- [2] N. Taneichi and Y. Sekiya, Improved transformed statistics for the test of independence in $r \times s$ contingency tables. *J. Multivariate Anal.*, **98**, (2007), 1630–1657.

単一強スパイク固有値モデルに対する 高次元平均ベクトルの2標本検定

石井 晶 (東京理科大学理工学部)
矢田 和善 (筑波大学数理物質系数域)
青嶋 誠 (筑波大学数理物質系数域)

1. はじめに

母集団が2つあり、各母集団 π_i は、平均に p 次のベクトル $\boldsymbol{\mu}_i$ 、共分散行列に p 次の正定値対称行列 $\boldsymbol{\Sigma}_i$ をもつとする。各母集団から、 n_i 個の p 次元データベクトル $\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{in_i}$ を無作為に抽出し、 $p \times n_i$ データ行列 $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{in_i}]$ を構成する。ここで、 $p > n_i$ である。 $\boldsymbol{\Sigma}_i$ の固有値を $\lambda_{i(1)} \geq \dots \geq \lambda_{i(p)} (> 0)$ とし、適当な直交行列 $\boldsymbol{H}_i = [\boldsymbol{h}_{i(1)}, \dots, \boldsymbol{h}_{i(p)}]$ で $\boldsymbol{\Sigma}_i$ を $\boldsymbol{\Sigma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i^T$, $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{i(1)}, \dots, \lambda_{i(p)})$ と分解する。各母集団において、標本共分散行列を $\boldsymbol{S}_i = (n_i - 1)^{-1} (\boldsymbol{X}_i - \bar{\boldsymbol{X}}_i)(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_i)^T$ と定義する。ここで、 $\bar{\boldsymbol{X}}_i = [\bar{\boldsymbol{x}}_i, \dots, \bar{\boldsymbol{x}}_i]$, $\bar{\boldsymbol{x}}_i = n_i^{-1} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}$ である。Aoshima and Yata [3] は、高次元データの固有値モデルを次の2つに分類した。1つ目は、強スパイク固有値モデル (strongly spiked eigenvalue (SSE) モデル) とよばれるもので、以下のように定義される。

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

2つ目は、弱スパイク固有値モデル (non-SSE モデル) とよばれるもので、以下のように定義される。

$$\frac{\lambda_{i(1)}^2}{\text{tr}(\boldsymbol{\Sigma}_i^2)} \rightarrow 0 \quad (p \rightarrow \infty) \quad (2)$$

本講演は、典型的な SSE モデル (1) として、次の固有値モデルを考える。

$$(A-i) \quad \frac{\sum_{s=2}^p \lambda_{i(s)}^2}{\lambda_{i(1)}^2} = o(1) \quad (p \rightarrow \infty)$$

Ishii et al. [6] は (A-i) を「単一強スパイク固有値モデル (単一 SSE モデル)」とよんで、単一 SSE モデル (A-i) のもとで高次元共分散構造の検定を与えた。本講演では、「 $p \rightarrow \infty$ だが、 n_1, n_2 は固定」の枠組みで高次元小標本漸近理論を展開し、単一 SSE モデル (A-i) における新たな2標本検定とその漸近的性質を論じた。

2. 単一 SSE モデルにおける2標本検定

次の仮説を考える。

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \quad (3)$$

高次元データの2標本検定に対して、次の検定統計量はスタンダードである。

$$T_n = \|\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2\|^2 - \sum_{i=1}^2 n_i^{-1} \text{tr}(\boldsymbol{S}_i)$$

Bai and Saranadasa (1996, *Stat. Sin.*), Chen and Qin (2010, *Ann. Stat.*), Aoshima and Yata [1, 2] は、 T_n の漸近正規性を non-SSE モデル (2) のもとで論じ、それぞれ異なる高次元

2 標本検定を与えた. 一方, SSE モデル (1) においては, T_n の漸近正規性は崩れることが, Aoshima and Yata [3] により示された. Ishii [4] は, 単一 SSE モデル (A-i) のもとで T_n を再評価し, 高次元 2 標本検定を与えた. Ishii [4] は, 第 1 固有空間について, 次を仮定した.

$$(C-i) \quad \frac{\lambda_{1(1)}}{\lambda_{2(1)}} = 1 + o(1), |\mathbf{h}_{1(1)}^T \mathbf{h}_{2(1)}| = 1 + o(1) \quad (p \rightarrow \infty)$$

実際のデータに対する仮定 (C-i) の妥当性は, Ishii et al. [5] が与えた第 1 固有空間の同等性検定を用いて確認することができる. Yata and Aoshima [7] のノイズ掃き出し法による $\lambda_{i(1)}$ の推定量を $\tilde{\lambda}_{i(1)}$ とおくと, 以下の検定統計量が考えられる.

$$F_a = u_n(n_1 + n_2 - 2) \frac{T_n + \sum_{i=1}^2 \tilde{\lambda}_{i(1)}/n_i}{\sum_{i=1}^2 (n_i - 1) \tilde{\lambda}_{i(1)}}$$

ここで, $u_n = (1/n_1 + 1/n_2)^{-1}$ とする. F_a の漸近帰無分布は次のようになる.

定理 2.1 (Ishii [4]). 各母集団に (A-i) を仮定する. また, (C-i) を仮定する. そのとき, いくつかの正則条件と H_0 のもとで, 次が成り立つ.

$$F_a \Rightarrow F_{1,\nu} \quad (p \rightarrow \infty \text{ だが, } n_1, n_2 \text{ は固定})$$

ここで, $\nu = n_1 + n_2 - 2$, F_{ν_1, ν_2} は自由度 (ν_1, ν_2) の F 分布に従う確率変数を表す.

与えられる $\alpha \in (0, 1/2)$ に対して, 仮説 (3) について, 次の検定方式を考える.

$$F_a \geq F_{1,\nu}(\alpha) \text{ のとき, } H_0 \text{ を棄却する} \quad (4)$$

ここで, $F_{\nu_1, \nu_2}(\alpha)$ は F_{ν_1, ν_2} の上側 α 点を表す. Ishii [4] は, 検定方式 (4) の第 1 種の過誤と検出力に関する理論的な結果を与え, 検定方式 (4) の漸近的な性能を保証した. しかし, 実際のデータ解析においては第 1 固有空間の同等性の仮定 (C-i) は必ずしも成立しない. そこで, 本講演では, (C-i) を仮定できない場合に, 新たな高次元 2 標本検定を与た.

参考文献

- [1] Aoshima, M. and Yata, K. (2011). Two-stage procedures for high-dimensional data, *Sequential Analysis (Editor's special invited paper)*, **30**, 356-399.
- [2] Aoshima, M. and Yata, K. (2015). Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions, *Methodology and Computing in Applied Probability*, **17**, 419-439.
- [3] Aoshima, M. and Yata, K. (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models, *Statistica Sinica*, **28**, 43-62.
- [4] Ishii, A. (2017). A two-sample test for high-dimension, low-sample-size data under the strongly spiked eigenvalue model, *Hiroshima Mathematical Journal*, **47**, 273-288.
- [5] Ishii, A., Yata, K. and Aoshima, M. (2016). Asymptotic properties of the first principal component and equality tests of covariance matrices in high-dimension, low-sample-size context, *Journal of Statistical Planning and Inference*, **170**, 186-199.
- [6] Ishii, A., Yata, K. and Aoshima, M. (2019). Tests for high-dimensional covariance structures, submitted.
- [7] Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations, *Journal of Multivariate Analysis*, **105**, 193-215.

13 の癌腫の 78 個の Microarray による LSD の確認と Method1 による SM と BGS 分割の評価

— Big データ時代における LSD の重要な 2 つの特徴 —

新村秀一 成蹊大学名誉教授

1999 年から 2004 年に米国の 6 医学研究が医学誌に論文を発表し、研究に用いたデータを公開している。これらの医学研究では癌と健常、あるいは異なった 2 種の癌症例を用いて、「癌の遺伝子診断に役に立つ癌の遺伝子の組」を特定することを目的としている。筆者は、2015 年 10 月 25 日に Jeffery らが、これらの 6 種のマイクロアレイを HP に公開していることを知った。そこで 10 月 28 日にそれらをダウンロードし、僅か 54 日間で判別理論の応用問題として簡単に次のように解決した。

- 1) **MNM** 基準による Revised IP-OLDF (**RIP**) で 6 個のデータを判別し、 $MNM=0$ であり、2 群が遺伝子空間で線形分離していることが分かった。線形分離可能なデータ (**LSD**) である遺伝子空間とその部分空間を Matryoshka と呼ぶ。スイス銀行紙幣データ等の LSD 判別研究で、LSD は「Matryoshka 構造」を持つことが分かった。すなわち LSD はその中に小さな Matryoshka を最小次元の Matryoshka である **BGS** まで入れ子状に含んでいる。iPS 細胞研究で、山中らは 2 万個以上といわれる遺伝子から、Rats の ES 細胞で活性化している 24 個の遺伝子で万能細胞の塊ができることを見つけた。そして重回帰分析の変数減少法に似た方法で、24 個から 4 個の山中遺伝子を特定した。4 個から 1 個を省くと万能細胞ができないことと BGS から 1 個省くと MNM が 1 以上になることは同じ概念である。高次元データから癌の遺伝子を特定することは NP-hard ではあるが、このような方法によるアプローチが必要である。ただし偏差平方和の最小化でなく MNM 基準で行う必要がある。
- 2) もし BGS が k 組あれば、簡単に k 組の小標本に分割でき高次元の呪いから解放される。このために、数理計画法ソフトの LINGO で BGS を求める Program4 を開発した。しかし、遺伝子数 p が大きいと計算時間がかかるので、Matryoshka Feature Selection Method (**Method2**) で Small Matryoshka (**SM**) を見つける Program3 を開発した。

すなわち高次元であろうが小標本であろうがデータが LSD であれば、**1) Matryoshka 構造を持ち、2) 複数の排他的な BGS (あるいは SM) と MNM が 1 以上のその他の部分空間に分割できる**。そして 6 個の高次元データから求めた小標本を統計分析し、マイクロアレイによる癌の遺伝子診断を世界で初めて提案した。

しかし用いたデータは 2004 年までのものであり、癌腫が限られている。また僅か 6 種類のデータが LSD であっても、他の多くのデータが LSD でないと、医学研究への適用が限られる。そこで Bruno らが開発した Curated Microarray Database (**CuMiDa**) が最適と判断した。そしてこの DB から 13 種類の異なる癌の 78 個のデータをダウンロードした。そして、2 群判別で全てが $MNM=0$ であることが分かった。すなわち、すでに分析した 6 個のデータに加えて、新しい 77 個のデータでも全て LSD である。これで高次元データの呪いから解放され、SM あるいは BGS を統計分析し癌の遺伝子診断が行えるという筆者の主張の正しさと汎用性が確認できた。CuMiDa では、今後 RNA-seq などの遺伝子データの搭載も検討している。今回の研究から、恐らく RNA-seq でも LSD であることが確認できると考えている。もし確認できない場合、マイクロアレイより RNA-seq の方が優れているという研究者もいるが、RNA-seq が癌の遺伝子診断に役に立たないことを示す。以上の結果は、マイクロアレイやその他の遺伝子データによる癌の遺伝子診断を今後大きく発展させることが期待できる。

また医学研究者が、新規の研究でマイクロアレイや RNA-seq を集めた場合、最初に LSD であることを確認すべきである。LSD であれば、これまでの多くの研究方法は全て無用になり、SM と BGS に分割して遺伝子診断を行えばよい。筆者の方法で重要な信号はもれなく列挙できる。しかし数多くある SM と BGS の遺伝子診断の重要度を検討する必要がある。このために、すでに開発した 100 重交差検証法 (**Method1**) で、検証標本の平均誤分類確率 $M2$ の小さな順に遺伝子診断の重要度のランク付けを行った。

ベイズモデルの母数：事前分布と estimand

柳本武美: 統計数理研究所

1. 序

標本分布の族を $p(x|\theta)|\theta \in \Theta \subset R^p$ とする。ベイズ推定では、事前密度 $\pi(\theta)$ を仮定して推論を行う。母数 θ の推定は事後平均

$$\hat{\theta} = E\{\theta; \pi(\theta)\}$$

で行うことが通常である。但し $E\{g(x); p(x)\}$ は関数 $f(x)$ の密度 $p(x)$ の下での期待値とする。推定量を得るための母数を estimand と呼ぶ。この場合は θ が estimand になる。ところでどのように estimand を選ぶと良いだろうか。確率モデルを仮定する場合には、母数の意味が取りやすいように選択する。特に informative な事前分布を仮定するためには、意味が取りやすい母数を選ぶ必要がある。しかし、母数を推定する段階では母数の意味について考えるより、リスクあるいはベイズリスクを小さくするように選ぶ必要がある。この違いは従来等閑視されてきたように思われる。別の理由として、応用的側面から母数が決まるだろうと考えることにある。この場合も応用的側面と推定量の効率をどう捉えるのかを明瞭にする必要がある。

2. 損失と estimand の選択

理論統計学の基本的枠組みでは、推定量はリスクによって評価される。リスクを小さくする推定量が良い推定量である。そうだとすると、損失 $L(\hat{\theta}, \theta)$ を定めて議論するのがよい。広く用いられる損失は二乗損失 $\|\hat{\theta} - \theta\|^2$ と Kullback-Leibler 損失

$$D(\hat{\theta}, \theta) = E\{\log\{p(\mathbf{y}|\hat{\theta})/p(\mathbf{y}|\theta); p(\mathbf{y}|\hat{\theta})\} \quad (2.1)$$

である。ここで $p(\mathbf{y}|\theta)$ の \mathbf{y} は予測分布を表す x と同じ次元の変数である。この損失を $L_e(\hat{\theta}, \theta)$ と書いたとき、その双対形 $L_m(\hat{\theta}, \theta) = D(\theta, \hat{\theta})$ も使われる。Kullback-Leibler 損失一対一に対応する母数の選択に依存しないが、二乗損失は依存する。しかも、二乗損失は母数が多次元の場合などには妥当性が理解しにくい問題がある。

3. 指数分布族の場合

指数分布族 $\mathcal{F} = \{p(x|\theta) = \exp\{n(\bar{x}\theta - M(\theta))\}a(x)|\theta \in \Theta\}$ をかんがえる。母数 θ は共役母数と呼ばれる。通常は、平均母数 $\mu = M'(\theta)$ の事後平均 $\hat{\mu}_m = E\{\mu|\pi(\mu|\pi(\theta|x))\}$ で推定する。対応する推定量を $\hat{\theta}_m = M'^{-1}(\hat{\mu}_m)$ と書く。この推定量の plug-in 予測子 $p(\mathbf{y}|M'(\hat{\mu}_m))$ は別の予測子 $p_m(\mathbf{y}|x) = E\{p(\mathbf{y}|\theta); \pi(\theta|x)\}$ より損失 $L_m(\hat{\theta}, \theta)$ の下でのベイズリスクが大きくなる。

この困難は estimand を共役母数 θ とすることにより解消される。plug-in 予測子は

$$p_e(\mathbf{y}|x) \propto \exp(E\{\log p(\mathbf{y}|\theta); \pi(\theta|x)\}). \quad (3.1)$$

となる。損失 $L_e(\hat{\theta}, \theta)$ の下で最適になる。この周辺は Corcuera and Giummole (1999), Yanagimoto and Ohnishi (2009, 2011) で調べられた。

次の等式はよく知られている。 $\log p(x|\hat{\theta}_{ML})/p(x|\theta) = D(p(\mathbf{y}|\hat{\theta}_{ML}), p(\mathbf{y}|\theta))$ がすべての x と θ に対して成り立つ。一方、すべての x に対して

$$E\left\{\log \frac{p_e(\mathbf{x}|\mathbf{x})}{p(\mathbf{x}|\theta)}; \pi(\theta|\mathbf{x})\right\} = E\{D(p_e(\mathbf{y}|\mathbf{x}), p(\mathbf{y}|\theta)); \pi(\theta|\mathbf{x})\}.$$

が成り立つ。このことは、MLE は損失を小さくし過ぎることを示している。

4. Laplace 分布の場合

平均が μ 分散が $1/\tau^2$ である Laplace 分布の密度関数は

$$p(x|\mu, \tau) = \frac{\tau}{2} \exp\{-\tau|x - \mu|\} \quad (4.1)$$

で与えられる。この分布は指数分布族に含まれない。歴史的には 標本中央値と MLE が一致することから、Laplace (1774), Keynes (1911) らの関心を惹いた。その結果この推定量の改善は Govindarajulu (1966) を始め多くの研究結果がある。この分布と推定量についての成果は Kotz ら (2001) に詳しく記述されている。

そこで、ベイズ推定における estimand を考察する。最初に考えつく (μ, τ) には問題が多い。 τ と $1/\tau$ の区別が難しいし、二乗誤差 $(\hat{\mu} - \mu)^2 + (\hat{\tau} - \tau)^2$ では二つの項の和を取ることの意味が取りにくい。

さて Kullback-Leibler 損失は

$$D((\hat{\mu}, \hat{\tau}), (\mu, \tau)) = \frac{\tau}{\hat{\tau}} \{\exp(-\hat{\tau}|\hat{\mu} - \mu|) + \hat{\tau}|\hat{\mu} - \mu| - 1\} + \frac{\tau}{\hat{\tau}} - \log \frac{\tau}{\hat{\tau}} - 1 \quad (4.2)$$

と書けるが扱いにくい。そこで、 $\hat{\mu} \doteq \mu$ で近似すると

$$L((\hat{\mu}, \hat{\tau}), (\hat{\mu}, \hat{\tau})) = \frac{1}{2} \tau \hat{\tau} (\hat{\mu} - \mu)^2 + \frac{\tau}{\hat{\tau}} - \log \frac{\tau}{\hat{\tau}} - 1. \quad (4.3)$$

となる。この損失の下でのベイズリスクを最小にする estimand を求めると $(\tau\mu, \tau)$ となる。

提案法は、損失 (4.3) の下でベイズリスクを最小にする意味での最適性が満たされている。この事実から、他の直感的あるいは強い制約条件の下での推定量よりも良い性能を示すと期待される。既存の推定量として

- 1) 最尤推定量同時に標本中央値
- 2) 最良線形不偏推定量 (Govindarajulu, 1966)
- 3) Estimand を μ, τ とした事後平均

が挙げられる。この中で上でも述べたように、1) は 2) に劣るので実質的には二つとの性能比較になる。そこで reference 事前分布を仮定した下で性能比較を実行して、良好な結果を得つつある。2) の推定量は線形不偏に制約する理由が分かりにくい。結果として性能が劣る。3) のベイズ推定量は estimand の選択の理由が直感的である。実際、 $n = 2$ の場合に μ の推定量が存在しなくなる。

実際に数値実験的に比較を行うとほぼ期待通りの結果が得られつつある。興味深いことに、 $1/\hat{\tau}$ が $1/\tau$ の推定量としてのバイアスが小さいことが観察された。この事実に対応する不偏性は、正規分布では解析的に示される。

References : 1) Corcuera, J. M. and Giummole F. (1999). *Scand. J. Statist.*, **26**, 265-279. 2) Garvan, C.W. and Ghosh, M. (1997). *Biometrika* **84**, 976-982. 3) Govindarajulu, Z. (1966). *J. Am. Statist. Assoc.*, **61**, 248-258. 4) Keynes, J.M. (1911). *J. Roy. Statist. Soc. A*, **74**, 322-328. 5) Kotz, S., Kozubowski, T. and Podgorski, K. (2001). *The Laplace Distribution and Generalizations*. Springer. 6) Laplace, P.S. (1774). *Mem. Acad. Roy. Sci. Paris*, **6**, 621-656. 7) Yanagimoto, T. and Ohnishi, T. (2009). *J. Statist. Plann. Inf.*, **139**, 3064-3075. 8) Yanagimoto, T. and Ohnishi, T. (2011). *J. Statist. Plann. Inf.*, **41**, 1990-2000.

On Gaussian semiparametric estimation for two-dimensional intrinsic stationary random fields

Yoshihiro Yajima
Tohoku University

We propose a Gaussian semiparametric estimator for semiparametric models of two-dimensional intrinsic stationary random fields (ISRFs) observed on a regular grid and derive its asymptotic properties. Originally this estimator is an approximate likelihood estimator in a frequency domain for long memory models of stationary and nonstationary time series (Robinson(1995); Velasco(1999)). We apply it to two dimensional ISRFs. These ISRFs include a fractional Brownian field, which is a Gaussian random field and is used to model many physical processes in space. The estimator is consistent and has the limiting normal distribution as the sample size goes to infinity. We conduct a computational simulation to compare the performance of it with those of different estimators.

Let $\{X(\mathbf{s}) : \mathbf{s} \in \mathbf{R}^d\}$ be a random field. If $\{X(\mathbf{s})\}$ satisfies that for any fixed $\mathbf{h}(\in \mathbf{R}^d)$, the increment $Z_{\mathbf{h}}(\mathbf{s}) = X(\mathbf{s} + \mathbf{h}) - X(\mathbf{s})$ is a stationary random field, $\{X(\mathbf{s})\}$ is called an ISRF. Then $\{X(\mathbf{s})\}$ is characterized by

$$\begin{aligned} E(X(\mathbf{s} + \mathbf{h}) - X(\mathbf{s})) &= 0, \\ \text{Var}(X(\mathbf{s} + \mathbf{h}) - X(\mathbf{s})) &= 2\gamma(\mathbf{h}), \end{aligned}$$

where $2\gamma(\mathbf{h})$ is the variogram function. Hereafter we also assume that $X(\mathbf{0}) = 0$.

Under some assumptions, the variogram function is expressed by

$$2\gamma(\mathbf{h}) = \int_{\mathbf{R}^d} \frac{1 - \cos((\boldsymbol{\lambda}, \mathbf{h}))}{(2\pi)^d} g(\boldsymbol{\lambda}) d\boldsymbol{\lambda},$$

where $g(\boldsymbol{\lambda})$ satisfies

$$\int_{\mathbf{R}^d} \frac{\|\boldsymbol{\lambda}\|^2}{1 + \|\boldsymbol{\lambda}\|^2} g(\boldsymbol{\lambda}) d\boldsymbol{\lambda} < \infty.$$

respectively.

Hereafter we assume that $d = 2$ and $\{X(\mathbf{s})\}$ is a two-dimensional Gaussian ISRF. Then we also denote $\boldsymbol{\lambda}$ by (λ_1, λ_2) and the spectral density function $g(\boldsymbol{\lambda})$ by $g(\lambda_1, \lambda_2)$ respectively.

We introduce the following assumptions.

Assumption 1 $g(\lambda_1, \lambda_2)$ is expressed by

$$g(\lambda_1, \lambda_2) = \|\boldsymbol{\lambda}\|^{-2H-2} g_o(\lambda_1, \lambda_2), \quad 0 < H < 1,$$

where $g_o(\lambda_1, \lambda_2)$ is a nonnegative with $g_o(0, 0) > 0$, symmetric, $g_o(\lambda_1, \lambda_2) = g_o(-\lambda_1, -\lambda_2)$, twice continuously differentiable function for $-\infty < \lambda_1, \lambda_2 < \infty$ and is bounded with bounded first and second order partial derivatives.

Then we have the following asymptotic properties of the estimator.

Theorem 1 Let \hat{H}_n be the Gaussian semiparametric estimator where n is the sample size. Under Assumption 1 and additional conditions, \hat{H}_n converges to H_0 in probability as $n \rightarrow \infty$. Moreover for $p \geq 2$, $m^{1/2}(\hat{H}_n - H_0)$ converges to $N(0, 1)$ in distribution as $n \rightarrow \infty$. where m is the number of the discrete Fourier transforms used to construct the estimator and p is the order of data tapers.

We conducted some computational simulations and compared the performance of our estimator with that of the estimator proposed by Zhu and Stein(2002) which are based on empirical variograms. Then our estimator is less biased than the estimator of Zhu and Stein if $g_o(\lambda_1, \lambda_2)$ is not a constant.

References

- Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630-1661.
- Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *J. Time Ser. Anal.* **20** 87-127.
- Zhu, Z. and Stein, M.L. (2002). Parametric estimation for fractional Brownian surfaces. *Statist. Sinica* **12** 863-883.

再生核ヒルベルト空間における射影のモーメントによる 二標本検定

千葉大・融合理工学府 牧草 夏実

1 はじめに

P, Q をヒルベルト空間 \mathcal{H} 上の確率分布とすると、二標本 $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} P, Y_1, \dots, Y_m \stackrel{i.i.d.}{\sim} Q$ に基づく検定

帰無仮説 $H_0 : P = Q$ vs. 対立仮説 $H_1 : P \neq Q$

を考える。ユークリッド空間での二標本検定はすでに様々な検定方法が議論されているが、ヒルベルト空間に値をとる確率変数に対する二標本検定を考えることで、高次元データに対する二標本検定の議論を与える。高次元データに対するアプローチとして、カーネル法という方法がある。カーネル法を用いた二標本検定として、Maximum Mean Discrepancy(MMD) に基づく二標本検定が [1] によりすでに議論されているが、この MMD と同様の考え方により、2 次のカーネルの MMD による二標本検定を考える。

2 再生核ヒルベルト空間でのモーメントの定義

確率変数 $X \sim P, Y \sim Q$ を正定値カーネル k によって、この k に対応する再生核ヒルベルト空間 $H(k)$ 上に、それぞれ $k(\cdot, X), k(\cdot, Y)$ により変換を行う。このとき、この $k(\cdot, X), k(\cdot, Y)$ の平均まわりの 2 次モーメント $\Sigma_k(P), \Sigma_k(Q)$ は、それぞれヒルベルト空間 $H(k)^{\otimes 2} = H(k) \otimes H(k)$ での期待値 $\Sigma_k(P) = \mathbb{E}_{X \sim P}[(k(\cdot, X) - \mu(P))^{\otimes 2}]$, $\Sigma_k(Q) = \mathbb{E}_{Y \sim Q}[(k(\cdot, Y) - \mu(Q))^{\otimes 2}]$ によって定められている。ここで、 $\mu(P), \mu(Q)$ は $k(\cdot, X)$ の 1 次モーメント $\mu(P) = \mathbb{E}_{X \sim P}[k(\cdot, X)]$, $\mu(Q) = \mathbb{E}_{Y \sim Q}[k(\cdot, Y)]$ であり、 \otimes はテンソル積を表しており、任意の $f \in H(k)$ に対し、 $f^{\otimes 2} = f \otimes f = \langle f, \cdot \rangle_{H(k)} f$ である。

3 検定統計量の構築

この $k(\cdot, X)$ と $k(\cdot, Y)$ の $f \in H(k)$ への射影のモーメント差

$$\sup_{\|f\|_{H(k)}=1} |\langle f, \mu(P) - \mu(Q) \rangle_{H(k)}| = \|\mu(P) - \mu(Q)\|_{H(k)}$$

により、2 つの分布の違いを測るのが、Maximum Mean Discrepancy (MMD) と呼ばれるものである。同様の考え方により、 $(k(\cdot, X) - \mu(P))^{\otimes 2}$ と $(k(\cdot, Y) - \mu(Q))^{\otimes 2}$ の $A \in H(k)^{\otimes 2}$ への射影のモーメント差

$$\sup_{\|A\|_{H(k)^{\otimes 2}}=1} |\langle A, \Sigma_k(P) - \Sigma_k(Q) \rangle_{H(k)^{\otimes 2}}| = \|\Sigma_k(P) - \Sigma_k(Q)\|_{H(k)^{\otimes 2}}$$

により2つの分布の違いを測る。これは、MMDのようなある種の分布の違いを測っている。この違い $\|\Sigma_k(P) - \Sigma_k(Q)\|_{H(k)^{\otimes 2}}^2$ は

$$\hat{T}^2 = \left\| \hat{\Sigma}_k(P) - \hat{\Sigma}_k(Q) \right\|_{H(k)^{\otimes 2}}^2$$

によって推定することができる。ただし、

$$\begin{aligned} \hat{\Sigma}_k(P) &= \frac{1}{n} \sum_{i=1}^n (k(\cdot, X_i) - \hat{\mu}(P))^{\otimes 2}, & \hat{\mu}(P) &= \frac{1}{n} \sum_{i=1}^n k(\cdot, X_i), \\ \hat{\Sigma}_k(Q) &= \frac{1}{m} \sum_{i=1}^m (k(\cdot, Y_i) - \hat{\mu}(Q))^{\otimes 2}, & \hat{\mu}(Q) &= \frac{1}{m} \sum_{i=1}^m k(\cdot, Y_i) \end{aligned}$$

である。

本発表では、この検定統計量 \hat{T}^2 の漸近挙動について報告を行った。特に、帰無仮説 $H_0 : P = Q$ のもとで、退化 V 統計量の結果に帰着させることで、 \hat{T}^2 の漸近分布が、独立な自由度1の χ^2 分布の重み付きつき無限和の形で得られること ([2] 参照)、その重みをデータに基づいて推定する方法について報告を行った。また、対立仮説 $H_1 : P \neq Q$ のもとでの \hat{T}^2 の漸近分布が平均0の正規分布になっていることについて報告を行った。

参考文献

- [1] Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. and Smola, A. (2007). A kernel method for the two sample problem. *Advances in Neural Information Processing Systems*, **19** of MIT Press, Cambridge.
- [2] Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley, New York.

非負データに対する再帰的な密度推定について

柿沢佳秀 (北大経済)

1. はじめに

位置尺度型のカーネル推定量 (Rosenblatt(1956), Parzen(1962)) は、畳み込み近似が背後にあって、 $\text{supp}(f) = \mathbb{R}$ のときに望ましい性質を持つことが知られているが、非負または区間データの密度推定においては、境界付近で一致性がない。そのために境界バイアス問題の対処も議論されている (リノーマライゼーション法, リフレクション法, 一般化ジャックナイフ法は Jones(1993) を参照のこと; なお, このカテゴリーに, 局所多項式法による密度推定法も含まれる)。Chen(1999,2000) 以降, 非対称カーネル密度推定法が境界バイアス問題の対処の1つとして脚光を浴び, ベータ, ガンマ, 逆ガウス (IG), 相反逆ガウス (RIG), バーンバウム・サンダース (BS), 対数正規, 逆ガンマ等個別に研究が進められ, 現在のところ, 次の (1)–(3) のように分類される (積カーネルを使えば, 多次元データも扱えるので, 簡単のため1次元データとする):

- (1) Amoroso カーネル密度推定量族 (Igarashi and Kakizawa(2018)) .
- (2) IG と RIG との ϵ -混合 (MIG) カーネル密度推定量族 (Igarashi and Kakizawa(2014)) , 及び, 加重対数正規カーネル密度推定量族 (Igarashi (2016)) . また, これらは $N(0, 1)$ ベースであったから, 対称分布ベースの q -MIG カーネル密度推定量族 (Kakizawa(2018)) へも拡張されている ($q \rightarrow 0$ としたサブクラスが, 対数対称カーネル密度推定量族である) .
- (2') (2) の対称分布ベースの q -MIG カーネル密度推定量族の類似だが, 歪分布ベースの q -BS カーネル密度推定量族 (柿沢; 2017 年度統計関連学会連合大会) へも拡張されている; なお, 対称分布を外した為, IG と RIG 相当の拡張は定義できない .
- (3) インデックスが $\nu \neq \pm 1/2$ の一般化逆ガウス (GIG) カーネル密度推定量 (なお, $\nu = -1/2$ (IG), $1/2$ (RIG) を (2) に分類した; Igarashi and Kakizawa(2014) は GIG の代わりに別名 “変形ベッセル” を採用) .

さて, 非対称カーネル密度推定量とは, 無作為標本 $X_1, \dots, X_n \sim \text{iid } f$ が与えられ, 付随する平滑化パラメータを $b_n \searrow 0$ ($n \rightarrow \infty$) とおくとき, 点 $x \in \text{supp}(f)$ で

$$\hat{f}_{b_n}(x) = \frac{1}{n} \sum_{i=1}^n k(X_i; b_n, x)$$

のような形式で定義されるが, これは “非再帰的” である。すなわち, 仮に追加データがあった場合に, 計算は最初からやり直さねばいけない。他方, 再帰的ならば, 計算結果を更新するので, 追加データを効率的に扱えるメリットがある。この種の話は, Rosenblatt–Parzen カーネル法が流行した頃に Yamato(1971) を初めとして, しばしば議論されていた (磯貝 (1993) を参照) が, 当時, Chen(1999) 以降の密度推定量は対象となっておらず, 本報告では, 再帰的な非対称カーネル法を考察した。

2. 再帰的な非対称カーネル密度推定量

正数列 γ_n (ステップサイズと呼ぶ) を選び, $\hat{f}_0(x) = 0$ として, $i = 1, 2, \dots, n$ に対し $\hat{f}_i(x) = (1 - \gamma_i)\hat{f}_{i-1}(x) + \gamma_i k(X_i; b_i, x)$ のように再帰的に更新する。任意の g について

$$\hat{f}_n(x) - g = \prod_n \sum_{i=1}^n \prod_{i=1}^{-1} \gamma_i \{k(X_i; b_i, x) - g\} + \prod_n \{\hat{f}_0(x) - g\}$$

を示すのは容易である．ここに， $\Pi_n = \prod_{i=1}^n (1 - \gamma_i)$ ．

以下，Galambos and Seneta(1973) の定義に従い，正数列 h_n について $(h_n) \in GS(d)$ であるとは， $\lim_{n \rightarrow \infty} \{n(1 - h_{n-1}/h_n)\} = d \in \mathbb{R}$ を満たすこととする (Bojanic and Seneta(1973) も参照) ．

主結果 $[\text{supp}(f) = [0, \infty)$ を念頭にしたが， $[0, 1]$ でも同様な結果になる] $(\gamma_n) \in GS(-\gamma)$ ， $(b_n) \in GS(-c)$ ， $\lim_{n \rightarrow \infty} (n\gamma_n)^{-1} = \xi$ を仮定する．ただし， γ, c, ξ について

(A) $x > 0$ ならば， $0 < \gamma \leq 1$ ， $0 < c < 2\gamma$ ， $0 \leq \xi < \max\{1/c, 2/(\gamma - c/2)\}$

(B) $x = 0$ ならば， $0 < \gamma \leq 1$ ， $0 < c < \gamma$ ， $0 \leq \xi < \max\{1/c, 2/(\gamma - c)\}$

とする．

(i) (1) $x > 0$ に対して $0 < c \leq (2/5)\gamma$ ならば $\text{Bias}[\hat{f}_n(x)] \approx \frac{b_n B(x)}{1 - c\xi}$ ， $2\gamma > c > (2/5)\gamma$ ならば $\text{Bias}[\hat{f}_n(x)] = o((\gamma_n b_n^{-1/2})^{1/2})$ ．

(2) $0 < c \leq (1/3)\gamma$ ならば $\text{Bias}[\hat{f}_n(0)] \approx \frac{b_n \delta}{1 - c\xi}$ ， $\gamma > c > (1/3)\gamma$ ならば $\text{Bias}[\hat{f}_n(0)] = o((\gamma_n b_n^{-1})^{1/2})$ ．

(ii) (1) $x > 0$ に対して $2\gamma > c \geq (2/5)\gamma$ ならば $V[\hat{f}_n(x)] \approx \frac{\gamma_n b_n^{-1/2} V(x)}{2 - (\gamma - c/2)\xi}$ ， $0 < c < (2/5)\gamma$ ならば $V[\hat{f}_n(x)] = o(b_n^2)$ ．

(2) $\gamma > c \geq (1/3)\gamma$ ならば $V[\hat{f}_n(0)] \approx \frac{\gamma_n b_n^{-1} v}{2 - (\gamma - c)\xi}$ ， $0 < c < (1/3)\gamma$ ならば $V[\hat{f}_n(0)] = o(b_n^2)$ ．

(iii)(1) $x > 0$ に対して $AMSE[\hat{f}_n(x)] \in \begin{cases} GS(-2c), & \text{if } 0 < c < (2/5)\gamma, \\ GS(-(4/5)\gamma), & \text{if } c = (2/5)\gamma, \\ GS(-(\gamma - c/2)), & \text{if } 2\gamma > c > (2/5)\gamma \end{cases}$

すなわち， $AMSE[\hat{f}_n(x)]$ を最小化するならば， $c = (2/5)\gamma$ でなければいけない；このとき， $AMSE[\hat{f}_n(x)] \in GS(-(4/5)\gamma)$ ， $\gamma \in (0, 1]$ なので， $\gamma = 1$ でなければいけない．また， $\gamma = 1$ ， $c = (2/5)\gamma = 2/5$ のとき

$$\begin{aligned} AMSE[\hat{f}_n(x)] &= \frac{b_n^2 B^2(x)}{\{1 - (2/5)\xi\}^2} + \frac{\gamma_n b_n^{-1/2} V(x)}{2 - (4/5)\xi} \\ &\geq \frac{5}{4^{4/5}} \left[\frac{\gamma_n^2}{4\{1 - (2/5)\xi\}^3} \right]^{2/5} V^{4/5}(x) \{B^2(x)\}^{1/5}. \end{aligned}$$

なお， $\gamma_n = 1/(\xi n)$ ($\xi \in (0, 5/2)$) とすれば，

$$\frac{\gamma_n^2}{\{1 - (2/5)\xi\}^3} = \frac{n^{-2}}{\xi^2 \{1 - (2/5)\xi\}^3} \geq \frac{n^{-2}}{(3/5)^3} \quad (\text{等号は } \xi = 1 \text{ に限る}).$$

(2) $AMSE[\hat{f}_n(0)]$ を最小化する場合，同様にして $\gamma = 1$ ， $c = (1/3)\gamma = 1/3$ となり， $\gamma_n = 1/(\xi n)$ ($\xi \in (0, 3)$) とすれば， $\xi = 1$ に至る．

(iv) AMISE については (iii-1) の積分版に相当の結果になる ($B^2(x), V(x)$ を，それぞれ $I^{B^2} = \int_0^\infty B^2(x) dx$ ， $I^V = \int_0^\infty V(x) dx$ とせよ) ．

(v) 漸近正規性，強一貫性も成立．

注意．AMSE あるいは AMISE の最小化で考えると，再帰的推定量は非再帰的推定量に劣る (上記のような最小値を，非再帰版と比較すれば分かる) ．他方，過小平滑化をすれば，分散 (分散積分) の主要項の最小化での比較になり，このときは再帰的推定量が良い．

ノンパラメトリックな極値解析法の漸近理論

鹿児島大学大学院理工学研究科 吉田拓真

1 はじめに

降水量, 人口推移, 金融リスクに関するデータでは, 平均よりも最大値や最小値, もしくはそれに近い分位点に興味がある. 別の言い方をすれば, 分布の中心やピークではなく, 裾に興味がある. 分布の裾は大きくわけて3つのタイプに分けられる. ひとつは, 正規分布などに代表される裾確率が標準的に小さい分布 (light tail), もうひとつは一様分布, ベータ分布に代表される裾が有限である分布 (short tail), そして t -分布などいわゆる正規分布よりも重い裾を持つ分布 (heavy tail) である. 本講演では, heavy tail を持つ分布の裾の情報を推測するための方法と理論について述べた.

2 モデル

一般に, $Y \sim F$ が heavy tail するとは, y が十分大きいところでは

$$P(Y > y) = 1 - F(y) = y^{-1/\gamma} L(y) \quad (1)$$

という構造を持つことをいう. ただし, $\gamma > 0$ は Extreme value index (EVI) と呼ばれるパラメータで, $L(y)$ は $L(y) \rightarrow c > 0, y \rightarrow \infty$ を満たすノンパラメトリック関数である. この分布は Pareto-type-tailed model と呼ばれており, t -分布, Burr 分布など多くの裾の重い分布を含む. 実際, (1) は分布の裾が次数 $1/\gamma$ の多項式オーダーで 0 に収束していくことのみを仮定しているにすぎない. このモデルを考える上で最も重要なのは, γ を推定することといえる. EVI は分布の裾の重さを表すパラメータで, 例えば t -分布の自由度を ν とすると, $\nu = 1/\gamma$ と対応する. すると, $\gamma = 1$ でコーシー分布の裾と同等であり, $\gamma = 0.5$ で分散を持たないほど重い分布であることがわかる.

さて, 仮に γ の推定量 $\hat{\gamma}$ を得たとすると, $\hat{\gamma} - \gamma$ の収束レートが気になる. しかし, それを数学的に示そうとすると結局 $L(y) \rightarrow c$ の収束オーダーが必要となってくる. よって, $\hat{\gamma}$ の数学的性質解明のため, L はしばしば, パラメータ $A > 0, \rho < 0$ とある関数 u を用いて

$$L(y) = 1 + u(y), \quad |u(y)| \leq Ay^{\rho/\gamma}$$

を満たすと仮定される. 特に, ρ は 2 次オーダーと呼ばれ, $\hat{\gamma} - \gamma$ の収束レートを決定する重要なパラメータである. この仮定をしてなお多くの分布が (1) に含まれることが知られている (see, Hall and Welsh 1984). 実際 $\hat{\gamma}$ を推定する方法としては最尤法があげられるが, ((1) から導かれる密度関数を $f(y|\gamma)$ と書く) そのまま $\sum_{i=1}^n \log f(y_i|\gamma)$ を最大化する γ を求めても良い推定量とはならない. なぜならば, (1) はあくまで裾の情報であって, 平均周りに分布するデータなどは (1) に従っている保証はないからである. ではどうするかというと, ある閾値 w を超えるデータのみで推定を行う. このとき, 考える分布は

$$P\left(\frac{Y}{w} > y | Y > w\right) = 1 - F_w(y) = y^{-1/\gamma} \tilde{L}_w(y)$$

と, 再び Pareto-type モデルになるため, このモデルに基づく最尤法などで γ を推定する. ただし, F_w は $Y > w$ の下での Y/w の分布関数, \tilde{L} は $\tilde{L}_w(y) \rightarrow \tilde{c} > 0 (y \rightarrow \infty)$ を満たす関数であ

る. 最尤法は $\sum_{i=1}^n f_w(y_i/w|\gamma)I(y_i > w)$ を最大化させる. ここで, f_w は F_w に対応する密度関数であり, I は指示関数である. このように, 閾値を超えるデータのみを用いて推定を行う方法は Peak-over-threshold(POT) 法と呼ばれ, EVI の一般的な推定方法となっている. 閾値 w の選び方に関する研究論文も数多く発表されている.

説明変数 x を用いた回帰分析では, 基本的には上記のモデルすべてが x の関数となり, Pareto-type-tailed regression model

$$P(Y > y|X = x) = 1 - F(y|x) = y^{-1/\gamma(x)}\{c(x) + u(y|x)\}, \quad |u(y|x)| \leq A(x)|y|^{\rho(x)/\gamma(x)} \quad (2)$$

というモデルに基づいて $\gamma(x)$ の推定を行う. ここで, 任意の $x \in \mathcal{X}$ に対して, $\gamma(x) > 0$, $A(x) > 0$, $\rho(x) < 0$ である. この回帰分析の枠組みでは, $\gamma(x)$ のパラメトリック推定・ノンパラメトリック推定が主な目的となる. 一方で, EVI が x に依存しない: $\gamma = \gamma(x)$ という仮定や仮説検定などは多くの研究者が議論している. EVI の推定に POT 法を用いるときは, 閾値も x に依存する ($w(x)$ と書く). そして, $\hat{\gamma}(x) - \gamma(x)$ の $n \rightarrow \infty$ の下での収束レートを調べるのが重要となる. ここで, n はデータ数である. 関連する先行研究は多いが, そこでは基本的に閾値関数 $w(x)$ は固定した下で行っている. しかし, 閾値関数 $w(x)$ は極値に位置するデータをピックアップするためにあることから, データ数が増えるとそれに対応して変化するのが自然である. もし, $n \rightarrow \infty$ で, $w(x)$ を固定したままだと, 相対的に極値ではない, 言い換えると, (2) に従っていないデータを推定に含むことになる. つまり, $n \rightarrow \infty$ の下では, $w(x) = w_n(x) \rightarrow \infty$ と仮定するのが自然である. 実際に, シミュレーションにおいて何らかの閾値選択 (クロスバリデーションや QQ-plot を用いた方法など) を行うとそのような傾向が見られる. 上記で述べた EVI 推定の先行研究などは Dey and Yan (2016) に詳しくまとめられている.

3 主結果

本研究では回帰分析の文脈で関数 $\gamma(x)$ をノンパラメトリック法で推定し, $\sup_{x \in \mathcal{X}} |\hat{\gamma}(x) - \gamma(x)|$ の収束レートを導く. 真の $\gamma(x)$ を m 次微分が連続な関数とし, $\hat{\gamma}(x)$ を最尤法をベースにしたノンパラメトリック推定量とする. このとき, 適切な条件の下で

$$\sup_{x \in \mathcal{X}} |\hat{\gamma}(x) - \gamma(x)| = O_P \left(\left(\frac{\log \{ \sup_{x \in \mathcal{X}} |w_n(x)^{-1/\gamma(x)} |n \} }{\sup_{x \in \mathcal{X}} |w_n(x)^{-1/\gamma(x)} |n} \right)^{m/(2m+1)} \right) + O \left(\sup_{x \in \mathcal{X}} |w_n(x)^{\rho(x)/\gamma(x)} | \right)$$

が成り立つ. 第1項はノンパラメトリック法によるバイアス-分散で, 第2項は(2)のモデルバイアスである. 特に, 閾値関数として, $X = x$ を与えた下での Y の分布関数 $F(\cdot|x)$ に対応する分位点 $w_n(x) = F^{-1}(\tau_n|x)$, $\tau_n \rightarrow 1(n \rightarrow \infty)$ とすると,

$$\sup_{x \in \mathcal{X}} |\hat{\gamma}(x) - \gamma(x)| = O_P \left(\left\{ \frac{\log \{ (1 - \tau_n)n \} }{(1 - \tau_n)n} \right\}^{\frac{m}{2m+1}} \right) + O \left((1 - \tau_n)^{-\rho} \right) \quad (3)$$

となることがわかった. ただし, $\rho = \sup_{x \in \mathcal{X}} \rho(x)$ である. (3) はノンパラメトリック分位点回帰をベースにした EVI 推定量の収束レート (Yoshida 2019) と一致していること, 回帰ではない場合の自然な展開になっていることを報告した. また, 分位点回帰への応用について報告した.

References

- 1.Dey,D.K. and Yan,J. (2016). *Extreme value modeling and risk analysis: Methods and applications*. Chapman & Hall/CRC.
- 2.Hall, P. and Welsh, A.H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079–1084.
- 3.Yoshida,T. (2019). Simultaneous confidence bands for extremal quantile regression with splines. *Extremes*. Accepted. DOI:10.1007/s10687-019-00360-4

ノンパラメトリック重回帰分析における Lasso 型推定の改良

島根大 自然科学研究科 松島 佑樹
千葉大 理学研究院 内藤 貫太

1 はじめに

ノンパラメトリック重回帰分析において、局所線形回帰の手法と Lasso 型推定を組み合わせた方法について考える。この方法は Bertin and Lecué (2008) ですでに研究されているが、先行研究では推定したい回帰関数の値にも罰則を付けた上で推定量を計算しているため推定精度が悪くなる問題が挙げられる。この問題を改善するため、回帰関数に関する罰則項を取り除いた推定手法について考え、先行研究で用いられた推定量と我々の提案手法による推定量、これら2つを比較した結果を報告する。

2 設定

目的変数 $Y \in \mathbb{R}$ 、 d 次元の説明変数 $X \in \mathbb{R}^d$ のデータ $(Y_1, X_1), \dots, (Y_n, X_n)$ にノンパラメトリック重回帰モデル

$$Y_i = f(X_i) + e_i, \quad i = 1, \dots, n$$

を想定し、回帰関数 f の推定を考える。ここで、 $e_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ は誤差であり、 $n \ll d$ を仮定しておく。 f の推定を行うため、カーネル関数を用いた局所重み付き線形回帰の手法を用いる。推定を行う点を $x \in \mathbb{R}^d$ とし、モデル式を線形式

$$Z = A\theta^* + \varepsilon \tag{1}$$

と変形する。ここで、 $Z = (Z_1 \cdots Z_n)^T$ 、 $A = (\tilde{A}_1^T \cdots \tilde{A}_n^T)^T = (A_0 \cdots A_d)$ 、 $\varepsilon = (\varepsilon_1 \cdots \varepsilon_n)^T$ であり、 $i = 1, \dots, n$ に対して、

$$Z_i = \alpha_i Y_i, \quad \alpha_i = \frac{1}{\sqrt{nh^d}} K\left(\frac{X_i - x}{h}\right), \quad \tilde{A}_i = \alpha_i \left(1 \quad \left(\frac{X_i - x}{h}\right)^T \right), \quad \varepsilon_i = \alpha_i e_i + \alpha_i f(X_i) - \tilde{A}_i^T \theta^*$$

である。また K はカーネル関数で原点对称な d 次元密度関数、 $h > 0$ はバンド幅である。 $\theta^* = (\theta_0^* \theta_1^* \cdots \theta_d^*)^T = (f(x) h \partial_1 f(x) \cdots h \partial_d f(x))^T$ がこのモデルでの回帰パラメータベクトルとなり、 ∂_i は第 i 変数による偏微分作用素を表す。(1)における回帰係数ベクトル θ^* の推定を行う上で、特に $f(x)$ の推定を考えているため、 $\theta_0^* = f(x)$ の推定が重要となることに注意しておく。

3 Lasso 型推定

Bertin and Lecué (2008) では、 $f(x)$ において本質的となる変数の選択のために Lasso を使った推定

$$\min_{\theta \in \mathbb{R}^{d+1}} \{ \|Z - A\theta\|_2^2 + 2\lambda \|\theta\|_1 \}, \theta = (\theta_0 \ \theta_1 \ \cdots \ \theta_d)^T \quad (2)$$

が用いられ、変数選択の一致性が示されている。しかし、この手法では、 θ_0 にも罰則を付けているため、 $f(x)$ の値にも制約を課した上で推定量が求められていることから、 $\theta_0^* = f(x)$ の推定精度が悪くなる可能性が考えられる。この問題を改善するため、 θ_0 には罰則を付けない推定方法

$$\min_{\theta \in \mathbb{R}^{d+1}} \{ \|Z - A\theta\|_2^2 + 2\lambda \|\theta_{-0}\|_1 \} \quad (3)$$

を提案する。ここで $\theta_{-0} \in \mathbb{R}^d$ は θ から θ_0 を取り除いた部分ベクトル $\theta_{-0} = (\theta_1 \ \cdots \ \theta_d)^T$ であり、 λ は正則化パラメータである。

(2), (3) の解の同値条件をそれぞれ導き比較することで、アクティブな変数とノンアクティブな変数に関しての条件が変わらないことがわかった。この事実から、提案手法 (3) から得られる推定量においても (2) と同様に変数選択の一致性が導かれることを報告した。

また、理論的比較をするため、しかるべき条件を満たした説明変数 X に対し、それぞれ (2) と (3) のユニークな解として求められる 2 つの理想的推定量を紹介した。この 2 つの理想的推定量は θ_0^* を含んでいたため、実装することはできず、その意味でこれらを「理想的」と呼んだ。 $n \rightarrow \infty$, $h \rightarrow 0$, $nh^{d+4} \rightarrow \infty$ における 2 つの理想的推定量の漸近挙動を見ることで、(2) と (3) の理論的比較を行った。その結果、(3) から得られる推定量のバイアスは $O(h\lambda) = O(h^2)$ であったが、(2) から得られる推定量のバイアスは $O(\lambda) = O(h)$ となり、このことから、提案手法が推定量のバイアスを改善していることを報告した。また分散は Ruppert and Wand (1994) の結果と同じになることを述べた。

研究集会においてはこのような理論的結果の詳細および数値実験の結果について報告を行った。

参考文献

- [1] Bertin, K. and Lecué, G. (2008). Selection of variables and dimension reduction in high-dimensional non-parametric regression. *Electronic Journal of Statistics*, 2, 1224–1241.
- [2] Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, vol. 22, Number 3, 1346–1370.

Semiparametric regression with local divergence

Kanta NAITO

Department of Mathematics and informatics, Chiba University, Japan

1 Setting

Let $(Y_1, \mathbf{X}_1), \dots, (Y_n, \mathbf{X}_n) \sim_{i.i.d.} f(y, \mathbf{x}) = p(y|\mathbf{x})q(\mathbf{x})$, where $(Y_i, \mathbf{X}_i) \in \mathbb{R} \times \mathbb{R}^d$, f is the joint density of (Y, \mathbf{X}) , p is the conditional density of Y given $\mathbf{X} = \mathbf{x}$, and q is the density of \mathbf{X} .

Let $\mathbf{t} \in \mathbb{R}^d$ be a target point at which we want to estimate the value of regression function $\mu(\mathbf{t}) = E[Y|\mathbf{X} = \mathbf{t}]$. Our parametric model for μ is having the form

$$m(\mathbf{x}, \boldsymbol{\theta}) = G^{-1}(\boldsymbol{\theta}^T \tilde{\mathbf{x}}),$$

where $\tilde{\mathbf{x}}^T = [1 \ \mathbf{x}^T] \in \mathbb{R}^{d+1}$ is the vector of explanatory variables, $\boldsymbol{\theta} = [\theta_0 \ \theta_1 \ \dots \ \theta_d]^T \in \Theta \subset \mathbb{R}^{d+1}$ is the parameter vector, and G is the link function.

2 The Bregman divergence

Now we fix a strictly convex function U . Then the discrepancy between $\mu(\cdot)$ and its parametric model $m(\cdot, \boldsymbol{\theta}) = m_{\boldsymbol{\theta}}(\cdot)$ can be measured by the Bregman divergence defined as

$$\begin{aligned} & D_{U^*}(u(m_{\boldsymbol{\theta}}), u(\mu)) \\ &= \int_{\mathbb{R}^d} [U^*(u(m(\mathbf{x}, \boldsymbol{\theta}))) - U^*(u(\mu(\mathbf{x}))) - \mu(\mathbf{x})\{u(m(\mathbf{x}, \boldsymbol{\theta})) - u(\mu(\mathbf{x}))\}] q(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (1)$$

$$\begin{aligned} &= \int_{\mathbb{R} \times \mathbb{R}^d} [U^*(u(m(\mathbf{x}, \boldsymbol{\theta}))) - y \cdot u(m(\mathbf{x}, \boldsymbol{\theta}))] f(y, \mathbf{x}) dy d\mathbf{x} \\ &\quad + \int_{\mathbb{R}^d} [-U^*(u(\mu(\mathbf{x}))) + \mu(\mathbf{x})u(\mu(\mathbf{x}))] q(\mathbf{x}) d\mathbf{x}, \end{aligned} \quad (2)$$

where $u = U'$: the derivative of U , U^* is the convex conjugate of U : $U^*(s) = \sup_{z \in \mathbb{R}} \{zs - U(z)\}$, and we have used a fact that $(U^*)' = u^{-1} = (U')^{-1}$ in (1).

The usual parametric regression can be carried out by using a certain estimator $\hat{\boldsymbol{\theta}}$ of the true value of $\boldsymbol{\theta}$. Necessary tools for this estimation scheme are

$$\rho(y, \mathbf{x}, \boldsymbol{\theta}) = U^*(u(m(\mathbf{x}, \boldsymbol{\theta}))) - y \cdot u(m(\mathbf{x}, \boldsymbol{\theta})), \quad (3)$$

$$\psi(y, \mathbf{x}, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \rho(y, \mathbf{x}, \boldsymbol{\theta}) = \{m(\mathbf{x}, \boldsymbol{\theta}) - y\} \frac{u'(m(\mathbf{x}, \boldsymbol{\theta}))}{G'(m(\mathbf{x}, \boldsymbol{\theta}))} \tilde{\mathbf{x}}, \quad (4)$$

by which the estimator $\hat{\boldsymbol{\theta}}$ and the true value $\boldsymbol{\theta}_*$ of $\boldsymbol{\theta}$ can be defined as

$$\boldsymbol{\theta}_* = \arg \min_{\boldsymbol{\theta} \in \Theta} \int_{\mathbb{R} \times \mathbb{R}^d} \rho(y, \mathbf{x}, \boldsymbol{\theta}) dF(y, \mathbf{x}), \quad (5)$$

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \int_{\mathbb{R} \times \mathbb{R}^d} \rho(y, \mathbf{x}, \boldsymbol{\theta}) dF_n(y, \mathbf{x}), \quad (6)$$

where F_n is the empirical distribution function based on $(Y_1, \mathbf{X}_1), \dots, (Y_n, \mathbf{X}_n)$, F is the cumulative distribution function with its density f and $\mathbf{0}_{d+1}$ is the zero vector in \mathbb{R}^{d+1} .

Note that $\boldsymbol{\theta}_*$ in (5) is the minimizer of (2), and the minimizer of the empirical version of (2) is nothing other than $\hat{\boldsymbol{\theta}}$ in (6). The regression function estimator can be obtained by plugging $\hat{\boldsymbol{\theta}}$ into $\boldsymbol{\theta}$ in $m(\cdot, \boldsymbol{\theta})$:

$$\hat{\mu}_G(\mathbf{x}) = m(\mathbf{x}, \hat{\boldsymbol{\theta}}). \quad (7)$$

3 The local Bregman divergence

We organize a scheme of estimation of $\boldsymbol{\theta}$ depending on \mathbf{t} locally. Necessary functions are listed as follows:

$$\rho(\mathbf{t}, y, \mathbf{x}, \boldsymbol{\theta}) = K\left(\frac{\mathbf{x} - \mathbf{t}}{h}\right) \rho(y, \mathbf{x}, \boldsymbol{\theta}), \quad (8)$$

$$\psi(\mathbf{t}, y, \mathbf{x}, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \rho(\mathbf{t}, y, \mathbf{x}, \boldsymbol{\theta}). \quad (9)$$

Here $K(z)$ is a smooth unimodal integrable function symmetric around $z = 0$ satisfying $K(0) = 1$, and $h > 0$ is the scalar bandwidth which controls the degree of localization.

We note that (8) and (9) are localized version of (3) and (4) respectively, with the use of the kernel K . Using these functions, we define the true parameter $\boldsymbol{\theta}_*(\mathbf{t})$ and its estimator $\hat{\boldsymbol{\theta}}(\mathbf{t})$ as follows:

$$\begin{aligned} \boldsymbol{\theta}_*(\mathbf{t}) &= \arg \min_{\boldsymbol{\theta} \in \Theta} \int_{\mathbb{R} \times \mathbb{R}^d} \rho(\mathbf{t}, y, \mathbf{x}, \boldsymbol{\theta}) dF(y, \mathbf{x}), \\ \hat{\boldsymbol{\theta}}(\mathbf{t}) &= \arg \min_{\boldsymbol{\theta} \in \Theta} \int_{\mathbb{R} \times \mathbb{R}^d} \rho(\mathbf{t}, y, \mathbf{x}, \boldsymbol{\theta}) dF_n(y, \mathbf{x}). \end{aligned}$$

This local estimator $\hat{\boldsymbol{\theta}}(\mathbf{t})$ of $\boldsymbol{\theta}_*(\mathbf{t})$ also suggests us to make a regression estimator defined as

$$\hat{\mu}_L(\mathbf{x}) = m(\mathbf{x}, \hat{\boldsymbol{\theta}}(\mathbf{x})), \quad (10)$$

which we call the *local estimator* of $\mu(\mathbf{x})$, because the involved estimator of parameter is determined locally. Since $\hat{\boldsymbol{\theta}}(\mathbf{x})$ can vary depending on \mathbf{x} , $\hat{\mu}_L$ would be expected to be more flexible than $\hat{\mu}_G$. On the other hand, we call the estimator $\hat{\mu}_G(\mathbf{x})$ in (7) the *global estimator* of $\mu(\mathbf{x})$.

4 Problem

Our interest is to investigate whether the local estimator $\hat{\mu}_L$ in (10) performs better than the global estimator $\hat{\mu}_G$ in (7). By its construction and the shape of K , it is seen intuitively that $\|\hat{\boldsymbol{\theta}}(\mathbf{t}) - \hat{\boldsymbol{\theta}}\|$ converges to zero in some sense, as $n \rightarrow \infty$ and $h \rightarrow \infty$, hence $\hat{\mu}_L \approx \hat{\mu}_G$. Even in such a situation, we think it is still not obvious that the global estimator $\hat{\mu}_G$ certainly defeats the local estimator $\hat{\mu}_L$.

In the symposium, several theoretical results which claim superiority of $\hat{\mu}_L$ over $\hat{\mu}_G$ were reported.

市場創造時点の特定に向けた文献調査：KIP を前提として¹

加藤淳一

(久留米大学 商学部)

1. 目的

本論文の目的は、先行研究のレビューにより、市場創造時点の特定方法の方向性を明らかにすることである。この目的に向けて、次のように議論を進める。2章で、議論に先立ち、市場創造の意味を共有する。3章で、2章で確認した市場創造について、これまでの研究がどのように扱ってきているのかを吟味する。4章で、本論文の結論をまとめる。

2. 市場創造とは

本論文は、市場創造についての事前の理解を必要とする。そこで、以下の議論に先立って、市場創造について共有しておく。まず、「市場」創造を扱うのであるから、本論文の分析単位は市場となる。消費者個人ではなく、それらを集計した市場に焦点を当てる。

次に、市場「創造」を扱う。ここで創造とは、評価軸（統計的には主成分軸）の移り変わりである。この「創造」について、既存のマーケティングとの対比を通じて述べる。マーケティングでは、新しい市場（製品）の創造をその一生に例えて、製品ライフサイクルという時系列のグラフによりとらえる。この製品ライフサイクルにおいて、市場創造は売上の水準の上昇（縦軸の高さ）により捉えられている。

このようにマーケティングでは、これまで市場創造を売り上げの急速な上昇としている。これに対して、本論文の定義は、評価（主成分）軸の継時的な移り変わりである。観測値の水準（縦軸で測ったときの高さ）ではなく、縦軸あるいは横軸という「軸」である。以上で、市場創造とは何かを説明した

3. 文献レビュー

時系列分析の分野において、構造変化(structure change)と呼ばれる研究が蓄積されている。それらは、とりわけ単位根過程との関連で例えば、黒住(2007)、市橋・金子(2009)、Byrne and Perman (2006)において議論されている。これら先行研究は、大まかにまとめれば、ある時点での傾きや切片の変化として構造変化をとらえようとしている。よって、これらの研究で想定していた構造変化は、本論文の問題意識（市場創造）とは異なっている。

そこで、より本論文の問題意識（市場創造）と近い研究を探索する。すると、近年になり状態空間モデルが研究されている。状態空間モデルは、例えば、近年マーケティング分野を

¹本論文の成果の一部は、平成31年（令和元年）度久留米大学の科研費申請・採択に係る個人研究費支援「観光目的地への消費者ニーズの経時的な変化点の特定に向けた文献調査」として研究費の支援を受けました。ここに記して感謝を表します。

含めて次のような様々な応用研究がなされている。山口・土屋・樋口(2004)、佐藤・北川(2002)、伊東(2016)、本橋・磯崎・長尾・樋口(2012)、青柳・佐藤(2015)、奥野・中村(2016)、佐藤・樋口(2008)、本橋・樋口(2013)などである。

ここで状態方程式が、(評価)軸の変化に最も関連しそうである。だが、状態方程式は、状態の数値の変化としてモデル構築しなければならない。このように理解すると、本論文の市場創造とは異なる。そこで、時系列データへのもう1つのアプローチとして回帰分析、とりわけ SCT(Stepwise Chow Test)に注目する。

さて、時系列データの全期間の端から Chow test を繰り返す。すると、説明変数で捉えられていない時点が分かれば、その時点が評価軸の移り変わった時点と考えられる。このようにして考えれば、SCTにより市場創造時点の特定できそうである。SCTは市場創造時点の特定方法の1つの選択肢として使用できるのではないかと考えられる。

以上で、時系列データを巡る既存研究を吟味し、現在のところ問題は残されているものの、評価軸の変化を捉える1つの方法として SCT の使用を提案した。最後に、本論文の目的に戻り、結論を整理する。

4. 結論

本論文の目的は、先行研究のレビューにより、市場創造時点の特定方法の方向性を明らかにすることであった。本論文の議論を通じて、これまでの時系列分析の構造変化の研究や、状態空間モデルの状態方程式は、どちらも軸を固定してレベルの変化を議論しており、本論文が想定した市場創造とは異なっていた。

そのような中でも、SCTの変化点は、別の軸を必要としている時点と解釈すれば、市場創造時点の特定にも結びつくと考えられた。まだ不十分なところも残されているものの、本論文のレビューを通じて市場創造時点の特定方法に一定の方向性を示せた。今後、更にこの解釈の適切さを巡って考察を深めたい。

参考文献

- 青柳憲治・佐藤忠彦(2015)、「3階層多変量状態空間モデリングによる動的市場反応形成メカニズムの解明」、『日本オペレーションズ・リサーチ学会和文論文誌』、Vol. 58、pp. 70-100。
- 市橋勝・金子慎治(2009)、「長期データ系列における構造変化の検定法比較分析」、『Discussion Paper IDEC, Hiroshima University』、pp.1-41。
- 伊東宏樹(2016)、「状態空間モデルの実行方法と実行環境の比較」、『日本生態学会誌』、Vol. 66、p. 361-374。
- Joseph P. Byrne & Roger Perman.(2006). "Unit Roots and Structural Breaks: A Survey of the Literature," Working Papers, pp.1-23.
- 黒住英司(2007)、「経済時系列分析と単位根検定：これまでの発展と今後の展望」、『Discussion Paper Series』、No.228、pp. 1-26。
- 本橋永至・樋口知之(2013)、「市場構造の変化を考慮したブランド選択モデルによる購買履歴データの解析」、『マーケティング・サイエンス』、Vol.21、N0.1、pp.37-59。
- 本橋永至・磯崎直樹・長尾大道・樋口知之(2012)、「状態空間モデルによるインターネット広告のクリック率予測」、『オペレーションズ・リサーチ』、Vol. 57、No. 10、pp. 574-583。
- 奥野拓也・中村和幸(2016)、「個人別セールスプロモーション効果の推定」、『情報処理学会論文誌数理モデル化と応用(TOM)』、9巻、3号、pp.61-74。
- 佐藤忠彦・樋口知之(2008)、「動的個人モデルによる消費者来店行動の解析」、『日本統計学会誌』、第38巻、第1号、pp. 1-19。
- 佐藤忠彦・北川源二郎(2002)、「マルコフ切り換え回帰モデルによる POS データの解析」、『行動計量学』、第29巻、第2号、pp.233-245。
- 山口類・土屋映子・樋口知之(2004)、「状態空間モデルを用いた飲食店売上の要因分解」、『オペレーションズ・リサーチ』、Vol. 49、No. 5、pp. 316-324。

Estimation, Diagnostics, and Extensions of Nonparametric Hawkes Processes with Kernel Functions

Jiancang Zhuang, Institute of Statistical Mathematics, Tachikawa, Tokyo 190-8562, Japan

The Hawkes self-exciting model has become one of the most popular point-process models in many research areas in the natural and social sciences because of its capacity for investigating the clustering effect and positive interactions among individual events/particles. It describes the stochastic excitations among a series of events that occur in a continuous time domain or in a spatiotemporal domain. A spatiotemporal marked Hawkes process, supposing its realization $N = \{t_i : i \in \mathbb{Z}\}$ with \mathbb{Z} being the set of all integers, has a conditional intensity in the form

$$\lambda(t, x, m) = \mu(t, x, m) + \int_{(-\infty, t) \times \mathbb{R}^d \times \mathbb{M}} g(t-s, x-u) f(m | m') dN(ds \times du \times dm') \quad (1)$$

where x and m denote the location in the space of \mathbb{R}^d and the mark in the space of \mathbb{M} , respectively, and $f(m | m')$ gives the conditional p.d.f. for the magnitudes of direct offspring from an event of magnitude m' .

For such a Hawkes process, the probability that an event, say j , is a background event, is

$$\varphi_j = \Pr\{\text{Event } j \text{ is a background event}\} = \frac{\mu(t_j, x_j, m_j)}{\lambda(t_j, x_j, m_j)} \quad (2)$$

and the probability that event j is triggered by another event i is

$$\rho_{ij} = \Pr\{\text{Event } j \text{ is triggered by } i\} = \frac{g(t_j - t_i, x_j - x_i) f(m_j | m_i)}{\lambda(t_j, x_j, m_j)}. \quad (3)$$

Making use of the above probabilities, we can construct an Expectation-Maximization algorithm for parametric Hawkes models and an iterative stochastic reconstruction algorithm for nonparametric or semi-parametric Hawkes models. When both the background rate and triggering term are nonparametric, for example, to estimate a Hawkes with conditional intensity

$$\lambda(t) = \nu\mu(t) + A \sum_{i:t_i < t} g(t - t_i) \quad (4)$$

as the conditional intensity function, and the estimates for μ and g become

$$\hat{\mu}(t) \propto \sum_i \varphi_i Z_{h_1}(t - t_i) \quad \text{and} \quad \hat{g}(t) \propto \sum_{i,j} \rho_{ij} Z_{h_2}(t_j - t_j - t). \quad (5)$$

with restrictions $\int_0^T \hat{\mu}(t) dt/T = 1$ and $\int_0^\infty \hat{g}(t) dt = 1$. In the above equations, the parameters ν and A are the relaxation coefficients and are estimated by the MLE or EM method.

Algorithm 1. *1. Stochastic declustering (Expectation). Calculate the background probability and triggering probabilities.*

2. Reconstruction (Maximization I). Estimate the nonparametric function in the model using nonparametric methods such as kernel functions.

3. Parametrization (Maximization II). Use the MLE method or EM algorithm to estimate the parameters in the parametric functions and the relaxation coefficients for the nonparametric functions.

The above method is applied to the data analysis in the studies of earthquakes and crimes. In the crime data, periodicity of the background rate is considered.

非定常クラスター過程のパラメータ推定法：Palm 尤度法の拡張

島谷健一郎（統計数理研究所）

空間点過程のパラメータ推定は、定常モデルから非定常へシフトしている。非定常モデルは定常モデルから定常性の仮定を外せば作られるが、どの定常性を外すかにより、いくつかの非定常化が考えられる。植物生態学では、定常ポアソンで分布する親の周りに散布された娘集団（親は既に死亡して位置は不明）という定常クラスター点過程 (homogeneous Neyman-Scott process) がよく使われる。ここに非定常性を入れる場合、

1. 定常ポアソンの親集団を非定常ポアソンにする。
2. 定常な親から散布された娘に非定常な生残率を与える。
3. 定常な親集団から、散布距離（散布カーネルのパラメータ、正規分布なら標準偏差）を非定常にする。
4. 1-3 の非定常性の 2 つ以上を同時に入れる。

などの選択肢がある。つまり、実データからのパラメータ推定では、同時に、どの非定常性を加味したモデルを考えるべきかという、モデル選択の問題に直面する。

ここでは、定常な場合に田中・尾形によって提唱された Palm 尤度法を、上記 1 と 2 の非定常に拡張し、それでモデル選択とパラメータ推定を行う手法を提唱する。Palm 最尤法に加え、Palm 尤度を用いるベイズ推定も提唱する。そして、簡単なシミュレーションで方法論の有効性と限界を検証した結果を報告する。人工点配置データとして、親集団の配置も娘集団に働く生残率も非定常とする full model に加え、いずれかを定常とするモデルでも点配置データを生成した。それらの人工データに、full model に加え、親集団または娘生残を定常（一定）とするモデルの Palm 尤度によるベイズ及び最尤推定も行った。

Palm 尤度を用いたベイズ推定では、MCMC は 1000 回程度できれいに単峰型に収束を示す場合と、振動する場合があり、親密度と娘生残の両方を非定常とするモデルでは、後者が多かった（図 1）。これは、現在の点配置データだけから、両者を分離して推定することが、一般には困難であることを示す。

一方、いずれかを定常として生成した点配置では、MCMC は明白な単峰型を示し、最尤推定値も求まった。さらに、点配置を生成したモデルの最大対数 Palm 尤度は、full model のものと 1.0 を下回る差しか持たず、赤池情報量規準 AIC で真の生成モデルが選ばれる結果となった（表 1）。

以上の結果をまとめると、一般に、親密度と娘の生残の同時推定は困難で、パラメータ推定値は振動したり、両者の中間になったりする。でも、うまく推定できる場合もある。また、単純化されたモデルで生成された点配置なら、AIC で真のモデルが選ばれ、パラメータ推定も悪くない。

なお、点配置データでは、サンプルサイズは、点の数でなく（これはデータの一部）、プロット (window) の大きさが対応する（漸近正規性も window の拡大で示されている）。点の数が多ければ推定精度が上がるわけではない。Neyman-Scott 過程では、親数が肝心で、実際、非定常ポアソン過程の推定では、200 点くらい欲しい。しかし、親は既いないので、今のデータからは不明で、推定して初めて目星がつく。人工データでは親数 > 200 で生成したからうまく推定できた。実データではやるまでわからない。

従って、いきなり複雑なモデルを膨大な MCMC で推定するより、本手法のような、簡略された推定法で何が推定でき何はそもそも無理そうか、試すのが賢明と言えよう。

図 1

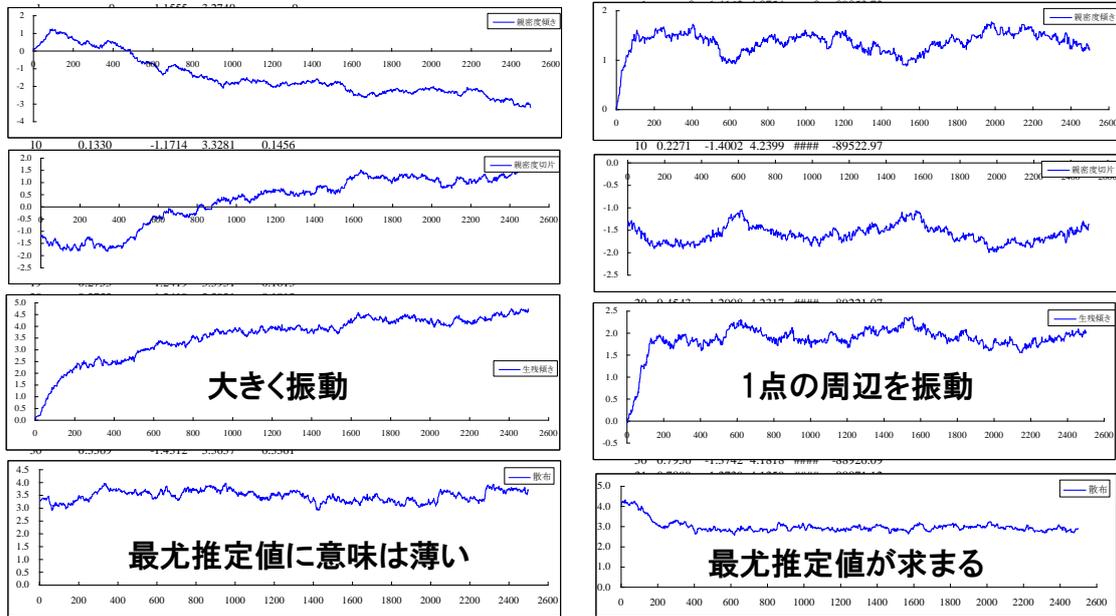


表 1

データ3	データ生成: 親も生残も非定常、 x に沿って減少					親と生残が同じ減少傾向 $\theta_1 = \varphi_1$				
	親密度傾き θ_1	親密度切片 θ_0	種子散布 σ	生残率傾き φ_1	最大対数尤度	親密度傾き θ_1	親密度切片 θ_0	種子散布 σ	生残率傾き φ_1	最大対数尤度
真値	1	-1.5	3	2		2	-2.2	3	2	
Full model	1.30	-1.51	2.91	1.97	-88148.8	2.55	-2.99	3.29	2.52	-142545.8
$\theta_1 = \varphi_1$	1.66	-1.85	2.94	1.66	-88149.3	2.54	-2.97	3.28	2.54	-142545.8
親: 非定常 生残: 定常						5.45	-5.29	3.08	0	-142640.2
親: 定常 生残: 非定常	0	-0.33	2.91	3.10	-88155.7	0	-0.57	3.26	4.58	-142565.4
データ2	親は x に沿って減少、生残は x に沿って増加					親は非定常、生残は定常				
真値	1	-1.5	3	2		2	-2.2	3	0	
Full model	-1.53	0.64	3.08	3.88	-69091.1	1.94	-2.80	3.95	-0.08	-84661.5
$\theta_1 = \varphi_1$	1.39	-1.99	3.12	1.39	-69135.6	0.72	-1.73	3.76	0.72	-84678.2
親: 非定常 生残: 定常						1.81	-2.69	3.94	0	-84661.9
親: 定常 生残: 非定常	0	-1.11	3.76	1.23	-84700.3					
真値	2	-2.2	3	-2		0	-0.8	3	2	
Full model	0.54	-1.61	3.20	-0.93	-46827.9	-0.31	-0.81	3.63	1.97	-68471.7
$\theta_1 = \varphi_1$	-0.30	-1.13	3.22	-0.30	-46841.4	0.95	-1.76	3.52	0.95	-68489.1
親: 非定常 生残: 定常	-0.59	-0.91	3.19	0	-46856.0	2.00	-2.34	3.26	0	-68540.6
親: 定常 生残: 非定常	0	-1.31	3.23	-0.55	-46833.2	0	-1.06	3.63	1.73	-68472.5

参考文献

Tanaka, U., Ogata, Y. & Stoyan, D. 2008. Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal*, 50, 43–57.