

シンポジウム「あたらしい統計科学」

報告書

開催日時：2024年9月23日（月・祝）～9月24日（火）

会場：石川県政記念 しいのき迎賓館 三階 セミナールーム

開催責任者：星野 伸明（金沢大学）

主催：科学研究費・基盤研究（A）「大規模複雑データの理論と方法論の革新的展開」

（課題番号：20H00576）

研究代表者：青嶋 誠（筑波大学）

— プログラムと目次 —

| 特別セッション「地震の大規模複雑データ解析」 | | | | | 座長：田中 勝人（一橋大学） | |
|------------------------|------|------|-------|---------|--|----|
| | 開始時刻 | 終了時刻 | 講演者 | 所属 | 演題 | 頁 |
| 9月23日 | 1255 | 1300 | | | 開会挨拶 | |
| | 1300 | 1400 | 熊澤 貴雄 | 統計数理研究所 | 点過程モデルを用いた異常地震活動の研究：M7.6能登半島地震に至る群発地震活動の時空間変動と異常活動の解釈について | 1 |
| | 1400 | 1500 | 長尾 大道 | 東京大学 | 地震ビッグデータ解析の最前線 | 3 |
| | 1515 | 1555 | 庄 建倉 | 統計数理研究所 | Probabilistic Earthquake Forecasting: How to Live Better Between Complete Randomness and Complete Predictability | 5 |
| | 1555 | 1635 | 矢野 恵佑 | 統計数理研究所 | 最小情報従属モデルの地震カタログへの適用 | 7 |
| | | | | | | |
| | | | | | | |
| 9月23日 | 1650 | 1730 | 竹内 努 | 名古屋大学 | 高次元統計解析の天文学への応用と展望 | 9 |
| | 1730 | 1810 | 間野 修平 | 統計数理研究所 | グラフ上の確率測度に値をとる拡散過程，衝突粒子系とその応用 | 11 |
| | | | | | 座長：張 元宗（統計数理研究所） | |
| 9月24日 | 930 | 1010 | 仲北 祥悟 | 東京大学 | 高次元ロジスティック回帰における最尤推定の一貫性 | 13 |
| | 1010 | 1050 | 平木 大智 | 東京大学 | Stochastic volatility in mean : Efficient analysis by a generalized mixture sampler | 15 |
| | 1050 | 1130 | 入江 薫 | 東京大学 | 逐次ベイズ予測統合 | 17 |
| | | | | | 座長：丸山 祐造（神戸大学） | |
| | 1300 | 1340 | 丸山 祐造 | 神戸大学 | Minimaxity under the half-Cauchy prior | 18 |
| | 1340 | 1420 | 張 元宗 | 統計数理研究所 | A revisit to the shrinkage estimator of normal mean vector restricted to a polyhedral cone | 19 |
| | 1420 | 1500 | 田中 勝人 | 一橋大学 | Brownian Motion, the Fredholm Determinant, and Time Series Analysis | 21 |
| | | | | | 座長：入江 薫（東京大学） | |
| | 1515 | 1555 | 二階 壮真 | 九州大学 | 精度行列の推定に基づく共分散構造のロバスト推定に関する研究 | 23 |
| | 1555 | 1635 | 佐藤 健一 | 滋賀大学 | Applying Non-negative Matrix Factorization with Covariates to the Longitudinal Data as Growth Curve Model | 25 |
| | | | | | 座長：柿沢 佳秀（北海道大学） | |
| | 1645 | 1725 | 柿沢 佳秀 | 北海道大学 | バイアスデータに対する非対称カーネル密度推定 | 27 |
| | 1725 | 1805 | 広瀬 修 | 金沢大学 | 形状位置合わせとその発展・応用 | 29 |
| | | | | | 座長：仲北 祥悟（東京大学） | |
| | 1805 | 1810 | | | 閉会挨拶 | |

点過程モデルを用いた異常地震活動の研究: M7.6 能登半島地震に至る群発地震活動の時空間変動と異常活動の解釈について

熊澤貴雄 統計数理研究所 リスク解析戦略研究センター 特任准教授

石川県能登半島の北東部では、2020年11月末から、それぞれに特徴のある活動性の高い群発地震が4つの地域に別れて発生した。群発地震の開始以前から微小地震活動の見られた南東部領域では、2020年11月末に活動が突如として14km以深へ移行し、数ヶ月の遅れで周辺の3領域に時計回りの順で活動が波及した。先行研究(Nakajima, 2022; Yoshimura et al, 2022; Yoshida et al., 2022; Nishimura et al., 2023)では当地域の深部から流体が供給されたことが指摘されており、地震発生地域が分断された理由は、この流体が地殻の強固な部分(地震空白地帯)を浸透して、周辺の脆弱な領域へと拡散していったことが示唆される。以降、北側2地域の地震活動性は高いまま推移し、両地域の境界付近で2022年6月19日にM5.4地震の発生を経て、2023年5月5日に北東部クラスター地域の北辺でM6.5の地震が発生し、その余震が北側2地域とその北方に広く分布した。一方、これに続く2024年元旦のM7.6は、群発地震の南北地域の境界部分、M6.5の余震分布の南辺を気象庁震源として発生し、南西—北東方向へと拡大した。

一連の地震発生パタンのフェイズ推移が示すように、発生様式は時間的に一様ではない。群発地震開始から本震M6.5とその余震活動を含む非一様な地震発生パターンに対応するために、今回報告した解析には以下の非定常ETASモデル(Kumazawa and Ogata, 2013)を用いた。

$$\lambda_{\theta}(t|H_t) = \mu(t) + \sum_{\{i: S \leq t_i < t\}} K_0(t_i) e^{\alpha(M_i - M_c)} / (t - t_i + c)^p \quad (1)$$

このモデルは定常ETASモデルの第1項の背景強度 μ と第2項の余震的誘発率のそれぞれに独立に時間変動要素を加味し、これらの変化を赤池ベイズ情報量規準(ABIC)法で推定するものである。背景強度関数 $\mu(t)$ は対象領域周辺でのゆっくりすべりによる応力変化、または断層内での流体貫入による断層弱化などを反映すると考えられる。他方 $K(t_i)$ は領域内の先行地震による近傍小断層群への地震連鎖効果(余震誘発率)を表わす。非定常ETASモデルと合わせて、デロネ型平滑化を用いた階層ベイズモデル(Ogata et al. 2020)で震源点密度分布及び b 値の空間分布を推定し、震源カタログを多角的に解析した。

これらの非定常ETASモデルの背景強度関数 $\mu(t)$ 及び地震密度分布推定から分かる地震活動異常が、幾つかのスロースリップまたは間隙流体圧変化のシナリオ

と矛盾しないことを報告した。本震の断層解が分かっていたら、回顧的に周辺でのスロースリップを仮定して、本報告の様に測地データや地震活動変化を解釈することはできるので、スロースリップの情報を事前に把握することは予測的観点から重要である。しかし多くの場合、特に内陸部でのスロースリップは規模が微小なのでインバージョンが難しく、精度の問題が付き纏う。その解決案として、可能なスロースリップの断層モデルを敷き詰めて地殻変動解析を実行できれば、どの断層モデルが地震発生パターンや GNSS 測地時系列の変化に適合しているか評価でき、それらの適合度を重みとして予測することが考えられる。このようなフォワード推論は組み合わせの問題だから大掛かりでモデル作成の労力や計算負荷が大きいであろうが、今の計算機だったら可能であろうかと思う。今回の報告での地殻変動解析には、地殻変動解析プログラム MICAP-G(内藤, 吉川, 1999)を用いた。データは 2024 年 8 月 1 日までの JMA カタログと、2023 年 5 月 16 日までの GNSS 観測時系列を使用した。

Kumazawa T, Ogata Y (2013) *J Geophys. Res. Solid Earth* **118** (12): 6165–6182.

<http://doi.wiley.com/10.1002/2013JB010259>

内藤宏人, 吉川澄夫 (1999) 地震 第 2 輯 52(1) DOI https://doi.org/10.4294/zisin1948.52.1_101

Nakajima J (2022) *Earth, Planets and Space* **74**, <https://doi.org/10.1186/s40623-022-01719-x>.

Nishimura T Hiramatsu Y, Ohta Y (2023) *Scientific Reports* 13, doi 10.1038/s41598-023-35459-z

Ogata Y. et al. (2020). *Computer Science Monograph*, No. 35, 統計数理研究所,

https://www.ism.ac.jp/editsec/csm/index_j.html

Yoshida K et al (2022) *J. Geophys. Res.* 128 (6), <https://doi.org/10.1029/2022jb026047>

Yoshimura, R et al. (2022) *Planetary and Space Sciences*, 3–7 November, Sagamihara, Japan 152th.

地震ビッグデータ解析の最前線

長尾 大道（東京大学地震研究所／東京大学大学院情報理工学系研究科）

地震分野においては、人工知能をはじめとする最先端の統計学的・情報科学的手法の積極的な導入が世界的に進められており、わが国でも文部科学省「情報科学を活用した地震調査研究プロジェクト」(STAR-E プロジェクト、研究期間：2021～2025 年度)をはじめとする複数の「情報×地震」の大型プロジェクトが実施されている。

人工知能と最も相性が良い地震研究のテーマは、地震波形データから P 波・S 波をはじめとする地震波の検出とその到来時刻を同定する「検出」と呼ばれる、全ての地震研究の第一歩となる作業である。現在でも気象庁が業務として実施している従来の検出手法は、局所定常自己回帰モデル、赤池情報量基準、および粒子フィルタ法などの従来の統計学的手法を組み合わせたものとなっているが、そう遠くない将来、これが人工知能に置き換えられることが確実視されている。その先駆けとなったのは、カリフォルニア工科大学が 2018 年に発表した深層学習モデル Generalized Phase Detection (GPD)法[1]である。著者を研究代表者とする STAR-E プロジェクトの研究課題「人工知能と自然知能の対話・協働による地震研究の新展開」(SYNTHA-Seis)においても、GPD 法を拡張することにより誤検出を減らすことに成功した地震波検出モデル（左図）などを開発した[2][3]。また、我々は 50～60 年前に稼働していたペン書き方式の地震計によって得られた紙記録からスロー地震[4]の一種である低周波微動を検出する深層学習モデル（右図）の開発も行った[5][6]。

このほか、本発表においては、マルチプル・クラスタリング法に基づく低周波地震検出のための観測点選択法[7]や 4 次元変分法データ同化に基づくスロー地震断層の摩擦パラメータの不確実性評価法[8][9]など、現在、SYNTHA-Seis で実施されている研究を紹介した。

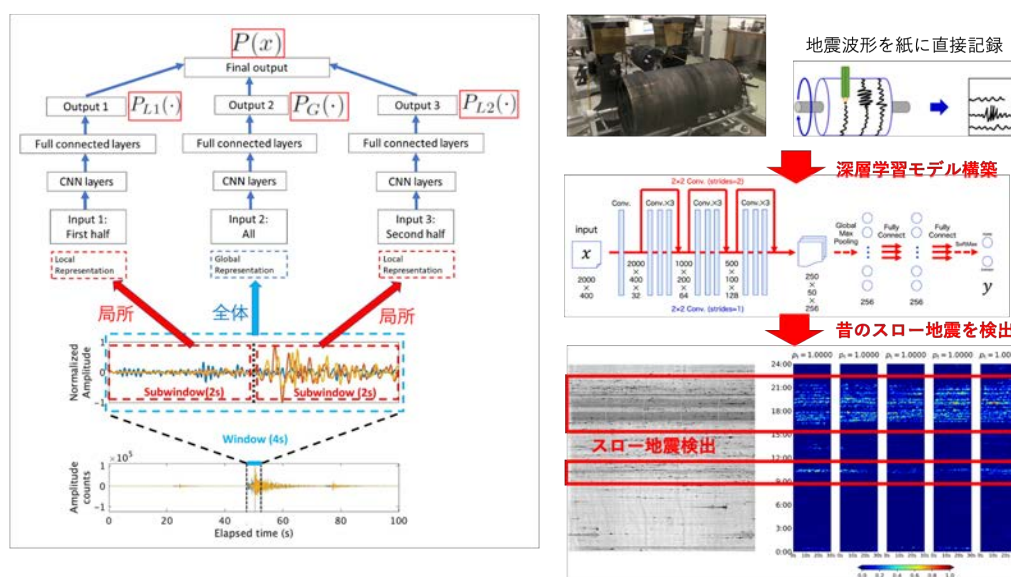


図: (左) GPD 法を拡張した地震波検出のための深層学習モデル[2]、(右) 深層学習を用いた約 50 年前の地震計紙記録からの低周波微動検出[5][6]。

謝辞

本発表内容は、文部科学省「情報科学を活用した地震調査研究プロジェクト」(STAR-E プロジェクト) No. JPJ010217、科研費基盤研究(A) No. 23H00466、国際共同研究加速基金(海外連携研究) No. 23KK018、挑戦的研究(萌芽) No. 20K21785、東大地震研特定共同研究(A) ERI JURP 2024-A-04、特定共同研究(B) 2022-B-06、2024-B-01 の支援を受けております。ご支援に対し、この場を借りて厚く御礼申し上げます。

参考文献

- [1] Ross, Z. E., M.-A. Meier, E. Hauksson, and T. H. Heaton, Generalized seismic phase detection with deep learning, *Bull. Seismol. Soc. Am.*, Vol. 108(5A), pp. 2894–2901, 2018.
- [2] Tokuda, T. and H. Nagao, Seismic-phase detection using multiple deep learning models for global and local representations of waveforms, *Geophys. J. Int.*, Vol. 235, Issue 2, pp. 1163–1182, doi:10.1093/gji/ggad270, 2023.
- [3] Katoh, S., Y. Iio, H. Nagao, H. Katao, M. Sawada, and K. Tomosaka, SegPhase: Development of Arrival Time Picking Models for Japan’s Seismic Network Using the Hierarchical Vision Transformer, *Earth Planets Space (in revision)*, doi:10.21203/rs.3.rs-4291782/v1.
- [4] Obara, K., Nonvolcanic deep tremor associated with subduction in southwest Japan, *Science*, Vol. 296, Issue 5573, pp. 1679–1681, doi:10.1126/science.1070378, 2002.
- [5] Kaneko, R., H. Nagao, S. Ito, H. Tsuruoka, and K. Obara, Detection of deep low-frequency tremors from continuous paper records at a station in southwest Japan about 50 years ago based on convolutional neural network, *J. Geophys. Res. Solid Earth*, Vol. 128, Issue 2, doi:10.1029/2022JB024842, 2023.
- [6] Kaneko, R., H. Nagao, S. Ito, K. Obara, and H. Tsuruoka, Convolutional neural network to detect deep low-frequency tremors from seismic waveform images, *Lecture Notes in Computer Science*, Vol. 12705, pp. 31–43, doi:10.1007/978-3-030-75015-2_4, 2021.
- [7] 徳田智磯, 長尾大道, 混合ウィシャートモデルに基づくマルチプル・クラスタリングによる低周波地震検出のための観測点選択, *応用統計学*, 52 巻, 2 号, pp. 99–112, doi:10.5023/jappstat.52.99, 2023.
- [8] Ito, S., M. Kano, and H. Nagao, Adjoint-based uncertainty quantification for inhomogeneous friction on a slow-slipping fault, *Geophys. J. Int.*, Vol. 232, Issue 1, pp. 671–683, doi:10.1093/gji/ggac354, 2023.
- [9] Ito, S., H. Nagao, A. Yamanaka, Y. Tsukada, T. Koyama, M. Kano, and J. Inoue, Data assimilation for massive autonomous systems based on a second-order adjoint method, *Phys. Rev. E*, 94, 043307, doi:10.1103/PhysRevE.94.043307, 2016.

Probabilistic Earthquake Forecasting: How to Live Better Between Complete Randomness and Complete Predictability

Jiancang Zhuang

Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

Earthquake hazards continue to pose serious threats worldwide, as seen in recent devastating events like the February 2023 Turkey-Syria earthquake, which resulted in over 50,000 deaths and widespread destruction. Such disasters underscore the urgent need for effective earthquake prediction and preparedness. Earthquakes can trigger tsunamis, landslides, and aftershocks, compounding their destructive impact on communities, infrastructure, and economies. People hope that scientific could predict these events before their occurrence so that we can take actions before hand to avoid our losses of lives and properties caused the earthquakes.

To forecast the occurrence of future disastrous earthquakes. Understanding the physical processes how earthquake ruptures are generated, accelerated, and stopped, is indispensable. In the 1990s, a series of papers by *Geller et al.* (1997) asserted that the occurrence of earthquakes cannot be precisely predicted. These papers led to a long argument on *Nature*. Even though the view of earthquake forecasting or prediction have changed over past 20 years, these papers still influencing the researches of earthquake forecasting studies. This is because, simply using Vere-Jones' branching crack model, we can simulate most of the empirical probability laws related to earthquake sources and magnitude and explain why earthquake occurrence cannot be deterministically predicted and discuss what are the potentially useful indices for evaluating the risk of future large earthquakes. However, earthquake source is not completely because of the existence of the critical zone in the crustal medium (*Zhuang et al.*, 2021).

As mention above, there is a significant gap between complete randomness and complete predictability in earthquake forecasting. In statistical seismology, the ETAS model has been accepted as the standard model for describing seismic activity (see special issue of *Huang et al.* (2016)) and adopted as the main earthquake prediction models by research institutes and government agencies in major earthquake-prone countries (*Schorlemmer et al.*, 2018). In particular, the United States Geological Survey (USGS) adopted the ETAS model as the UCERF3-ETAS model for short-term forecasts in the third all-California earthquake probability forecasting model (UCERF3) (*Field et al.*, 2017). This model the conditional intensity of the frequently

used spatiotemporal ETAS model is adopted from *Ogata* (1998):

$$\lambda(t, x, y) = \mu(x, y) + \sum_{i: t_i < t} \kappa(m_i) g(t - t_i) f(x - x_i, y - y_i; m_i), \quad (1)$$

where where $g(u)$ is the normalized form of the Omori-Utsuu formula $(p-1)/c(1+t/c)^{-p}$, the weighting function $\kappa(m) = A \exp[\alpha(m - m_0)]$ is the expected number of earthquakes directly triggered by an earthquake of magnitude m , and the spatial response kernel

$$f(x, y; m) = \frac{1}{\pi \sigma(m)} f_0 \left(\frac{x^2 + y^2}{\sigma(m)} \right) \quad (2)$$

is the density function for the relative location of the triggered earthquakes from an event of magnitude m , with f_0 being a normal density function $f_0(\omega) = \frac{1}{2D^2} e^{-\frac{\omega^2}{2D^2}}$ or a scaled inversed power law $f_0(\omega) = \frac{q-1}{D^2} (1 + \omega/D^2)^{-q}$ is considered, and the scaling function is $\sigma(m) = \kappa(m)$ or $\sigma(m) = [\kappa(m)]^{\gamma/\alpha}$. The magnitude of aftershocks need not be smaller than the triggering earthquake. An independent exponential distribution, i.e., the frequency of occurrence of the Gutenberg-Richter magnitude-frequency relationship is usually assumed for the purpose of theoretical discussion in simulation experiments of earthquake series. That is, $s(m) = \beta e^{-\beta(m-m_0)}$, $m \geq m_0$ is the probability density function form for the G-R law.

The usefulness of the ETAS model encourage researchers to extend this model to higher dimension and higher resolution of seismicity, including (1) *Impact of mainshock rupture geometry* (*Guo et al.*, 2015b); (2) *Impact of hypocenter depths* (*Guo et al.*, 2015a); (3) *Location-dependent and time-dependent ETAS parameters* (*Ogata*, 2004; *Zhuang*, 2015); (4) *Self-similar ETAS models* (*Vere-Jones*, 2005).

Currently, the predictability of earthquake occurrences that we have understood is illustrated in Figure 1. This figure situates predictability between complete randomness, represented by the Poisson process for time occurrences and the Gutenberg-Richter relationship for magnitude distribution, and complete determinism, where the target can be predicted with 100% precision. In recent years, the ETAS model has become a de facto standard model or null hypothesis for comparing with and testing other models and ideas (*Huang et al.*, 2016; *Zhuang et al.*, 2021; *Zhuang*, 2023), which in fact implies

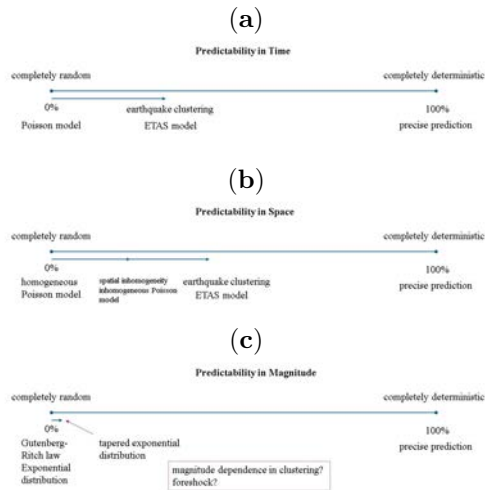


Figure 1: An illustration of earthquake predictability.

that the clustering is the largest predictable component in seismicity. In the RELM and CSEP projects, the highly scored models are almost among different versions of the ETAS models. Though there reports declaring non-seismicity precursors, their performance has not been fully validated.

In summary, owing to our inability to observe many of the fundamental processes of the system, as well as its inherent randomness, it is difficult to deterministically predict individual earthquakes. Therefore, statistical seismology, which places a larger focus on probabilistic forecasting, represents the best quantification method with respect to our state of knowledge. Furthermore, to provide more reliable earthquake forecast models, the challenge becomes the construction of models that can yield increased information gain with respect to a reference of the ETAS model. With the rapid development of observation technologies, an increasing amount of observational data has been obtained. These new observations provide new theories and approaches to help us understand seismicity. Based on these new observations, seismologists can develop new methods to more efficiently analyze these data and new models to connect them to the earthquake process and tectonic environments, thus providing more knowledge on the earthquake occurrence process and the ability to obtain reliable forecasts.

References

- Field, E. H., K. R. Milner, J. L. Hardebeck, M. T. Page, N. van der Elst, T. H. Jordan, A. J. Michael, B. E. Shaw, and M. J. Werner (2017), A Spatiotemporal Clustering Model for the Third Uniform California Earthquake Rupture Forecast (UCERF3)ETAS): Toward an Operational Earthquake Forecast, *Bulletin of the Seismological Society of America*, 107(3), 1049–1081, doi:10.1785/0120160173.
- Geller, R. G., D. D. Jackson, Y. Y. Kagan, and F. Mulargia (1997), Earthquakes cannot be predicted, *Science*, 275, 1616–1617.
- Guo, Y., J. Zhuang, and S. Zhou (2015a), A hypocentral version of the space-time ETAS model, *Geophysical Journal International*, 203(1), 366, doi:10.1093/gji/ggv319.
- Guo, Y., J. Zhuang, and S. Zhou (2015b), An improved space-time ETAS model for inverting the rupture geometry from seismicity triggering, *Journal of Geophysical Research: Solid Earth*, 120(5), 3309–3323, doi:10.1002/2015JB011979, 2015JB011979.
- Huang, Q., M. Gerstenberger, and J. Zhuang (2016), Current challenges in statistical seismology, *Pure and Applied Geophysics*, 173(1), 1–3, doi:10.1007/s00024-015-1222-7.
- Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics*, 50(2), 379–402, doi:10.1023/A:1003403601725.
- Ogata, Y. (2004), Space-time model for regional seismicity and detection of crustal stress changes, *Journal of Geophysical Research*, 109(B3), B03,308, doi:10.1029/2003JB002621.
- Vere-Jones, D. (2005), A class of self-similar random measure, *Advances in Applied Probability*, 37(4), pp. 908–914.
- Zhuang, J. (2015), Weighted likelihood estimators for point processes, *Spatial Statistics*, 14(B), 166–178, doi:http://dx.doi.org/10.1016/j.spasta.2015.07.009.
- Zhuang, J. (2023), Statistical seismology, in *Encyclopeida of Earth Sciences Series: Encyclopeida of Mathematical Geosciences*, edited by B. S. Daya Saha, Q. Cheng, J. McKinley, and F. Agterberg, Springer Nature Switzerland AG, Cham, doi:10.1007/978-3-030-26050-7_34-1.
- Zhuang, J., M. Matsu'ura, and P. Han (2021), Critical zone of the branching crack model for earthquakes: Inherent randomness, earthquake predictability, and precursor modelling, *European Physics Journal Special Topics*, 230, 409–424, doi:10.1140/epjst/e2020-000272-7.

最小情報従属モデルの地震カタログへの適用

女野 恵佑
統計数理研究所 数理・推論研究系
(東京大学 情報学教授との共同研究)

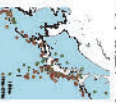
① シンポジウム「あたらしい統計科学」

Paper Barnouli, 30 (4) 2023, 3643 <https://doi.org/10.3150/23-BEJ1687>
Code (R and Python): <https://zenodo.org/record/8017290>



混合ドメインデータとは？

データセット：質的データ、数値データ、多様体値データ(方向データ)
例：地震カタログ



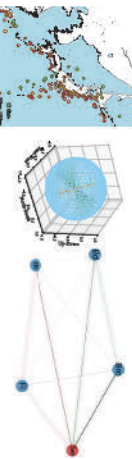
| No. | 経度 | 緯度 | 深さ [km] | 主方向 (az) [°] |
|-----|-------|------|---------|--------------------|
| 1 | 134.1 | 19.3 | 12 | 0.96, -0.63, 0.10 |
| 2 | 119 | 42.1 | 47 | (0.96, 0.13, 0.09) |
| 3 | 119 | 42.1 | 47 | (0.96, 0.13, 0.09) |

| No. | 種 | 性別 | 年齢 | 種別 |
|-----|--------|------|------|-------|
| 1 | Adelie | 18.7 | 18.0 | Adult |
| 2 | Adelie | 17.4 | 18.6 | Adult |
| 3 | Adelie | 17.4 | 18.6 | Adult |

→ 振数の形質のデータが混ざると依存性の統計解析は難しい

最小情報従属モデルリング

混合ドメインデータと高次の従属関係の統計モデルリング
変数値変数と多様体値変数の従属関係
高次元属性ネットワーク



多様体値変数：地震高次元情報(振数の向き)
変数値変数：震源の深さ
ノード：質的変数、質的変数
エッジ：次の従属
ハイパーエッジ：高次の従属

「最小情報」の意味

・最小情報従属モデルは次の最適化の解

$$\text{Maximize } -\int p(x) \log p(x) dx$$

$$\text{w.r.t. } p(x) \text{ satisfying}$$

$$\int p(x) h(x) dx = \text{const}$$

$$\int p(x) dx = 1, x \in \mathcal{X}$$

・相互情報量の一般化である総相関 (Mutualk, 1900) との関係

$$\int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx = \theta^T \nabla_{\theta} \log p(\theta, y) - \psi(\theta, y)$$

5

存在性

最小情報従属分布の存在性は $\theta^T h(x)$ のある種の可積分性に帰着される

- ① ある $h(x) \in L_1(\mathcal{X}, \mathcal{H}(x))$ が存在し

$$\int \exp(\theta^T h(x)) p(x) dx < \infty$$
 ならば最小情報従属分布は存在する。
- ② さらに、 $\sum \alpha_i \langle h_i, y \rangle + \psi(\theta, y)$ は一意に存在する。

・弱い条件のもと存在性が言えるのが強み
 ・例えば、多項式関数や距離空間の Lipschitz 関数の存在が保証される
 $h(x)$ が多項式関数や距離空間の Lipschitz 関数の存在が保証される

6

存在性から分かること：高次の従属関係の表現

☆ 高次の情報従属モデルは高次の振数の定数項の従属関係を表示できる

- ・同様の $d = 3$ 振数の変数データ ($X_1 = \theta$) を考える
- ・それぞれの振数の周辺分布を標準正規分布 $r_1(x) = \phi(x)$ とする
- ・ $h(x_1, x_2, x_3) = x_1 x_2 x_3$

$$p(x, \theta) = \exp(\theta(x_1, x_2, x_3) - 3\theta(x_1, \theta) - \psi(\theta))$$

$$r_1(x) = \phi(x), r_2(x) = \phi(x)$$



7

最小情報従属モデル

・ 確率密度 $X = \Pi_{i=1, \dots, d} X_i$ を任意の連続空間とする
 ・ 変数値変数 $X \in \mathcal{X} = \Pi_{i=1, \dots, d} \mathcal{X}_i$ の幾何分布として

$$p(x, \theta, y) = \exp\left(\theta^T h(x) - \sum_{i=1}^d \alpha_i \langle x_i, \theta \rangle - \psi(\theta, y)\right) \prod_{i=1}^d r_i(x_i, y)$$

$\theta \in \mathbb{R}^d$: 従属パラメータ
 $h(x) \in \mathbb{R}^d$: 所与の関数
 $r_i(x_i, y)$: 周辺分布族

・ 閉領域 $\alpha_i \langle x_i, \theta \rangle$ とテンソル積関数 $\psi(\theta, y)$ は次のように決まる:

$$\int p(x, \theta, y) dx = 1, x \in \mathcal{X}, y \in \mathcal{Y} \Rightarrow \int \sum_{i=1}^d \alpha_i \langle x_i, \theta \rangle p(x, \theta, y) dx = 0$$

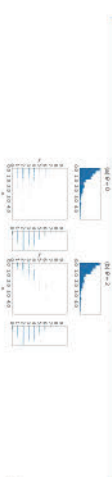
4

存在性から分かること：混合ドメインデータの従属関係の表現

☆ 最小情報従属モデルは混合ドメインデータの従属関係を表示できる

- ・ $X = h_x$ とし、 $y = h_y$ を自然数とする
- ・ 変数の周辺分布を $r_1(x) = \text{Exp}(1)$, $r_2(y) = \text{Poi}(\theta)$ とする
- ・ $h(x, y) = xy$

$$p(x, \theta) = \exp(\theta(x, y) - \theta(x, \theta) - \psi(\theta, y), r_1(x), r_2(y))$$



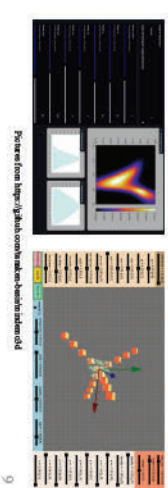
8

* 分布形状に関するデモ

MindemoQ4, MindemoQ4 by 田中研太郎先生@成蹊大学

> 2変数、3変数における分布形状の可視化

> <https://github.com/manekbc/mindemoQ4>



9

モデルの活用方法その1: 離散-連続混合&高次の従属関係

ペンギン-ペンギンの重なり、質的特徴量の高次の従属関係の抽出

$$\exp\left(\sum_{i=1}^d \alpha_i \langle h_i, x \rangle + \sum_{i=1}^d \alpha_i \langle h_i, y \rangle + \sum_{i=1}^d \alpha_i \langle h_i, z \rangle + \sum_{i=1}^d \alpha_i \langle h_i, w \rangle + \text{marginal part}\right)$$



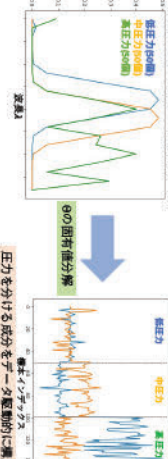
10

モデルの活用方法その2: 従属性に着目した特徴量抽出

分光スペクトルは通常により変化(例えば岩石の種類): Chang, Kishin, X et al., arXiv:2002.12321

分光スペクトル $F = (F(\lambda))_{\lambda \in \Lambda}$ と圧力の従属関係を抽出

$$\exp(\theta^T F + \text{marginal part}), \theta \in \mathbb{S}^m(M, \mathbb{R})$$



11

性質: コピュラモデルとの関係

コピュラ
 ● 高度な研究 (Sklarの定理: 変数変換)
 ● 変数値以外で非直線的、扱いが難しい
 ● 変数値以外で非直線的、扱いが難しい



14

性質: グラフイカルモデルとの関係

ノード任意変数ドメインの場合のグラフイカルモデルのクラスを定める



16

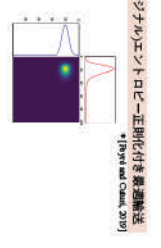
性質：エントロピー正則化付き最速輸送との関係

最小情報量の式を整理すると以下を得る：

$$\text{Minimize } \int p(x) (-\theta^T \text{Th}(\alpha)) dx + \int p(x) \log \prod_{i=1}^d \pi_i(x_i) dx$$

w.r.t. $p(x)$ satisfying

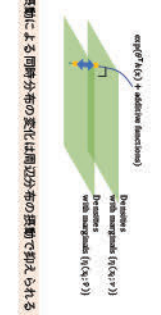
$$\int p(x) dx = 1, \pi_i(x_i) = 1, \dots, d$$



性質：エントロピー正則化付き最速輸送との関係から分かること

周辺分布の振動に対する Kullback-Liebler divergence での感度解析

$$D(p(x; \theta, \nu) \| q(x; \theta, \nu)) \leq \text{KL}(p(x; \theta, \nu) \| q(x; \theta, \nu)) + \theta^T (\text{Th}(\nu) - \text{Th}(\nu_0))$$



扱いにくい尤度の解決策：条件付き尤度最大化

- 1: 条件付き尤度

$$p(x^{(1)}, \dots, x^{(d)} | M_1, \dots, M_d) = \frac{\exp(\sum_{i=1}^d \theta^T \text{Th}(\alpha_i))}{\sum_{x \in \mathcal{X}} \exp(\sum_{i=1}^d \theta^T \text{Th}(\alpha_i))}$$

条件付き尤度最大化により従属パラメータの推定値(CLE)を得る：

$$\text{Max}_{\theta} \text{CLE}(\theta) = \frac{\exp(\sum_{i=1}^d \theta^T \text{Th}(\alpha_i))}{\sum_{x \in \mathcal{X}} \exp(\sum_{i=1}^d \theta^T \text{Th}(\alpha_i))}$$

*正規化定数の計算を回避するため、MKMNCにより推定を求める(CLE via MKMNC) (Dempster, 1969)

PLEの漸近分散

PLEを求める推定方程式は線形な取り扱いが難しい

$$E_{x \sim p} \sum_{i=1}^d \frac{1}{1 + \exp(\theta^T \text{Th}(X_i))} = 0$$

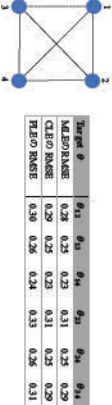
⊗ Xとで同じX(X)が得られる(独立でない)

→ 独立変数の関数の和の空間加法関数の空間に直交射影する (Hankel projection)

→ 追加制約を制約することにより尤度を示すことができる

$$A(\theta) = \sum_{i=1}^d \frac{1}{1 + \exp(\theta^T \text{Th}(X_i))} + \dots + A(\theta)$$

$$B_{\text{Hankel}} = \theta_0 \sim N\left(\frac{1}{d}, I\right)$$

$$I = \sum_{i=1}^d \frac{1}{1 + \exp(\theta^T \text{Th}(X_i))} \sum_{j=1}^d \frac{1}{1 + \exp(\theta^T \text{Th}(X_j))}$$


| Input θ | θ_1 | θ_2 | θ_3 | θ_4 | θ_5 | θ_6 |
|----------------|------------|------------|------------|------------|------------|------------|
| MAE of RMSE | 0.28 | 0.25 | 0.23 | 0.31 | 0.25 | 0.29 |
| CLE of RMSE | 0.29 | 0.25 | 0.23 | 0.31 | 0.25 | 0.29 |
| PLE of RMSE | 0.30 | 0.26 | 0.24 | 0.33 | 0.26 | 0.31 |

CLEはPLEも良い推定値を示す

扱いにくい尤度の解決策：Breggの類似尤度

- 2: Breggの類似尤度 (xを定数だけ入れ替えたものとして)

$$p(x^{(d)}, x^{(1)} | M_d, \dots, M_1, x^{(d)}) = \frac{1}{1 + \exp(\theta^T (\text{Th}(x^{(d)}) - \text{Th}(x^{(1)})) - \theta^T \alpha)} \prod_{i=1}^d \frac{1}{1 + \exp(\theta^T (\text{Th}(x^{(i)}) - \text{Th}(x^{(d)}) - \theta^T \alpha)}$$

類似尤度最大化により従属パラメータの推定値(CLD)を得る：

$$\text{Max}_{\theta} \text{PL}(\theta) = \prod_{i=1}^d \frac{1}{1 + \exp(\theta^T (\text{Th}(x^{(i)}) - \text{Th}(x^{(d)}) - \theta^T \alpha)}$$

*ロジスティック回帰に帰着できるので学習が容易/高速/高次元が容易

Contrastive divergence learning (Hinton, 2002) との関係

CLEとPLEを求める式を推定方程式を置き下してみるとどちらも(漸近的な)メーモントネットワークになる

$$\text{CLE} = \sum_{x \sim p} h(\theta \cdot \pi(x)) - E_{q(x)} \left[\sum_{i=1}^d h(\theta \cdot \pi_i(x)) \right] = 0$$

→ PLは MKMNC を途中でやめた学習 (Contrastive divergence learning)

$$\text{PLE} = \sum_{x \sim p} h(\theta \cdot \pi(x)) - E_{q(x)} \left[\sum_{i=1}^d h(\theta \cdot \pi_i(x)) \right] = 0$$

*PLは MKMNC を途中でやめた学習 (Contrastive divergence learning)

従属パラメータに対する推論

扱いにくい尤度 (unimodal likelihood)

$$p(x; \theta, \nu) = \exp\left(\theta^T \text{Th}(\nu) - \sum_{i=1}^d q_i(\theta; \theta, \nu) - \psi(\theta; \nu)\right) \prod_{i=1}^d r_i(x_i; \nu)$$

この部分を求めるのは大変...

最尤法を行う場合、エントロピー正則化付き最速輸送を利用した以下が考えられる：

1. 推定空間を離散化した Sinkhorn algorithm (Bedford and Wilson, 2014) or
2. Iterative scaling based propagation (Fussler et al., 2021)

⊙いずれも更新ごとに求める必要があるが大変

CLEの一意性・漸近分散

次の性質に着目する：正則条件のもと

$$\sup_{\theta} \frac{1}{n} \log \prod_{i=1}^n \frac{1}{1 + \exp(\theta^T (\text{Th}(x^{(i)}) - \text{Th}(x^{(d)})) - \theta^T \alpha)} \rightarrow 0 \text{ in mean}$$

→ 対数尤度比と対数条件付き尤度比は漸近的に等価

この性質を利用すると、CLE θ_{CLE} の異なる解を数枚推定値 θ_{CLE} の異なる解に帰着でき、漸近分散の候補として指数型分布 $\text{QL}(\theta)$ がソフマのソフマが利用可能

$$\theta_{\text{CLE}} \rightarrow \theta_0 \text{ in prob.}$$

サンプリング：交換アルゴリズム

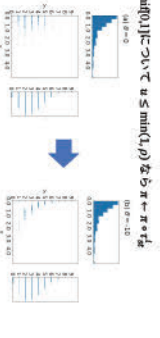
交換：LとRを一様ランダムに選ぶ

$r_{L,R} \in \mathcal{X}$ を \mathcal{X} を一様ランダムに選ぶ

$\rho = \frac{p(r_{L,R} | \theta)}{p(r_{L,R} | \theta_{\text{CLE}})}$ with $H_L(\rho) = \theta^T \text{Th}(M \cdot \pi(\rho))$

に基いて ρ と θ_{CLE} の交換に関する棄却法を行う

棄却法： $u \sim \text{Unif}(0, 1)$ について $u \leq \min(L, \rho)$ なら $\theta \leftarrow \theta_{\text{CLE}}$

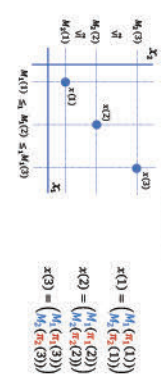


周辺経路分布による条件付け

n 個の推定 $x^{(1)}, \dots, x^{(n)}$ の順序で定ると...

推定 $x^{(i)} \in \mathcal{X}$ と周辺経路分布 $M_i, i=1, \dots, d$ を使い

$$x^{(i)} = \begin{pmatrix} M_i(x^{(1)}, \dots, x^{(i-1)}) \\ M_i(x^{(i)}, \dots, x^{(i-1)}) \end{pmatrix}$$



PLEの一意性

Breggの類似尤度の表示に着目：

$$w^T(x^{(d)}, x^{(1)}) = h^T(x^{(d)}) - h^T(x^{(1)}) - h^T(x^{(d)}) + h^T(x^{(1)})$$

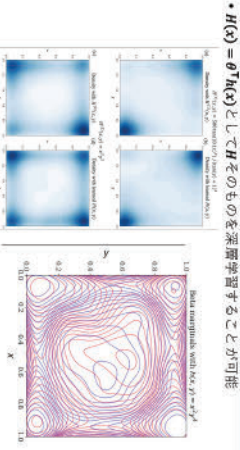
$$\frac{1}{2} \log \text{PL}(\theta) = \frac{1}{2} \sum_{i=1}^d \log \frac{1}{1 + \exp(\theta^T w^T(x^{(i)}, x^{(d)}))}$$

$$= E_{x \sim p} \sum_{i=1}^d \log \frac{1}{1 + \exp(\theta^T w^T(x^{(i)}, x^{(d)}))}$$

→ Hinton の U-Net 計算量 (Bridging, 1988)

このことに着目すると $\theta_{\text{CLE}} \rightarrow \theta_0 \text{ in prob.}$

* h がうまく与えられない時



まとめ

最小情報量推定モデルについてその応用性・性質・推論法を紹介した

→ 混合ドメインデータの同時分布のサンプリング

→ 活用例

混合ドメインデータの同時分布のサンプリング

→ 今後の課題：高次元性・欠損への挑戦

→ Paper: <https://doi.org/10.3150/23-BEJ1687>

→ Code: <https://zenodo.org/record/80172980>



Application of High-Dimensional Statistics to Astrophysics and its Perspective

Tsutomu T. TAKEUCHI^{1,2*}, Kazuyoshi YATA, Kento EGASHIRA,
Makoto AOSHIMA, Kohji YOSHIKAWA, Aki ISHII,
Ryusei R. KANO, Wen E. SHI, Aina May So, Hai-Xia MA,
Sena A. MATSUI, Koichiro NAKANISHI, Sucheta COORAY, Kotaro KOHNO

1. Division of Particle and Astrophysical Science, Nagoya University, Japan,
2. Research Center for Statistical Machine Learning, Institute of Statistical Mathematics, Japan,

1 Main Result

If we denote the dimension of data as d and the number of samples as n , we often meet a case with $n \ll d$. Traditionally in astronomy, such a situation is regarded as ill-posed, and they thought that there was no choice but to throw away most of the information in data dimension to let $d < n$. The data with $n \ll d$ is referred to as high-dimensional low sample size (HDLSS). To deal with HDLSS problems, a method called high-dimensional statistics has been developed rapidly in the last decade.

In this work, we first introduce the high-dimensional statistical analysis. We apply two representative methods in the high-dimensional statistical analysis methods, the noise-reduction principal component analysis (NRPCA) and automatic sparse principal component analysis (A-SPCA), to a spectroscopic map of a nearby archetype starburst galaxy NGC 253 taken by the Atacama Large Millimeter/Submillimeter Array (ALMA). The ALMA map is a typical HDLSS dataset. First we analyzed the original data including the Doppler shift due to the systemic rotation. The high-dimensional PCA could describe the spatial structure of the rotation precisely. We then applied to the Doppler-shift corrected data to analyze more subtle spectral features. The NRPCA and A-SPCA could quantify the very complicated characteristics of the ALMA spectra. Particularly, we could extract the information of the global outflow from the center of NGC 253. This method can also be applied not only to spectroscopic survey data, but also any type of HDLSS data. The main result is published in Takeuchi et al. (2024) and Takeuchi et al. (2024), Toukei SUuri, in press.

2 Further Development for the Next Generation Data

The original data of this study were recently updated to the one with much higher quality. The new data contains information of very weak spectral lines from molecules or radicals (ionized

*E-mail: tsutomu.takeuchi.ttt@gmail.com.

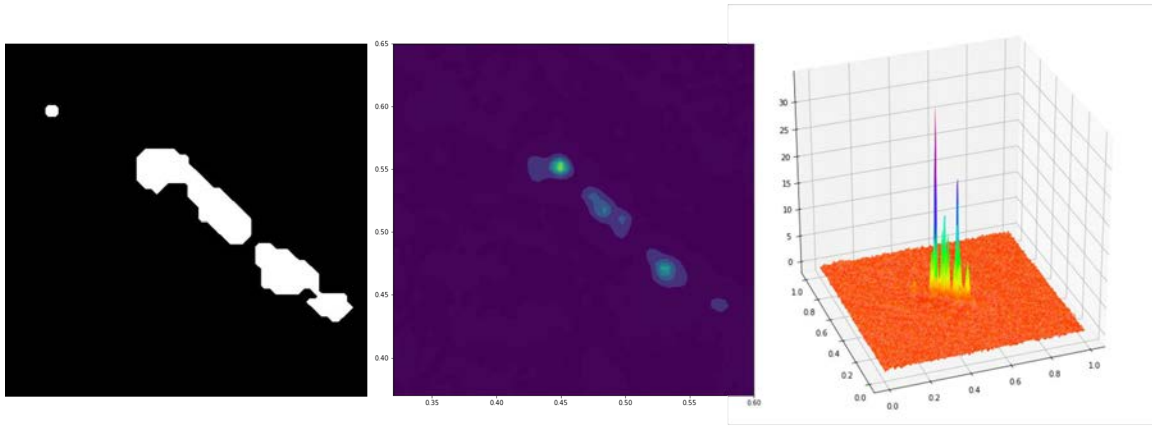


Figure 1: The bright regions of NGC 253 map cut out by the mask. Left: the mask region map. White regions have significant intensity signals. Center: the cut-out region with significantly bright emission. Right: the bird's view of the signal.

molecules) in NGC 253. To analyze such data, it would make sense to apply an analysis method which can deal with nonlinear correlation of data features. Kernel PCA is one of such possibilities. We will develop this study with such methods as our next step.

References

Takeuchi, T. T., Yata, K., Egashira, K., et al. 2024, *ApJS*, 271, 44

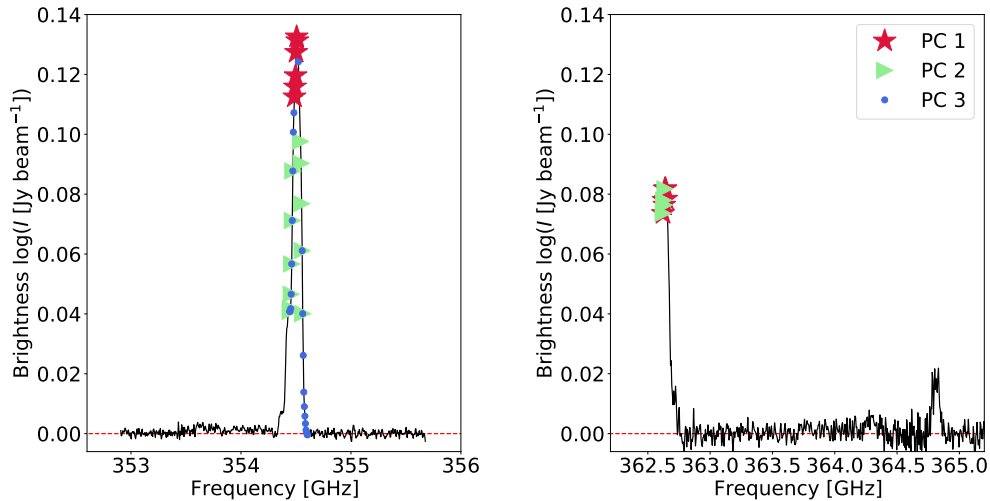


Figure 2: Responsible features to characterize PCs from the RPCA for the ALMA map of NGC 253, after the Doppler shift correction due to the systemic rotation. Information on the details of this figure is found in Takeuchi et al. (2024).

A measure-on-graph-valued diffusion: a particle system with collisions and their application

By

SHUHEI MANO

The Institute of Statistical Mathematics, Tokyo 190-8562, Japan

E-mail: smano@ism.ac.jp

A diffusion taking value in probability measures on a graph with a vertex set V , $\sum_{i \in V} x_i \delta_i$, was established. The masses on each vertices satisfy the stochastic differential equation (SDE) of the form $dx_i = \sum_{j \in N(i)} \sqrt{x_i x_j} dB_{ij}$ on the simplex, where $\{B_{ij}\}$ are independent standard Brownian motions with skew symmetry and $N(i)$ is the neighbour of the vertex i . The dual Markov chains on integer partitions were effectively used. Then, the SDE with the linear drift $\frac{\alpha}{2}(1 - |V|x_i)dt$, $\alpha > 0$, which gives killing of the dual chain with a linear rate, was discussed. The main result was

Theorem 1. *For the adjoint Markov semigroup $\{\tilde{T}_t^*\}$ associated with the diffusion on a graph $\mathcal{G} = (V, E)$, the unique stationary state ν_α in the set of probability measures on the simplex $\Delta_{|V|-1}$, is absolutely continuous with respect to the Lebesgue measure on $\Delta_{|V|-1}$ and admits a probability density that is strictly positive in $\text{int}(\Delta_{|V|-1})$ and of $C^\infty(\Delta_{|V|-1})$ -class.*

If the graph is complete, the stationary state is the symmetric Dirichlet distribution of parameter α . In this sense, the diffusion generalizes the Dirichlet distribution. No explicit expression of the density is available. However, an algorithm based on the dual Markov chain with the coupling from the past enables computing the marginal likelihood using the stationary states as priors.

Here, I report unpublished results on the application mentioned in this talk. The dataset was from “Shugi-in no ugoki 2023” (the report of movement of house of representatives of Japan). The forces are 10, 107, and 260 for the communist party, the democratic party (constitutional and national), and the liberal democratic party, respectively. This order represents left to right in their political stances. I used the Dirichlet distribution (complete graph K_3) and the stationary state of the S_2 graph as the priors, where S_2 is the graph consisting of three vertices in which two of the vertices are not neighbours. Table 1 shows the marginal likelihoods and the Bayes factors compared with the Dirichlet distribution obtained by the algorithm. The

results are averages of 10,000 trials. Any prior reduces to the multinomial sampling from the uniform distribution on the graph as $\alpha \rightarrow \infty$.

Table 1: Forces of parties

| | α | log ML | $\log_{10}\text{BF}$ |
|-------------------------|--------------|--------|----------------------|
| $K_3(\text{Dirichlet})$ | 1 | -810.5 | |
| $S_2(\text{C-D-L})$ | 1 | -677.9 | 58 |
| $K_3(\text{Dirichlet})$ | 85 (optimal) | -437.8 | |
| $S_2(\text{C-D-L})$ | 84 (optimal) | -439.4 | -0.7 |
| $K_3(\text{Dirichlet})$ | 10^6 | -537.5 | |
| $S_2(\text{C-D-L})$ | 10^6 | -537.5 | 0 |

For comparison, I analyzed an artificial data $a_1 = \dots = a_7 = a_{13} = \dots = a_{21} = 0$ and $a_8 = \dots = a_{12} = 5$ with anticipation that the data fits graph L_{21} consisting of 21 vertices aligned in a line (Table 2).

Table 2: An artificial data

| | α | log ML | $\log_{10}\text{BF}$ |
|----------------------------|--------------|--------|----------------------|
| $K_{21}(\text{Dirichlet})$ | 1 | -699.2 | |
| L_{21} | 1 | -483.6 | 94 |
| $K_{21}(\text{Dirichlet})$ | 81 (optimal) | -461.1 | |
| L_{21} | 79 (optimal) | -465.6 | -2 |
| $K_{21}(\text{Dirichlet})$ | 10^6 | -465.8 | |
| L_{21} | 10^6 | -465.8 | 0 |

These results agree with our anticipation if α is small: S_2 (the left and the right never interact) and L_{21} (the vertices occupied by particles are in the line) are supported. However, if we optimize α such that the marginal likelihoods increase, we cannot select a graph. Moreover, for such α , the Dirichlet distributions always perform the best.

I introduced the hyperparameter α as a “regularizer” to stop the algorithm in finite time, and $\alpha = 1$ is the uniform distribution on the simplex if we use the Dirichlet distribution. Because of this reason, I did not intend to take α large. Moreover, strict Bayesian will not set the hyperparameter to the “true” value. But the usefulness of the priors is indeed doubtful. The difficulty comes from the identifiability issue in the diffusion: we can explain the variability of the data by either bias in the diffusion or the drift.

高次元ロジスティック回帰における最尤推定の一致性

仲北祥悟

ロジスティック回帰は、統計学・機械学習におけるもっとも基本的な2値分類モデルの1つである。 $\mathbb{R}^p \times \{0, 1\}$ 値確率変数列 $\{(\mathbf{x}_i, y_i); i = 1, \dots, n\}$ を教師ありデータとして受け取り、 y_i を次の条件つき Bernoulli 分布に従う確率変数としてモデリングする：すなわち、ある $\boldsymbol{\theta} \in \mathbb{R}^p$ が存在して、任意の $i = 1, \dots, n$ および $a \in \{0, 1\}$ に対して、

$$\mathbb{P}(y_i = a | \mathbf{x}_i) = \sigma(\langle \mathbf{x}_i, \boldsymbol{\theta} \rangle)^a (1 - \sigma(\langle \mathbf{x}_i, \boldsymbol{\theta} \rangle))^{1-a}$$

が確率1で成り立つとするのがロジスティック回帰である。ここで、 $\sigma(t) = 1/(1 + \exp(-t))$ とする。パラメータ $\boldsymbol{\theta}$ は、 \mathbf{x}_i の実現値がどのように条件付き確率 $\mathbb{P}(y_i = 1 | \mathbf{x}_i)$ に影響するかを表現するパラメータであり、ロジスティック回帰での推定対象となる。その解釈の簡明さから、広く応用されている統計・機械学習モデルの1つである。

本講演では、特に高次元設定下での $\boldsymbol{\theta}$ の推定を議論する。低次元設定で一般的な推定方法は、負の対数尤度関数、すなわち経験リスク関数を最小化する最尤推定である。ここで、経験リスク関数 $\mathcal{R}_n : \mathbb{R}^p \rightarrow [0, \infty)$ は

$$\mathcal{R}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n (-y_i \log(\sigma(\langle \mathbf{x}_i, \boldsymbol{\theta} \rangle)) - (1 - y_i) \log(1 - \sigma(\langle \mathbf{x}_i, \boldsymbol{\theta} \rangle)))$$

と表される。最尤推定のモチベーションは、特にこの経験リスク関数 $\mathcal{R}_n(\boldsymbol{\theta})$ の最小化問題は母集団リスク関数 $\mathcal{R}(\boldsymbol{\theta}) = \mathbb{E}[\mathcal{R}_n(\boldsymbol{\theta})]$ の最小化問題に対する近似と見なすことにある。入力次元 p がサンプル数 n に対して十分小さいならば、この考えが正しいことは古典的な統計的漸近理論の枠組みで示される。一方で、 p が n に対してある程度大きいならば、この考えの正当化は難しくなる。この最小化問題が高い確率で不良設定問題となり得ることや (Candès and Sur, 2020)、良設定問題であっても真値に対してバイアスを持ち得ること (Sur and Candès, 2019) が近年示されている。本講演では、どのような設定であれば確かに経験リスク関数の最小化が母集団リスク関数の最小化問題に対する近似となるかを考える。

特に、半径 $R \geq 0$ の閉球 $B[R] = \{\boldsymbol{\theta} \in \mathbb{R}^p; \|\boldsymbol{\theta}\| \leq R\}$ 上での $\mathcal{R}_n(\boldsymbol{\theta})$ の $\mathcal{R}(\boldsymbol{\theta})$ に対する一様大数の法則を調べる。制約条件のない \mathbb{R}^p 上での $\mathcal{R}_n(\boldsymbol{\theta})$ の最小化の代わりに、次の制約最小化問題を考える。

$$\text{minimize } \mathcal{R}_n(\boldsymbol{\theta}) \text{ subject to } \|\boldsymbol{\theta}\| \leq R. \quad (1)$$

これは有界凸集合上での平滑凸最適化問題であり、その解は常に存在し、計算機で効率的に求めることができる。このような問題の代替は、実際的であり、先行研究でもしばしば置かれる設定である (Kuchelmeister and van de Geer, 2024; Hsu and Mazumdar, 2024)。もし $\mathcal{R}_n(\boldsymbol{\theta})$ が $B[R]$ 上で $\mathcal{R}(\boldsymbol{\theta})$ に一様に近いのであれば、凸最適化問題 (1) は母集団リスク関数 \mathcal{R} に対する次の凸最小化問題に対して良い近似となる。

$$\text{minimize } \mathcal{R}(\boldsymbol{\theta}) \text{ subject to } \|\boldsymbol{\theta}\| \leq R. \quad (2)$$

このアイデアは、一様大数の法則が成り立つならば漸近的に正当化される。すなわち、次の一様概収束

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in B[R]} |\mathcal{R}_n(\boldsymbol{\theta}) - \mathcal{R}(\boldsymbol{\theta})| = 0 \right) = 1 \quad (3)$$

が成り立つならば、確かに経験リスク関数の最小化は母集団リスク関数の最小化問題の近似と見なせる。このような最小化問題の漸近的な同値性を導くために、一様集中不等式を議論する。

1つの有望な仮説は、最小化問題 (1) の固有次元が n に対して十分小さい時には、確かに $\mathcal{R}_n(\boldsymbol{\theta})$ は $\mathcal{R}(\boldsymbol{\theta})$ に対して近いだろうと考えることである。有力な固有次元としては、 $\boldsymbol{\Sigma} := \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top]$ の有効ランク $\mathbf{r}(\boldsymbol{\Sigma}) = \text{tr}(\boldsymbol{\Sigma}) / \|\boldsymbol{\Sigma}\|$ が挙げられる。ここで、行列 $\boldsymbol{\Sigma}$ に対して $\|\boldsymbol{\Sigma}\|$ はスペクトルノルムを表す。古典的な Rademacher 複雑度と McDiarmid の不等式を用いた議論により、次の次元非依存一様集中不等式を得ることができる (Bach, 2023)：適当な正数 $K > 0$ に対して有界性条件 $\mathbb{P}(\|\mathbf{x}_i\|^2 \leq K^2 \text{tr}(\boldsymbol{\Sigma})) = 1$ の下、任意の正数 $\delta \in (0, 1]$ に対して、 $1 - \delta$ 以上の確率で、

$$\sup_{\boldsymbol{\theta} \in B[R]} |\mathcal{R}_n(\boldsymbol{\theta}) - \mathcal{R}(\boldsymbol{\theta})| \leq 2\sqrt{\frac{R^2 \|\boldsymbol{\Sigma}\| \mathbf{r}(\boldsymbol{\Sigma})}{n}} + \sqrt{\frac{8(1 + R^2 K^2 \|\boldsymbol{\Sigma}\| \mathbf{r}(\boldsymbol{\Sigma})) \log \delta^{-1}}{n}} \quad (4)$$

が成り立つ。ここで、入力次元 p は右辺に現れていないことは注目に値する。この不等式から、確かに $\mathbf{r}(\boldsymbol{\Sigma})$ は p よりもこの問題の固有次元をよく表現していると言える。

しかし、この一様集中不等式 (4) は一様大数の法則 (3) を考える上ではタイトではない。Borel–Cantelli の補題と一様集中不等式 (4) によって、一様大数の法則の十分条件 $\mathbf{r}(\boldsymbol{\Sigma}) \log n / n \rightarrow 0$ を得るが、この $\log n$ の項は明らかに余分である。例えば極端なケースとして、 $\mathbf{r}(\boldsymbol{\Sigma}) \asymp n / \log n$ を考えると、 $\mathbf{r}(\boldsymbol{\Sigma}) / n \rightarrow 0$ であるから、一様大数の法則が成り立つと期待できる。しかし、 $\mathbf{r}(\boldsymbol{\Sigma}) \log n / n \asymp 1$ であるから一様集中不等式 (4) は一様大数の法則を導かない。

本講演では、これを改善する新しい次元非依存一様集中不等式を示す。正則条件の下、 $\boldsymbol{\Sigma}$, n , および p に依存しない定数 $c > 0$ が存在して、任意の正数 $\delta > 0$ に対して、 $1 - \delta$ 以上の確率で、

$$\sup_{\boldsymbol{\theta} \in B[R]} |\mathcal{R}_n(\boldsymbol{\theta}) - \mathcal{R}(\boldsymbol{\theta})| \leq c\sqrt{\frac{\|\boldsymbol{\Sigma}\| \mathbf{r}(\boldsymbol{\Sigma}) + (1 + \|\boldsymbol{\Sigma}\|)^2 (1 + \log \delta^{-1})}{n}} \quad (5)$$

が成り立つ (Nakakita, 2024)。この上界により、 $\mathbf{r}(\boldsymbol{\Sigma}) / n \rightarrow 0$ が一様大数の法則の十分条件となる。

参考文献

- Bach, F. (2023). *Learning Theory from First Principles*. https://www.di.ens.fr/~fbach/ltfp_book.pdf
- Candès, E. J. and Sur, P. (2020). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *The Annals of Statistics*, 48(1):27–42.
- Hsu, D. and Mazumdar, A. (2024). On the sample complexity of parameter estimation in logistic regression with normal design. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 2418–2437. PMLR.
- Kuchelmeister, F. and van de Geer, S. (2024). Finite sample rates for logistic regression with small noise or few samples. *Sankhya A*. Advance online publication.
- Nakakita, S. (2024). Dimension-free uniform concentration bound for logistic regression. [arXiv:2405.18055 \[math.ST\]](https://arxiv.org/abs/2405.18055).
- Sur, P. and Candès, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29):14516–14525.

シンポジウム「あたらしい統計科学」報告書

平木 大智（東京大学経済学研究科）

作成日：2024年10月12日

1 基本情報

- シンポジウム名：あたらしい統計科学
- 主催：科学研究費・基盤研究（A）「大規模複雑データの理論と方法論の革新的展開」（研究代表者：青嶋 誠（筑波大学）、課題番号：20H00576）
- 開催日時：2024年9月23日 - 9月24日
- 会場：石川県政記念 しいのき迎賓館 三階 セミナールーム B

2 シンポジウム内容・所感

2.1 9月23日 特別セッション「地震の大規模複雑データ解析」

1日目のセッションは、2024年1月に発生した能登地震にちなんだものであった。本セッションでは統計数理研究所・東京大学から4名の先生がご講演され、先端的な深層学習の手法から古典的なモデルを用いた手法まで幅広い研究結果をご報告されていた。地震は空間・時間の方向に従属性を持ち、それらの複雑な情報を適切にモデル化する難しさを実感した。

空間・時間方向の従属性の扱いは、私の専攻する経済学に関連する分野でも非常に重要である。その意味で最も興味深かったのは統計数理研究所の熊澤先生・床先生が扱っていらっしゃる ETAS モデルであった。本モデルは自己励起性（ある現象の発生がその後の現象の発生確率に影響を及ぼす）を適切に描写するモデルであり、地震の分野から提案・発展してきたものである。一方で ETAS モデル、もしくは似た考え方により提案されたモデルは、地震と同様の現象を見せる一部の金融データ・マクロデータに対しても適用可能であるように考えられ、実際にそのようなアプローチをとる計量経済学分野の先行研究が少数ながら見られた。個人的な感想となるが、私自身の研究に対し新たなアプローチを提示していただける、とても良い機会となった。

特別セッション終了後は通常セッションとなり、竹内先生（名古屋大学）・間野先生（統計数理研究所）によるご講演がなされた。天体や拡散過程に関する、理論応用の両側面で興味深いご講演であった。

2.2 9月24日 通常セッション（報告者講演日）

2日目のセッションでは、統計数理研究所、および各大学から多くの先生・学生が発表された。発表テーマは様々であるが、概ね「大規模複雑データ」に関連するものであった。

24日は丸山先生（神戸大学）が座長を務める最初のセッションにて報告者は講演を行った。演題は「Stochastic Volatility in Mean: Efficient Analysis by a Generalized Mixture Sampler」であり、内容は金融データの特徴を捉えるのに優れた stochastic volatility in mean model のパラメータを推定するための手法を与えるものである。本講演において、長尾先生（東京大学）、矢野先生（統計数理研究所）、および間野先生（統計数理研究所）から非常に有益なご質問・関連情報のご提示をいただいた。特に矢野先生には講演後に私の講演内容をより改善するための、関連する先行研究をいくつか共有していただいた。非常に収穫のある講演をさせていただき、今後の研究活動も一層邁進する所存である。

24日は 報告者含め 10名の先生・学生が講演を行い、23日に続き多くの質問もなされる活発な時間となった。私の専攻するベイズ統計学に関連するものとして、入江先生（東京大学）、丸山先生（神戸大学）のご講演は非常に興味深いものであった。特に入江先生による「逐次ベイズ予測統合」は私の扱う高頻度の金融データに対しても重要なトピックであり、直接的な応用の可能性を感じさせる実りのあるご講演であった。その他では二階さん（九州大学）は「精度行列の推定に基づく共分散構造のロバスト推定に関する研究」の題でご講演をされており、当該分野で観測される非自明な現象の原因を解明するご研究をなされていた。二階さんのご研究自体が興味深い一方、彼が修士2年生で私と同学年であるということで、同年代にも研究に励んでいる方がいるという点で、私自身の研究のモチベーションを保つのにも良い機会となった。

3 おわりに

開催責任者である星野先生（金沢大学）、主催の青嶋先生（筑波大学）、また私の研究にアドバイスをいただいた先生方をはじめとする参加者の方々にあらためて感謝いたします。

「逐次ベイズ予測統合」

東京大学経済学部 入江 薫

予稿の通り、ベイズ予測統合の逐次分析に関する研究について講演した。予稿の内容を講演するのに先立ち、ベイズ予測統合に関する基礎事項について、Masuda and Irie (2024) arXiv:2409.09660 に基づいて解説した。

講演後には、以下の質疑および議論があった。

- 逐次モンテカルロ法の実施においては、有効サンプルサイズを随時計算し、一定水準を下回った場合には粒子退化の問題があると判断して、マルコフ連鎖モンテカルロ法を用いて計算をやり直すことを提案している。実際の有効サンプルサイズは時点ごとに上下するが、この値は更新回数とともに単調減少するものではないかと質問があった。これは粒子をリサンプリングせず重みのみを更新する場合には正しいが、本研究では毎回粒子のリサンプリングを行っているためこのような挙動となっている。また、それにより粒子退化の問題が完全に払拭されるわけではないが、逐次モンテカルロ法によって計算した事後分布は、マルコフ連鎖モンテカルロ法を繰り返し行った計算した「真の事後分布」に極めて近いことから、計算の正確性を確認している。また関連して、パラメータが高次元の場合に逐次モンテカルロ法が本質的に困難になる問題についても指摘があった。この出張も一般的には正しいが、本研究における応用においては、上述の事後分布の比較などを通じて、近似精度の正確性を担保していると返答した。
- ベイズ予測統合はモデルにではなく、予測分布に関して確率を考える枠組みであると説明した。その点を踏まえると、もはやモデルを考える必要はなく、それどころか統計モデルに基づかない予測分布も統合可能であるのかと質問があった。この指摘は正しく、先行研究でも同様の指摘がなされており、実際に人間の専門家による予測（エコノミストによる経済予測）をベイズ予測統合により合成して予測を行った先行研究があることを紹介した。
- 自然科学分野への応用可能性について議論があった。ベイズ予測統合は経済および公衆衛生分野で応用されているが、他分野でも広く活用されることが期待される。

Minimaxity under the half-Cauchy prior

神戸大学・経営 丸山 祐造 東京大学・計数 & 理研・CBS 松田 孟留

正規分布によるベイズ階層モデル

$$y | \beta \sim \mathcal{N}_p(\beta, I_p), \quad \beta | \kappa \sim \mathcal{N}_p(0, \frac{1-\kappa}{\kappa} I_p), \quad \kappa \sim \pi(\kappa), \text{ for } \kappa \in (0, 1)$$

を考える。Polson & Scott (2012, Bayesian Analysis) は、上記モデルにおけるベイズ推論で

$$\pi(\kappa) \propto \kappa^{-1/2}(1-\kappa)^{-1/2}$$

を推奨した。 $\pi(\kappa)$ は $\lim_{\kappa \rightarrow 0} \pi(\kappa) = \lim_{\kappa \rightarrow 1} \pi(\kappa) = \infty$ を満たす U 字型であり、連続型の spike and slab prior と理解できる。また変数変換 $\lambda = \sqrt{1/\kappa - 1} \in (0, \infty)$ により、密度が

$$\pi(\lambda) \propto \frac{1}{1+\lambda^2} I_{(0,\infty)}(\lambda)$$

と変換されることが half-Cauchy prior と呼ばれる所以である。

さて $\beta \in \mathbb{R}^p$ の推定問題において、推定量の良さを平均二乗誤差で測るとき、MLE である y は $p \geq 3$ のとき非許容的である。Polson & Scott (2012) が half-Cauchy prior を推奨する根拠は、half-Cauchy prior のもとでのベイズ推定量

$$\hat{\beta}_{\text{HC}} = \left(1 - \frac{\int_0^1 \kappa^{p/2+1/2} (1-\kappa)^{-1/2} \exp(-\kappa \|y\|^2/2) d\kappa}{\int_0^1 \kappa^{p/2-1/2} (1-\kappa)^{-1/2} \exp(-\kappa \|y\|^2/2) d\kappa} \right) y$$

が良い性質を持つことである。特に、彼らは $\hat{\beta}_{\text{HC}}$ が MLE y を優越すること、つまり

$$E[\|\hat{\beta}_{\text{HC}} - \beta\|^2] \leq E[\|y - \beta\|^2] = p \quad (*)$$

を数値的に示唆した。本発表では以下のように、理論的に (*) を示す。

定理 $p \geq 7$ のとき、half-Cauchy prior のもとでのベイズ推定量 $\hat{\beta}_{\text{HC}}$ は y を優越する。

(*) であるための Stein's unbiased risk estimates に基づく十分条件は、

$$\frac{p-5}{2} + (p+3) \frac{M(-1/2, p/2+2, w)}{M(1/2, p/2+2, w)} - \frac{p+1}{2} \frac{M(-1/2, p/2+1, w)}{M(1/2, p/2+1, w)} \geq 0 \text{ for all } w \geq 0 \quad (**)$$

で与えられる。ただし、 $M(b, c, w)$ は合流型超幾何関数

$$M(b, c, w) = 1 + \sum_{i=1}^{\infty} \frac{b \cdots (b+i-1)}{c \cdots (c+i-1)} \frac{w^i}{i!}.$$

である。定理の証明においては、(**) に登場する合流型超幾何関数の比を

$$\frac{M(-1/2, p/2+2, w)}{M(1/2, p/2+2, w)} \geq -\frac{1}{5} \text{ for } p \geq 11$$

のように下から抑えることが肝要である。その際に、区間演算により精度保証をして、厳密に不等式が成立することを確認する。我々の知る限り、統計的決定理論・数理統計学において、区間演算を用いた証明は見当たらず、その点においても新規性があると考えられる。

A revisit to the shrinkage estimator of normal mean vector restricted to a polyhedral cone

Yuan-Tsung Chang (ISM) and Shinozaki Nobuo (Keio University)

Grant-in-Aid for Scientific Research (A) 20H00576 Symposium, Sep. 23-24/2024 at Kanazawa

Introduction : Here we focus on the simultaneous estimation of p -dimensions normal means which are restricted on polyhedral cone. In simultaneous estimation of normal means when there are no restrictions given on the normal means, Stein (1961) has proposed an estimator called Stein estimator which shrinks the UE to the origin and dominates the UE under squared error loss when dimensions are greater than or equal to 3. Alternatively, when there are restrictions given on normal means, does Stein type estimators still improve upon the UE ? Based on the above motivations, Chang (1981) first proposed the Stein type estimator when there are linear inequalities given on the normal means. Sengupta and Sen (1991) extended Chang's results to the positive homogenous set. Fourdrinier, Strawderman and Wells (2018) also discussed the normal mean vector restricted to a polyhedral cone by applying the lemma of Sengupta and Sen (1991). Amirdjanova and Woodroffe (2004) have also proposed Stein-type estimators when normal means are restricted to convex polyhedral cones. They showed that shrinking MLE towards an appropriate target(including the origin and sample mean) can also reduce the mean squared error.

The summary: Here we summarize Chang's result about the simultaneous estimation of normal means under linear inequality restrictions. We introduce Sengupta and Sen's (1991) results about the estimation of normal mean in positive homogenous set, the results of Fourdrinier et al. (2018) and Amirdjanova and Woodroffe (2004).

(i) Let $\mathbf{X} \sim N_p(\boldsymbol{\theta}, I)$, there are restriction $T\boldsymbol{\theta} \geq 0$, is given on unknown $\boldsymbol{\theta}$, where T is known matrix. Chang (1981) proposed a Stein type estimator when unknown mean vector $\boldsymbol{\theta}$ is restricted $T\boldsymbol{\theta} \geq 0$ as follows.

Stein type estimator which shrinks \mathbf{X} to the origin if $T\mathbf{X} \geq 0$:

$$\delta^O(\mathbf{X}) = \begin{cases} \left(1 - \frac{a}{\mathbf{X}'\mathbf{X}}\right)\mathbf{X}, & \text{if } T\mathbf{X} \geq \mathbf{0} \\ \hat{\boldsymbol{\theta}}(\mathbf{X}), \text{ the MLE of } \boldsymbol{\theta}, & \text{otherwise,} \end{cases} \quad (1)$$

where a is constant.

(ii) Sengupta and Sen (1991) have considered a positive orthant model with arbitrary covariance matrix.

Let $\mathbf{X} \sim N_p(\boldsymbol{\theta}, \Sigma)$ and restriction $\boldsymbol{\theta} \geq \mathbf{0}$ is given, where Σ is assumed to be known and positive definite. For $p \geq 1$, let $N_p = \{1, \dots, p\}$, A is a subset of N_p ordering by natural ordering and $B = N_p \setminus A$. We denote the cardinality of A by $|A|$. For p -dimensional vector \mathbf{X} and a subset $A \subseteq N_p$. Let \mathbf{X}_A denote $|A|$ -dimensional vector consisting of the components whose indices belong to A . Let

$$\mathbf{X}_{A:B}(\Sigma) = \mathbf{X}_A - \Sigma_{AB}\Sigma_{BB}^{-1}\mathbf{X}_B,$$

where Σ_{AB} denotes $|A| \times |B|$ submatrix of Σ consisting of the rows in A and columns in B , respectively, Σ_{BB} is defined similarly. For $\emptyset \subseteq A \subseteq N_p$, let

$$\mathcal{X}_A(\Sigma) = \{\mathbf{X} \in R^p | \Sigma_{BB}^{-1}\mathbf{X}_B \leq \mathbf{0}, \mathbf{X}_{A:B}(\Sigma) > \mathbf{0}\}.$$

Then from Kudo's (1963) result, the subset $\mathcal{X}_A(\Sigma), A \subseteq N_p$ are disjoint and $\bigcup_{A \subseteq N_p} \mathcal{X}_A(\Sigma) = R^p$. For every $A \subseteq N_p$, if \mathbf{u} and \mathbf{v} are $|A|$ and $p - |A|$ vector, respectively, we define

$$[\mathcal{P}_A(\mathbf{u}, \mathbf{v})]_i = \begin{cases} u_i, & \text{if } i \in A \\ v_i, & \text{if } i \in B, \end{cases}$$

Then the RMLE of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}}^{RMLE} = \Sigma_{\emptyset \subseteq A \subseteq N_p} \mathcal{P}_A(\mathbf{X}_{A:B}(\Sigma), \mathbf{0}) I_A(\mathbf{X} \in \mathcal{X}_A(\Sigma)) \quad (2)$$

where $I_A(\cdot)$ is indication function. Sengupta and Sen (1991) have proposed the shrinkage estimator

$$\hat{\boldsymbol{\theta}}^{SRMLE} = \Sigma_{\emptyset \subseteq A \subseteq N_p} I_A(\mathbf{X} \in \mathcal{X}_A(\Sigma)) \left(1 - c_a \frac{1}{\|\hat{\boldsymbol{\theta}}^{RMLE}\|_{\Sigma}^2}\right) \hat{\boldsymbol{\theta}}^{RMLE}, \text{ where } \|\hat{\boldsymbol{\theta}}^{RMLE}\|_{\Sigma}^2 = \left(\hat{\boldsymbol{\theta}}^{RMLE}\right)' \Sigma^{-1} \left(\hat{\boldsymbol{\theta}}^{RMLE}\right).$$

(iii) Fourdrinier et al. (2018) consider the $\boldsymbol{\theta}$ is restricted to the polyhedral cone, $\mathcal{C} = \{\mathbf{x} | \boldsymbol{\gamma}'_i \mathbf{x} \geq 0, i = 1, 2, \dots, m\}$

for fixed linear independent vectors $\gamma_i \in R^p, i = 1, \dots, m$, and assume that \mathcal{C} has non-empty interior \mathcal{C}^0 .

Some properties about polyhedral cone : P.1) \mathcal{C} is positively homogeneous, close and convex. **P.2)** For each $\mathbf{x} \in R^p$, the projection of \mathbf{x} onto $\mathcal{C}, P_{\mathcal{C}}\mathbf{x} \in \mathcal{C}$, which is a unique point in \mathcal{C} such that $\|P_{\mathcal{C}}\mathbf{x} - \mathbf{x}\| = \inf_{\mathbf{y} \in \mathcal{C}} \|\mathbf{y} - \mathbf{x}\|$. **P.3)** \mathcal{C} may be partitioned into $\mathcal{C}_i, i = 0, \dots, m$, that is $\mathcal{C} = \bigcup_{i=0}^m \mathcal{C}_i, \mathcal{C}_i \cap \mathcal{C}_j = \emptyset, i \neq j, i, j = 1, \dots, m$, where $\mathcal{C}_0 = \mathcal{C}^0$ and $\mathcal{C}_i, i = 1, \dots, m$ are relative interior of the proper faces of \mathcal{C} . **P.4)** Let $D_i = P_{\mathcal{C}_i}^{-1}\mathcal{C}_i$, then $D_i, i = 0, \dots, m$ form a partition of R^p . **P.6)** For each $\mathbf{x} \in D_i$ we have $P_{\mathcal{C}}\mathbf{x} = P_i\mathbf{x}$, where P_i is the orthogonal linear projection onto the s_i -dimensional subspace L_i spanned by \mathcal{C}_i . Also, for each \mathbf{x} , the orthogonal projection on L_i^\perp , is equal to $P_{\mathcal{C}}^*\mathbf{x}$, where $\mathcal{C}^* = \{\mathbf{y} | \mathbf{x}'\mathbf{y} \leq 0\}$ is the polar cone corresponding to \mathcal{C} . **P.6)** Additionally, if $\mathbf{x} \in D_i$, then $aP_i\mathbf{x} + P_i^\perp\mathbf{x} \in D_i$ for all $a > 0$, so D_i is positively homogenous in $P_i\mathbf{x}$ for fixed $P_i^\perp\mathbf{x}$. Hence, we may express.

$$\delta_{\mathcal{C}} = \sum_{i=0}^m I_{D_i}(\mathbf{X}) P_i \mathbf{X}. \quad (3)$$

They proposed shrinkage estimator as

$$\delta(\mathbf{X}) = \sum_{i=0}^m \left(1 - \sigma^2 \frac{g_i(\|P_i\mathbf{X}\|^2)(s_i - 2)^+}{\|P_i\mathbf{X}\|^2} \right) P_i \mathbf{X} I_{D_i}(\mathbf{X}) \quad (4)$$

dominates the rule $\delta_{\mathcal{C}}(\mathbf{X})$ provided $0 < g_i(t) < 2, g_i(\cdot)$ is absolutely continuous and $g_i'(t) \geq 0$, for each $i = 0, 1, \dots, m$.

(iv) Amirdjanova and Woodroofe (2004) have proposed a shrinkage estimator which shrinks to an appropriate target when θ is restricted to polyhedral cone, $\Omega = \{z | \gamma_i'z \geq 0, i = 1, 2, \dots, m\}$ for fixed linear independent vectors $\gamma_i \in R^p, i = 1, \dots, m$. Thus $m \leq p$. Given \mathbf{x} , let $\tilde{\theta} = \theta(\mathbf{x})$ be the projection of \mathbf{x} onto Ω .

Let L be the orthogonal complement of $span\{\gamma_1, \dots, \gamma_m\}$, and $\Gamma = [\gamma_1, \dots, \gamma_m]_{(p \times m)}$, let $\Gamma(\Gamma'\Gamma)^{-1} = [\delta_1, \dots, \delta_m]_{(p \times m)}$, and let $\delta_{m+1}, \dots, \delta_p$ be an orthonormal basis for L . Then $\delta_1, \dots, \delta_p$ is a basis for R^p and $\gamma_i'\delta_j = 1$, if $i = j$ and $\gamma_i'\delta_j = 0$, if $i \neq j$, since $\Gamma'[\delta_1, \dots, \delta_m] = I_m$ and $\Gamma'[\delta_{m+1}, \dots, \delta_p] = 0$. So if $\mathbf{z} \in R^p$ is written as $\mathbf{z} = \sum_{i=1}^p c_i \delta_i$, then $c_j = \gamma_j'\mathbf{z}$ for $j = 1, \dots, m$, and $\mathbf{z} \in \Omega \iff c_1, \dots, c_m \geq 0$.

Recall that $\tilde{\theta} = \theta(\mathbf{x})$ is the projection of \mathbf{x} onto Ω , and write $\tilde{\theta} = \sum_{j=1}^p \tilde{c}_j \delta_j$, where $\tilde{c}_j = \tilde{c}_j(\mathbf{x}) = \gamma_j'\tilde{\theta} \geq 0$ for $j = 1, \dots, m$.

For subsets $a \subseteq \{1, \dots, m\}$. and $a^c = \{1, \dots, m\} \setminus a$ and $B_a = \{\mathbf{x} \in R^p | \tilde{c}_j > 0 \text{ for } j \in a \text{ and } \tilde{c}_j = 0 \text{ for } j \in a^c\}$. We list some properties about B_a as follows: **P.1')** The set B_a partition R^p and their boundaries are of Lebesque measure zero (Meyer and Woodroofe(2000)). **P.2')** If $\mathbf{x} \in B_a$, then $\tilde{\theta}$ is projection on \mathbf{x} onto $span\{\delta_j, j \in a \cup m+1, \dots, p\}$ and $\mathbf{x} - \tilde{\theta}$ is the projection of \mathbf{x} onto $span\{\gamma_j, j \in a^c\}$.

If $a = \{k_1, \dots, k_r\}$, where $0 = k_0 < k_1 < k_2 < \dots < k_r < k_{r+1} = \min\{(m+1), p\}$, let

$$X_a = [\delta_{k_1}, \dots, \delta_{k_r}]_{(p \times r)}. \quad (5)$$

Then $P_a = X_a(X_a'X_a)^{-1}X_a'$ is the projection operator onto $span\{\delta_j, j \in a\}$ and $\tilde{\theta}(\mathbf{x}) = P_a\mathbf{x} + \Pi_L\mathbf{x}$, for $\mathbf{x} \in B_a$, where Π_L is the projection of \mathbf{x} onto L .

The target estimator Let $1 \leq \ell \leq m$ be a fixed integer and define the target estimator $\tilde{\theta}^0 = \sum_{j=\ell+1}^p \tilde{c}_j \delta_j$. Equivalently, for $\mathbf{x} \in B_a$, let $a = \{k_1, \dots, k_r\}$ be as (5) and s denote the unique integer for which $k_s \leq \ell < k_{s+1}$. Then

$$\tilde{\theta}^0 = \sum_{j=s+1}^r \tilde{c}_{k_j} \delta_{k_j} + \sum_{j=m+1}^p \tilde{c}_{k_j} \delta_{k_j},$$

where an empty sum is to be interpreted as zero. Thus since $\gamma_i'\delta_j = 0$ for $i \in a^c$ and $j \in a \cup \{m+1, \dots, p\}$,

$$\tilde{\theta} - \tilde{\theta}^0 = \sum_{j=1}^s \tilde{c}_{k_j} \delta_{k_j} = Q_a \mathbf{x} \text{ and } (\mathbf{x} - \tilde{\theta})'(\tilde{\theta} - \tilde{\theta}^0) = 0,$$

where $M_a = \begin{bmatrix} I_s & 0 \\ 0 & 0 \end{bmatrix}$, $Q_a = X_a M_a (X_a' X_a)^{-1} X_a'$, $Q_a^2 = Q_a$ and $tr(Q_a) = s$. To be noted that s depends on a and if $\ell = m$ then $\tilde{\theta}^0 = \Pi_L \mathbf{x}$ is the projection of \mathbf{x} onto L and for any $a \subseteq \{1, \dots, m\}$.

They proposed the shrinkage estimator is

$$\tilde{\theta} = \tilde{\theta}^0 + \left[1 - (D - 2) + \frac{g(\|\tilde{\theta} - \tilde{\theta}^0\|^2)}{\|\tilde{\theta} - \tilde{\theta}^0\|^2} \right] (\tilde{\theta} - \tilde{\theta}^0),$$

where $D = tr(Q_a)$ and g be an absolutely continuous, non-decreasing function on $[0, \infty)$, for which $0 \leq g \leq 2$.

Brownian Motion, the Fredholm Determinant, and Time Series Analysis

田中 勝人 (一橋大学名誉教授)

[0,1] 上で定義されたブラウン運動 $\{W(t)\}$ の 2 次および 1 次の汎関数, さらにそれらの比の統計量の分布を導出する方法について, 以下の (1), (2), (3) の報告をした.

(1) フレドホルム行列式 (FD) の一般的な導出方法

$K(s, t)$ が対称, 連続, ほぼ定符号の核関数のとき, 統計量

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

の特性関数は, Anderson-Darling (1952) の定理により,

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

で与えられる. ここで, $D(\lambda)$ は $K(s, t)$ の FD, $\{\lambda_n\}$ は $K(s, t)$ の固有値で, 重複度の分だけ繰り返される.

FD は, 本来, 積分方程式に基づいて無限級数

$$D(\lambda) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n d_n}{n!} \lambda^n, \quad d_n = \int_0^1 \cdots \int_0^1 \begin{vmatrix} K(t_1, t_1) & K(t_1, t_2) & \cdots & K(t_1, t_n) \\ \vdots & \vdots & & \vdots \\ K(t_n, t_1) & K(t_n, t_2) & \cdots & K(t_n, t_n) \end{vmatrix} dt_1 \cdots dt_n$$

で定義されるが, この定義から FD を求めるのは困難であるので, 導出可能な一般的方法を提案した. それは, 積分方程式を同値な微分方程式および境界条件に変換して固有値がみだす方程式から導出する方法である.

(2) 特殊な場合の FD の効率的な導出方法

核関数が

$$\begin{aligned} K_1(s, t) &= \min(s, t) - st - \sum_{j=1}^m \psi_j(s)\psi_j(t), \quad \psi_j(0) = \psi_j(1) = 0, \\ K_2(s, t) &= 1 - \max(s, t) - \sum_{j=1}^m \psi_j(s)\psi_j(t), \quad \psi_j(1) = 0 \end{aligned}$$

の場合に, これらの FD $D_1(\lambda)$, $D_2(\lambda)$ を微分方程式を使わずに, より効率的に導出できる次の公式を求めた.

$$D_1(\lambda) = \frac{\sin \sqrt{\lambda}}{\sqrt{\lambda}} |P_1(\lambda)|, \quad D_2(\lambda) = \cos \sqrt{\lambda} |P_2(\lambda)|$$

ここで, $P_1(\lambda)$ と $P_2(\lambda)$ は m 次の対称行列で, 各要素は積分で定義されるが, 数式処理言語を使うことにより容易に計算できる場合が多い. 例えば, Tanaka (1996) において MA モデルの単位根検定で使われた

$$\begin{aligned} K_1(s, t) &= \min(s, t) - st - \sum_{j=1}^3 \psi_j(s)\psi_j(t), \\ \psi_1(t) &= \sqrt{2}it(1-t), \quad \psi_2(t) = \sqrt{10}t^2(1-t), \quad \psi_3(t) = \sqrt{10}t(1-t)^2 \end{aligned}$$

に対する FD は、数式処理を使って次のように求めることができる。

$$\begin{aligned} |P_1(\lambda)| &= \frac{720}{\lambda^3 \sqrt{\lambda} \sin \sqrt{\lambda}} \left(4(6 + \lambda) - 8(3 - \lambda) \cos \sqrt{\lambda} - \sqrt{\lambda}(24 - \lambda) \sin \sqrt{\lambda} \right), \\ D_1(\lambda) &= \frac{\sin \sqrt{\lambda}}{\sqrt{\lambda}} |P_1(\lambda)| \\ &= \frac{720}{\lambda^4} \left(4(6 + \lambda) - 8(3 - \lambda) \cos \sqrt{\lambda} - \sqrt{\lambda}(24 - \lambda) \sin \sqrt{\lambda} \right) \\ &= 1 - \frac{3}{70} \lambda + \frac{1}{1260} \lambda^2 - \dots \end{aligned}$$

(3) さまざまな統計的問題への適用

上記の方法は、次のような統計的問題に適用できることを報告した。

- (a) 分布の適合度検定統計量の分布. ただし、局所対立仮説のもとでの分布は、FD の他に、レゾルベントに関連する計算が必要になる。例えば、

$$S = \int_0^1 (X(t) + m(t))^2 dt, \quad X(t): \text{平均 } 0 \text{ の正規過程}, \quad m(t): \text{連続関数}$$

と表される統計量の場合には、レゾルベントに関連する計算が必要となる。

- (b) 状態空間モデルを使ったパラメータの constancy に関する LBI 検定統計量の帰無および局所対立仮説のもとでの分布. 例えば、

$$y_j = \beta_j + \varepsilon_j, \quad \beta_j = \beta_{j-1} + \eta_j, \quad \begin{pmatrix} \varepsilon_j \\ \eta_j \end{pmatrix} \sim \text{NID} \left(\mathbf{0}, \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix} \right) \quad (j = 1, \dots, n)$$

において、 β_0 が未知定数の場合、局所対立仮説 $H_1: \rho = \sigma_\eta^2 / \sigma_\varepsilon^2 = c/n$ (n は標本サイズ、 c は正定数) のもとで、LBI 統計量は

$$S = \int_0^1 \int_0^1 [K(s, t) + cK^{(2)}(s, t)] dW(s) dW(t), \quad K^{(2)}(s, t) = \int_0^1 K(s, u)K(u, t) du$$

と表すことができる。

- (c) MA モデルにおける単位根 LBIU 検定統計量の分布. 局所対立仮説のもとでの統計量は、上記 (b) と同様の表現となる。例えば、 $y_j = \mu + \varepsilon_j - \alpha \varepsilon_{j-1}$ における α の単位根検定の LBIU 統計量は、局所対立仮説 $\alpha = 1 - c/n$ のもとで、上記 (b) の統計量 S で $K(s, t) = \min(s, t) - st - 3st(1-s)(1-t)$ となる。

- (d) AR モデルにおける単位根検定統計量の分布. 最小 2 乗推定量に基づく単位根検定統計量の局所対立仮説のもとでの分布は、Ornstein-Uhlenbeck 過程を含む比の形の表現となる。例えば、 $y_j = \rho y_{j-1} + \varepsilon_j$ の ρ の単位根検定において、対立仮説 $H_1: \rho = 1 - c/n$ のもとでの統計量は $\int_0^1 Y(t) dY(t) / \int_0^1 Y^2(t) dt$ に収束する。ここで、 $dY(t) = -cY(t) dt + dW(t)$ である。 $c = 0$ の場合が帰無仮説で、 $Y(t)$ はブラウン運動となる。

- (e) 離散的な多重単位根過程の連続バージョン

$$F_g(t) = \int_0^t F_{g-1}(s) ds = \frac{1}{g!} \int_0^t (t-s)^g dW(s)$$

に対して、統計量

$$S_g = \int_0^1 F_g^2(t) dt = \int_0^1 \int_0^1 K_g(s, t) dW(s) dW(t), \quad K_g(s, t) = \int_{\max(s, t)}^1 \frac{(s-u)(t-u)^g}{(g!)^2} du$$

の分布を導出した。

- (f) フラクショナル・ブラウン運動の 2 次汎関数の分布は未解決問題であるが、その近似分布を導出した。

精度構造の推定に基づく共分散構造のロバスト推定に関する研究

Soma Nikai¹, Yuichi Goto², and Koji Tsukuda³

¹*Graduate School of Mathematics, Kyushu University*

²*Faculty of Mathematics, Kyushu University*

³*Faculty of Mathematics, Kyushu University*

一般に共分散行列の任意のスカラー倍は共分散構造や Scatter Matrix, Shape Matrix と呼ばれる。共分散構造の推定は、単に変数間の関係に関する記述を得るだけでなく、主成分分析や因子分析など様々なデータ解析手法を用いることで、より細かなデータの解釈を得ることに役立つという観点からも重要である。

p 次元母集団分布 F を考える。単純のために $E[\mathbf{x}] = \mathbf{0}$ とし、共分散行列 $\Sigma := E[\mathbf{x}\mathbf{x}^\top] \in \mathcal{S}_+^p$ が存在するとする (ここで、 $\mathbf{x} \sim F$)。そして、分布 F からのサイズ n の i.i.d. サンプル $\mathbf{x}_1, \dots, \mathbf{x}_n$ について考える。このとき、 Σ の推定量として標本共分散行列 (SCM) $S := \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top \in \mathcal{S}_+^p$ がよく用いられる。また、共分散構造の推定量としてシンプルに SCM のスカラー倍がよく用いられる。ここで \mathcal{S}_+^p は $p \times p$ 半正定値対称行列全体の成す集合であり、 \mathcal{S}_{++}^p は $p \times p$ 正定値対称行列全体の成す集合であるとする。また、 $(\cdot)^\top$ で行列やベクトルの転置を表す。

しかしながら、SCM は外れ値の影響を受けやすいことが知られており、したがって外れ値が混入している恐れがあるデータに対して用いることは推奨されていない [2]。また、データの次元が高い場合、SCM の推定精度は良くないことも知られている [1]。更に、データの次元が高くなるほど外れ値の検出は一般に難しくなる [2]。

講演では、上記の問題に対処する方法を議論した。特に、 $\text{Tr}[V] = p$ を満たす共分散構造 $V \in \mathcal{S}_{++}^p$ (i.e. 共分散行列 Σ が存在するとき $V = (p/\text{Tr}[\Sigma]) \cdot \Sigma$ を満たす V) を外れ値の混入に対してロバストに推定する方法について、まずは代表的な既存手法を説明し、特に既存手法がうまくいかない状況を指摘した。そして、その状況に対処するための手法を提案した。ここで、任意の正方行列 A について $\text{Tr}[A]$ で A のトレースを表している。なお、制約 $\text{Tr}[V] = p$ は共分散構造を識別するためによく用いられる基準の1つである。

参考文献

- [1] Chen, Y., Wiesel, A., Elder, C. Y., and Hero, O. A. III. (2010), Shrinkage Algorithms for MMSE Covariance Estimation., *IEEE Transactions on Signal Processing* **58**, 5016–5029.
- [2] Ollila, E., and Tyler, E. D. (2014), Regularized M-Estimators of Scatter Matrix., *IEEE Transactions on Signal Processing* **62**, 6059–6070.
- [3] Taskinen, S., Frahm, G., Nordhausen, K., and Oja, H. (2023), A Review of Tyler’s Shape Matrix and Its Extensions. In: Yi, M., Nordhausen, K. (eds) *Robust and Multivariate Statistical Methods*, Springer, Cham. pp.23–41.

Applying Non-negative Matrix Factorization with Covariates to the Longitudinal Data as Growth Curve Model¹

Kenichi SATOH

Faculty of Data Science, Shiga Univ.

Last update: August 31, 2024

¹<https://arxiv.org/abs/2403.05359>



Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

1 / 31

Daily temperature in Canada



- Using the Canadian weather dataset in the `fda` package, we present our method.
- These are the daily temperatures at 35 different locations in Canada, averaged from 1960 to 1994.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

1 / 31

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

3 / 31

Non-negative Matrix Factorization

On the other hand, **Non-negative Matrix Factorization (NMF)** can be also applied.

- For non-negative observation matrix Y , we approximate it by the product of the **basis matrix X** and the **coefficient matrix B** , both having non-negative elements (see, Lee and Seung (1999, 2000)):

$$Y_{P \times N} \approx X_{P \times Q} B_{Q \times N}$$

where the number of bases Q satisfies $Q \leq \min(P, N)$ and it is also referred to as "rank".

- Unlike MLM, only Y is given to the NMF, which optimizes **not only B but also X** .
- To make the decomposition unique, **the column sum of X is 1**.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

7 / 31

R package: nmfkc

We have developed and published R package: `nmfkc` (NMF with Kernel Covariate) on GitHub². In MLM, $Q = 5$ was assumed for the fourth degree polynomial, whereas in NMF $Q = 2$. Then, R^2 was 0.985 which is still high.

```
> library(fda)
> data(CanadianWeather)
> d <- CanadianWeather$dailyAv[, , 1]
> Y <- d - min(d)
> library(nmfkc)
> result <- nmfkc(Y, Q=2)
> result$r.squared
[1] 0.9853615
```

²<https://github.com/ksatohds/nmfkc>

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

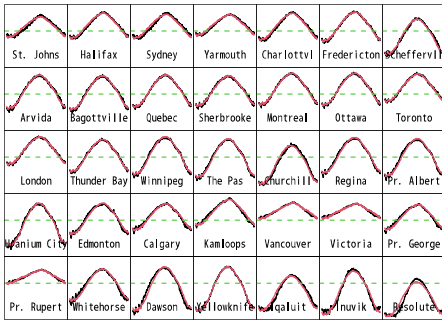
8 / 31

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

8 / 31

Approximation by NMF



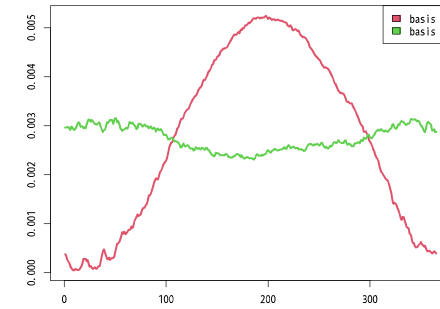
- Fitted values by NMF with $Q = 2$ are illustrated.
- Note that $Q = 2$ is the same as the number of bases in a simple linear regression, but makes the fit good.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

9 / 31

Basis vectors



- Two basis vectors in $X = (x_1, x_2)$ are illustrated.
- Basis 1 represents the temperature variation in the Northern area, with lower temperatures in winter and higher temperatures in summer.
- Basis 2, conversely, is shaped to push up winter temperatures.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

10 / 31

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

10 / 31

Non-negative coefficients at Pr. Rupert

- Temperature at Pr. Rupert can be approximated by two **coefficients** and bases,

$$y_{29} \approx 3826 x_1 + 11429 x_2$$

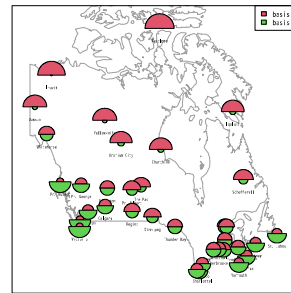
- The contribution of basis 2 is much larger than that of basis 1.
- The two coefficients are non-negative and can be interpreted as **probabilities**, i.e., the contribution of basis 1 is $3826 / (3826 + 11429) = 0.251$ and that of basis 2 is $11429 / (3826 + 11429) = 0.749$.
- The sum of the observations can be approximated by the sum of the coefficients, i.e., $15258 \approx 3826 + 11429 = 15254$.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

12 / 31

Soft-clustering using coefficients



- The probabilities obtained from the coefficients are shown on the map.
- The probability of basis 1 is higher north of the interior, and the probability of basis 2 increases southward.
- The probability of basis 2 tends to be higher on the west coast than on the east coast, even in the same southern area.
- These trends suggest that **the coefficient can be explained by longitude and latitude**.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

13 / 31

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

13 / 31

NMF with covariates

Next, we explain the coefficient matrix B using covariates. In other words, we use the **known covariate matrix $A = (a_1, \dots, a_n)$** and the **unknown parameter matrix Θ** , both having nonnegative elements, to write

$$B_{Q \times N} = \Theta_{Q \times R} A_{R \times N}$$

Thus,

$$Y_{P \times N} \approx X_{P \times Q} B_{Q \times N} = X_{P \times Q} \Theta_{Q \times R} A_{R \times N}$$

This approximation model, expressed as the product of three matrices, is included in the **tri-NMF**, proposed by Ding et al. (2006). Note that in tri-NMF, the three matrices are generally unknown; however, the covariate matrix in this study is known. Next, we show how to optimize the basis matrix X and the parameter matrix Θ .

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

14 / 31

Comparison with the Growth Curve Model

A comparison of NMF with the covariates and GCM is summarized in the table. The X optimized by NMF can be used in GCM as known design matrix.

| | NMF with covariates | GCM |
|----------|--------------------------------|--------------------|
| Model | $Y \approx X\Theta A$ | $E(Y) = X\Theta A$ |
| Y | non-negative values | continuous values |
| X | unknown, optimized, restricted | known |
| Θ | unknown, optimized | unknown, estimated |
| A | known | known, $R < N$ |

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

14 / 31

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

16 / 31

Objective function

We measure the discrepancy between Y and $\hat{Y} = X\Theta A$ by the **squared Euclidean distance**, or the squared Frobenius norm $\text{tr}(Y - \hat{Y})'(Y - \hat{Y})$, and consider the minimization with an L^2 regularization term:

$$D_{EU}(Y, \hat{Y}) = \text{tr}(Y - \hat{Y})'(Y - \hat{Y}) + \gamma \text{tr}\Theta'\Theta,$$

where γ is a non-negative penalty parameter, and Θ' denotes the transposed matrix of Θ .

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

18 / 31

Update formula

The update formula for the basis matrix X remains the same as that without using covariates, which was presented by Lee and Seung (1999),

$$X \leftarrow X \odot (YB' \oslash \hat{Y}B')$$

where \odot and \oslash correspond to the element-wise multiplication and division of matrices, respectively. Then, standardise X so that each column sum is 1. The update formulas for the parameter matrix Θ in tri-NMF were presented by Ding et al. (2006).

$$\Theta \leftarrow \Theta \odot \{(X'YA') \oslash (X'\hat{Y}A' + \gamma\Theta)\}$$

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

19 / 31

NMF with direct covariates

```
> u <- CanadianWeather$coordinates[,2:1]
> u[,1] <- -u[,1]; r <- apply(u,2,range)
> U <- t(scale(u,center=r[1,],scale=r[2,]-r[1,]))
> A <- rbind(rep(1,ncol(Y)),U)
> A[,c(1:2,35)]
      St. Johns   Halifax   Resolute
(Intercept) 1.000000 1.0000000 1.0000000
LON          1.000000 0.87410735 0.5149735
LAT          0.152208 0.05981835 1.0000000
> nmfkc(Y,A,Q=2)$r.squared
[1] 0.7151039
```

The approximation is **much worse** than in cases without covariates.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

21 / 31

Improve approximation using a Gaussian kernel function

By using individual data, u_1, \dots, u_N , we defined a **Gaussian kernel** for u

$$K(u_n, u) = \exp(-\beta|u_n - u|^2), \quad n = 1, \dots, N,$$

where $\beta > 0$ and Satoh (2023) proposed the NMF using the **kernel covariates**,

$$Y_{P \times N} \approx X_{P \times Q} \Theta_{Q \times N} A_{N \times N} \quad \text{where } A = (a_1, \dots, a_N) = \begin{pmatrix} K(u_1, u_1) & \dots & K(u_1, u_N) \\ \vdots & \ddots & \vdots \\ K(u_N, u_1) & \dots & K(u_N, u_N) \end{pmatrix}.$$

Additionally, Chen et al. (2022) conducted a study using a general kernel-based method for transformed observation $\phi(y)$ in NMF without covariates.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

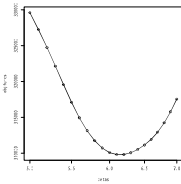
NMF as GCM

22 / 31

Cross-validation for optimization of β

The function `nmfkc.cv` can be used for **cross-validation**.

```
> betas <- 50:70/10
> objfuncs <- 0*betas
> for(i in 1:length(betas)){
+   A <- nmfkc.kernel(U,beta=betas[i])
+   result <- nmfkc.cv(Y,A,Q=2,div=10)
+   objfuncs[i] <- result$objfunc
+ }
> (best.beta <- betas[which.min(objfuncs)])
[1] 6.1
```



The optimized β is 6.1.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

25 / 31

Using Gaussian kernel function with optimized β

We applied NMF with kernel covariates using $\beta = 6.1$.

```
> A <- nmfkc.kernel(U,beta=6.1)
> result <- nmfkc(Y,A,Q=2)
> result$r.squared
[1] 0.9340371
```

0.715 (direct covariates) \ll 0.934 (kernel covariates) $<$ 0.985 (without covariates)
Using kernel covariates greatly improved the approximation accuracy over using direct covariates, and was no worse than the case without covariates.

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

26 / 31

Prediction of coefficients on new covariates

The covariate matrix A was created from **new covariates** $V = (v_1, \dots, v_M)$ on existing covariates $U = (u_1, \dots, u_N)$, and it is given by

$$A_{N \times M}^U(U, V) = \begin{pmatrix} K(u_1, v_1) & \dots & K(u_1, v_M) \\ \vdots & \ddots & \vdots \\ K(u_N, v_1) & \dots & K(u_N, v_M) \end{pmatrix}.$$

Using the covariate, the **coefficients at any given location** and **those probability** for each basis can be obtained.

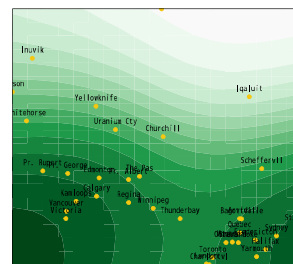
```
> v <- seq(from=0,to=1,length=20)
> V <- t(cbind(expand.grid(v,v)))
> A <- nmfkc.kernel(U,V,beta=6.1)
> B <- result$C %%% A
> B.prob <- prop.table(B,2)
```

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

27 / 31

Soft-clustering on any location



- The probabilities of basis 2 for any locations were obtained.
- The probability increases southward and it tends to be higher on the west coast than on the east coast.
- Daily temperature at given location can be also predicted. As a result, the daily geographic distribution is obtained³.

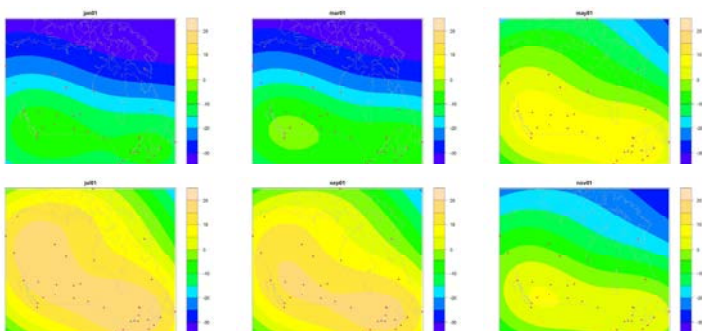
³<https://x.gd/KUBhB>

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

28 / 31

Daily geographic distribution



Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

29 / 31

Summary

- NMF with covariates was applied to the longitudinal measurement as GCM
- NMF with covariates is a less accurate approximation than NMF without covariates, but the accuracy can be improved by using Gaussian kernel functions.
- We have developed an R package: `nmfkc` (NMF with Kernel Covariates).
- NMF with covariates is applicable to various types of data³ including images⁴.



Thank you for your attention

³<https://x.gd/wMkYZ>

⁴<https://x.gd/1IfEX>

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

26 -

Kenichi SATOH (Faculty of Data Science, Shiga Univ.)

NMF as GCM

30 / 31

バイアスデータに対する非対称カーネル密度推定

柿沢 佳秀 (北大経済)

1. はじめに 台を半無限区間 $[0, \infty)$ とする密度 f を持つ母集団からの標本 $\{X_1, \dots, X_n\}$ を直接にはサンプリングできないが, その代わりに, 観測される $\{Y_1, \dots, Y_n\}$ は重み付き密度

$$f_w(x) = \frac{w(x)f(x)}{\mu_w}, \quad x \geq 0$$

に従うような状況を考えて¹. ここに w は正值関数で $\mu_w = \int_0^\infty w(t)f(t) dt (> 0)$ とした. 特に $w(t) = t$ の場合がレングス・バイアス (LB) な設定で, その密度を f_{LB} と書く.

先行研究 (e.g., Bhattacharyya et al.(1988), Jones(1991), Richardson et al.(1991)) では, Rosenblatt–Parzen による尺度型カーネルを用いたアプローチで LB サンプリングに対して f の推定量が 2 種類考察されたが, Jones の推定法では, いわゆる, 原点付近での境界バイアス問題が起こりうる. 実際, 技術的には f, f', f'' の有界性と $f(0) = 0$ を仮定すれば², $f'(0) \neq 0$ の場合, 境界付近のバイアスのオーダーが内点でのバイアスのオーダーよりも悪い, という意味の境界バイアス問題が起きる.

一方, 通常の直接サンプリングに対しては, 古くから, 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も議論されており, Rosenblatt–Parzen のカーネル密度推定量の境界バイアスに回避策があった (e.g., Jones(1993); この中には, “局所多項式法” も含まれていた). 近年, 推定するターゲット関数が区間 $[0, 1]$ または半無限区間 $[0, \infty)$ の台を持つ場合についてベータ・ガンマカーネル (Chen(1999,2000)) が脚光を浴びている.

本報告では, 非負バイアスデータに対して非対称カーネル法を応用した. すなわち, 非対称カーネル $k(\cdot; \beta, x)$ の台を $[0, \infty)$ とし, 平滑化パラメータ $\beta = \beta_n (> 0)$ は $\beta_n \rightarrow 0$ ($n \rightarrow \infty$) を満たし, 技術的に, 正数 $\epsilon = n^{-1/2}$ (say) も導入した.

2. 非対称カーネルと記号 (詳細は省略するが) 非対称カーネルに応じた, 個別的な議論を避け, 少なくとも非負カーネル $k(s; \beta, x)$, $s, x \geq 0$ は以下の 3 つの要請を満たすとした (e.g., Igarashi and Kakizawa(2020), Kakizawa(2021)):

1. $\mu_j(k(\cdot; \beta, x)) = \int_0^\infty (s-x)^j k(s; \beta, x) ds$ と $\int_0^\infty k^2(s; \beta, x) ds$ の近似, すなわち,

$$\begin{aligned} 1) J_b^{(k,u)}(x) &= \int_0^\infty k(t; b, x)u(t) dt \approx u(x) + b \left[\zeta_{1,1} u'(x) + \frac{\zeta_{2,1}}{2} x u''(x) \right], \\ 2) J_b^{(k^2,u)}(x) &= \int_0^\infty k^2(t; b, x)u(t) dt \approx b^{-1/2} \zeta \frac{u(x)}{x^{1/2}}; \end{aligned}$$

¹この例としてブレーキパッドを考えた. これは車の減速, 及び, 停止の要となる部品であり, その疲労が蓄積されると, いっそう摩耗する傾向がある. このような故障時間分布の推測 (生存時間分析) においてはバイアスサンプリング設定が重要である. Efromovich(1999; 92 頁) は飲酒運転者のアルコール濃度の比率の区間 $[0, 1]$ データの例を挙げているが, 致命的事故を起こしてしまった飲酒運転者の場合, 非常に高いアルコール濃度が検出される可能性が高い (バイアスデータ).

²このとき $f_{LB}(0) = f'_{LB}(0) = 0$ であり, $f_{LB}(x)$ を Rosenblatt–Parzen カーネル密度推定するような Bhattacharyya et al.(1988) の推定法では, 境界バイアス問題が生じないようにみえるが, 彼らの密度推定量の定義そのものには $1/x$ のファクターがあり ($x = 0$ で推定量は定義できていない), そのため, Jones(1991; 図 1(a)–(c)) の数値実験で報告されたように, $x = 0$ の近くでは必然的に発散してしまう.

2. 上界 $m_{\beta,\nu}(x) = \sup_{s \geq 0} \{s^{-\nu} k(s; \beta, x)\}$ のオーダー (直接サンプリングの場合は $\nu = 0$ のみを仮定すればよい) ;

3. $k(s; \beta, x)$ の x に関する裾積分の減衰 .

上記の 1 で現れる定数 $\zeta_{1,1}, \zeta_{2,1}$ と ζ (ただし, $\zeta_{2,1}, \zeta > 0$) を用いて

$$B_u(x) = \zeta_{1,1} u'(x) + \frac{\zeta_{2,1}}{2} x u''(x), \quad V_u(x) = \zeta \frac{u(x)}{x^{1/2}},$$

$I^{B_u^2} = \int_0^\infty B_u^2(x) dx$, $I^{V_u} = \int_0^\infty V_u(x) dx$ を定義した .

例 1. Amoroso カーネル族 (Igarashi and Kakizawa(2018)) .

例 2. 対称分布ベースの q-MIG カーネル族 (Kakizawa(2018)) . このサブクラスとして, $q \rightarrow 0$ は対数対称カーネル族である .

例 2'. 歪分布ベースの q-BS カーネル族 (Kakizawa(2021)) . ただし, 例 2 で対称分布を外したため, IG と RIG 相当の拡張を定義できないが, このサブクラスとして, $q \rightarrow 0$ は対数歪カーネル族である .

例 3. インデックスが $\nu \neq \pm 1/2$ の GIG カーネル (例外的に, $\nu = -1/2$ (IG) と $1/2$ (RIG) を, 正規分布ベースの例 2 へ分類した ; Igarashi and Kakizawa(2014) は GIG の代わりの別名 “変形ベッセル” を採用) .

例 1 から例 3 の詳細は柿沢 (2024) を参照 .

3 . 主結果 [Kakizawa(2024; to appear) による LB の場合を僅かに拡張した結果である] 既知の正数 r に対し $w(t) = t^r$ とする³ . オリジナルの密度 f の推定量として

$$\hat{f}_\beta(x) = \frac{n^{-1} \sum_{i=1}^n Y_i^{-r} k(Y_i; \beta, x)}{n^{-1} \sum_{i=1}^n (Y_i^r + \epsilon)^{-1} + \epsilon}, \quad x \geq 0$$

を提案するとき⁴, そのバイアスと分散について, $E[Y^{-2r}] = \mu_w^{-1} E[X^{-r}] < \infty$ の下で

$$\text{Bias}[\hat{f}_\beta(x)] \approx \beta B_f(x), \quad x \geq 0 \quad (\text{境界バイアス問題が回避された}),$$

$$V[\hat{f}_\beta(x)] \approx \frac{\mu_w V_{f_{-r}}(x)}{n\beta^{1/2}} = \left(\frac{\mu_w}{x^r}\right) \frac{V_f(x)}{n\beta^{1/2}}, \quad x > 0, \quad V[\hat{f}_\beta(0)] \approx \text{constant} \times (n\beta)^{-1}.$$

ここに $f_{-r}(\cdot) = f(\cdot)/(\cdot)^r$ とした . また, $B_f \neq 0$ ならば, 平均積分 2 乗誤差について

$$\begin{aligned} E\left[\int_0^\infty \{\hat{f}_\beta(x) - f(x)\}^2 dx\right] &\approx \beta^2 I^{B_f^2} + \frac{\mu_w I^{V_{f_{-r}}}}{n\beta^{1/2}} \\ &\geq \frac{5}{4^{4/5}} (I^{B_f^2})^{1/5} (\mu_w I^{V_{f_{-r}}})^{4/5} n^{-4/5}. \end{aligned}$$

注意. (i) 漸近正規性・強一貫性も成立する . (ii) 多次元のバイアスデータも議論できる .

³現実には r を未知とすべきだが, 識別性の問題がある . 実際, $w(t) = t^r$ の場合, ペア $(r, f(\cdot))$ と $(r+c, f_{-c}(\cdot))$ は同じ重み付き密度 $\frac{t^r f(x)}{\int_0^\infty t^r f(t) dt} = \frac{t^{r+c} f_{-c}(x)}{\int_0^\infty t^{r+c} f_{-c}(t) dt}$ に至る (e.g., Gilbert et al.(1999)) .

⁴分母に対し, 概収束 $\mathbb{P}(n^{-1} \sum_{i=1}^n Y_i^{-r})^{-1} \xrightarrow{a.s.} \mu_w$ を動機とし, この近似が $O_p(n^{-1/2})$ であることから, 主結果の証明では $\hat{f}_\beta^*(x) = n^{-1} \sum_{i=1}^n \mu_w Y_i^{-r} k(Y_i; \beta, x)$ の漸近的性質に帰着させた .

形状位置合わせとその発展・応用

広瀬 修

金沢大学 理工研究域 生命理工学類
hirose@se.kanazawa-u.ac.jp

形状位置合わせは2つの類似形状に対し、一方の形状を変形させもう一方の形状に重ね合わせる作業である。この作業により、類似形状の「対応点」を推定することができる。形状の対応点とは例えば左目の中心を表す点といった、同じ意味や役割を表す点の組とされる。形状の対応点は植物の葉の幅・高さの自動計測や、既存の形状から新たな形状を作る自動合成技術に応用することが可能である。講演では、形状位置合わせの様々な応用例について紹介し、非線形形状位置合わせに関する自身の貢献を紹介した。また、最近、形状位置合わせを「関数位置合わせ」に拡張し、関数位置合わせを経由して形状位置合わせを行うというアプローチを試行している。関数位置合わせとは、2つのよく似た関数が与えられた時に、一方の関数の定義域を変形しもう一方の関数に重ね合わせる作業である。この研究の進捗についても紹介を行った。