

位相データ解析における因果性とその系統学への応用

研究代表者 早水 桃子
(基幹理工学部 応用数理学科 准教授)

1. 研究課題

位相的データ解析 (TDA) はトポロジーを応用してデータの「かたち」を捉える解析手法の総称であり、近年注目されている応用数理の一分野である。複雑なデータの可視化や探索的データ解析の道具として様々な活用例が報告されているが、数理統計的な観点では多くの興味深い課題があり、実社会への応用の観点ではまだ TDA の手法が浸透していない分野も多い。例えば、TDA は材料や結晶の原子配置のような三次元データの「かたち」を分析する分野には自然な応用例が多数あるが、それと比較すると、時系列データのように物理的な「かたち」を持たないデータへの有用性には、まだ研究や考察の余地が多くある。共同研究者の谷口は Granger 因果性を一般化して時系列データ以外にも適用できる極めて一般的な因果指標を導入し、既存の因果指標では捉えられなかった潜在要因の抽出を可能にしたが、このような因果推論と TDA の関係を調べることは興味深い。一方で、近年の生命科学の研究分野では、シングルセルの遺伝子発現データから細胞分化のツリーを推定するタスクのように、非時系列データにおけるサンプル同士のつながりからサンプルの系統や擬似的な時系列を推定するデータ解析課題がある。既存研究の中には例えばパーシステントホモロジーを用いたウイルスの進化の解析や、Mapper を用いた細胞の分化の軌跡推定などがあるが、応用研究の蓄積が十分ではなく、生物や細胞の系統解析における TDA の有用性はまだ解明しきれていない。本研究では TDA・因果推論・系統解析の融合を目指し、①生物班と②統計班で研究を進めた。

①生物班： 複雑な進化の系統解析を可能にする系統ネットワークに関する基礎理論を構築し、それに基づいて系統樹と系統ネットワークを統計的な解析に活用するための高速アルゴリズムの開発を行った (例：系統ネットワークに含まれる最尤系統樹の高速計算、尤度値による系統樹ランキングの高速計算)。生物学の研究者とも協力して、系統ネットワークを実際の生物の系統解析に適用する研究も進めた。また、細胞生物学や関連分野の研究者とも協力し、シングルセルの遺伝子発現データに基づく細胞分化の軌跡推定への TDA の応用も行った。組合せ最適化アルゴリズムを応用した細胞分化のランドスケープ解析なども行った。

②統計班： Granger 因果性を含む極めて一般的な因果指標を、極めて一般的な乖離度から導入し、今まで捉えられなかった潜在要因の抽出を行った。具体的には、2 次モーメントを持たない安定過程に対する L_p ノルムでの最適予測、補間の基礎理論を構築し、安定過程の予測指標、因果指標に対する仮説を検定する経験尤度を提案し、因果性のあるなしを検定できるようにした。また 2 次モーメントを持たない多次元線形安定過程に対し、スペクトル行列は定義できないので、応答関数行列、行列型 power-transfer 関数を使い、この非線形積分汎関数で、多次元安定過程に対して極めて一般的な因果性指標を定義した。これに対してある種の周波領域的アプローチから推測・検定論の基礎もできた。また時間領域でのアプローチを行い L_1 ノルムでの 2 変量からの予測離反度を導入し、それによる、経験尤度に基づく因果性のある、なし、を検定する検定の提案、その漸

近分布の導出も行った。

2. 主な研究成果

①生物班： 早水の「系統ネットワークの構造定理」を応用し、系統ネットワークの全域系統樹に関する色々な計算問題に対する高速アルゴリズムを設計した。各辺が確率で重みづけられた系統ネットワークから最尤系統樹を線形時間で計算するアルゴリズムだけでなく、よりプラクティカルな応用を見据えて、最尤系統樹から順に（尤度の値に応じて降順に）全域系統樹を線形遅延で列挙する高速アルゴリズムを設計した。本研究で開発したアルゴリズムは、系統ネットワークの統計的な応用を初めて切り拓くものであり、しかも計算の効率性の意味で最適である（理論的に存在する中で最も効率的である）。この他にも、全域でない系統樹のサイズや計算方法と、無向グラフから有向系統ネットワークを得る向き付け問題を考察した論文プレプリントを2編発表した。系統ネットワークを活用したデータ解析としては鳥類の専門家との共同研究で鳥の系統解析を行った。その他、細胞生物学の実験研究者と連携して TDA の手法である Mapper アルゴリズムを用いてシングルセルの遺伝子発現データ解析を行い、生物学的に有用な知見を見出した。

②統計班： 基礎理論的なところでは、2次モーメントを持たない安定過程や、定常性を持たない場合は局所定常過程への極めて一般的因果性を導入した。またそれに基づく位相データ解析と、その応用等、基本的流れは順調に來ていると思われる。コロナ禍で、2020年に予定していた本資金での欧州ワークショップ：Rome-Waseda Workshop, Bologna-Waseda Workshop, Luxembourg-Waseda Workshop, Bergamo-Waseda Workshop は2021年度に延期されたが、再度、コロナ禍により、2022年度に再延期になったが、無事、遂行され、本研究成果の発表、また欧州の先端的研究者達との研究交換ができ、発展的展開を得た。また、当初は予期してない研究成果として、研究代表者：谷口は、高次のスペクトルから、極めて一般的な円周分布を定義して、モデル候補が少ない分野に、情報量基準でモデル選択ができる新地平も、築いた。また、時系列解析において、真の意味での Hellinger 距離は、導出されていなかったが、谷口は、スペクトル密度関数の言葉で、記述し、最小 Hellinger 距離推定量が、漸近有効で最尤推定量にくらべて頑健さを持つことも示した。そのほかには、初期解析としては、局所定常時系列の局所時間の尤度から求められる局所時間推定量を導き、そのバーコードからパーシステント・ランドスケープを求め、そこから位相情報を抽出し、まずは、シミュレーションで基礎モデルの有効性の検証をおこなった。また結果を脳波データや金融データに適用して新しい因果性を見た。さらに生物系統学や感染症医学の研究者とも協働して一般化因果性の生物学的な意義や有用性を専門的に考察し、因果性解析・TDA・系統解析の融合可能性について多角的に検討する。

3. 共同研究者

谷口 正信 早稲田大学・理工学術院・名誉教授

清水 泰隆 早稲田大学・理工学術院・教授

劉 言 早稲田大学・数理科学研究所・講師

後藤 祐一 早稲田大学・理工学術院講師（任期付き）

Xu Xiaofei 早稲田大学・理工総研・次席研究員 基盤（S）（谷口）による雇用

Xue Yujie 早稲田大学・理工総研・次席研究員 基盤（S）（谷口）による雇用

4. 研究業績

4.1 学術論文

Hayamizu, M. and Makino, K. (2022). Ranking top-k trees in tree-based phylogenetic networks, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), To appear. DOI: 10.1109/TCBB.2022.3229827

Taniguchi, M. and Xue, Y., (2023) Hellinger Distance Estimation for Non-Regular Spectra, Theory Probab. Appl. (to appear)

Goto, Y., Arakaki, K., Liu, Y. and Taniguchi, M. (2022). Homogeneity tests for one-way models with dependent errors under correlated groups, TEST DOI: 10.1007/s11749-022-00828-9

Y. Goto, T. Kaneko, S. Kojima, and M. Taniguchi (2022). ~Likelihood ratio processes under nonstandard settings, Theory Probab. Appl. 67 (2) , DOI: 10.1137/S0040585X97T990903

Goto, Y., Suzuki, K., Xu, X. and Taniguchi, M. (2022) Two-way random ANOVA model with interaction disturbed by dependent errors The Annals of the Institute of Statistical Mathematics. DOI: 10.1007/s10463-022-00853-3.

Xu, X.*, Taniguchi, M. and Murata, N. (2023) "UMVU estimation for Time Series". Papers in Statistical Inference for Time Series and Related Models - Essays in Honor of Masanobu Taniguchi. (to appear).

Xu, X., Liu, Y. and Taniguchi, M. (2023) "Higher order asymptotic theory for minimax estimation of time series". Journal of Time Series Analysis. 44(2): 247-257. .

Xu, X., Li, Z. and Taniguchi, M. (2022) Comparison between the exact likelihood and Whittle likelihood for moving average processes. STATISTICA. 82(1) 3-13. <https://doi.org/10.6092/issn.1973-2201/13609>

Xue, Y. and Taniguchi, M. (2023). A New Generalized Estimator for AR(1) Model Which Improves MLE Uniformly. Research Papers in Statistical Inference for Time Series and Related Models - Essays in Honor of Masanobu Taniguchi (to appear)

Liu, Y. (2023). A minimum contrast estimation for spectral densities of multivariate time series. Research Papers in Statistical Inference for Time Series and Related Models - Essays in Honor of Masanobu Taniguchi (to appear)

Goto, Y., Kley, T., Hecke, R. V., Volgushev, S, Dette, H., Hallin, M. (2022). The integrated copula spectrum. to appear in Ann. Statist.

Shimizu, Y.; Shirai, K.; Kojima, Y.; Mitsuda, D. and Inoue, M. (2023). Survival energy models for mortality prediction and the future prospects, *ASTIN Bulletin*, 53, (2), 377-391.

He, Y.; Kawai, R. Shimizu, Y. and Yamazaki, K. (2022). The Gerber-Shiu discounted penalty function: From practical perspectives, *Insurance: Mathematics and Economics*, 109, 1-28.

Nakajima, S. and Shimizu, Y. (2022). Parameter estimation of stochastic differential equation driven by small fractional noise, 2022; *Statistics*, 56, (4), 919-934.

Nakajima, S. and Shimizu, Y. (2022). Asymptotic normality of least squares estimators to stochastic differential equations driven by fractional Brownian motions, *Statistics and Probability Letters*. 187, 109476.

Kobayashi, M. and Shimizu, Y. (2022). Least-squares estimators based on the Adams method for stochastic differential equations with small Lévy noise, *Japanese Journal of Statistics and Data Science*. 5, 217-240.

Nakajima, S. and Shimizu, Y. (2022). Asymptotic inference for stochastic differential equations driven by fractional Brownian motion, to appear in *Japanese Journal of Statistics and Data Science*.

Kobayashi, M. and Shimizu, Y. (2022). Threshold estimation for jump-diffusions under small noise asymptotic, to appear in *Statistical Inference and Stochastic Processes*.

4.2 総説・著書

該当なし

4.3 招待講演

早水桃子 進化の系統樹と系統ネットワークに関する組合せ論 日本数学会 2022 年度秋季総合分科会 応用数学分科会（特別講演） 2022 年 9 月 13 日 日本数学会

早水桃子 進化の系統樹と系統ネットワークに関する組合せ論への招待 FRIS/TI-FRIS Life Science Seminar 2022 年 7 月 11 日 東北大学・学際科学フロンティア研究所

Hayamizu, M. Treefit: a method to measure the robustness of differentiation trajectories inferred from single-cell RNA-seq data, The 55th Annual Meeting of the Japanese Society of Developmental Biologists Symposium 4 : Novel insights into cellular plasticity in development, regeneration, and aging, 2022 年 6 月 3 日, 日本発生生物学会 (Japanese Society of Developmental Biologists)

早水桃子, 細胞分化の軌跡推定に関するソフトウェア Treefit の概要, Bio"Pack"athon2022#4, 2022 年 4 月 13 日 <https://doi.org/10.7875/togotv.2022.043>

清水泰隆, レヴィ保険リスクモデルにおける破産確率の区間推定問題, 応用数理学会・待ち行列研究部会@早稲田大学西早稲田西早稲田キャンパス, 令和 5 年 2 月 4 日.

Shimizu, Y.: Threshold estimation for jump-diffusions under small noise asymptotics, The 15th International Conference on Computational and Methodological Statistics (CMStats), King's College London, UK, December 19, 2022.

清水泰隆, ランダム・フォレストを用いた確率微分方程式の推定 ～金融実務への応用可能性は? ～ / 第 9 回金融シンポジウム「金融が直面する新環境への対応と方法論」, 統計数理研究所・リスク解析センター, 令和 4 年 12 月 12 日.

清水泰隆, 死亡率予測の新手法 ～ 生命エネルギー仮説と SEM project ～, 九州大学統計科学セミナー@九州大学理学部数学科, 令和 4 年 11 月 25 日.

4.4 受賞・表彰

- ・谷口正信 令和 4 年度 科学技術分野の文部科学大臣表彰科学技術賞 研究部門
- ・早水桃子 令和 4 年度 科学技術分野の文部科学大臣表彰 若手科学者賞
- ・村田悠, 早水桃子 2022 年度日本応用数理学会 優秀ポスター賞「横断歩道における歩行者対向流の組合せ論的考察」

4.5 学会および社会的活動

早水桃子:

- ・民間企業とのデータサイエンス系の共同研究: 1 件 (研究代表者: 早水桃子)
- ・学会活動: ICIAM 2023 Mini-symposium Challenges and Recent Advances in Phylogenetics
企画立案・運営等, パネル司会・セッションチェア等, 査読
- ・学会活動: ICIAM 2023 (International Council for Industrial and Applied Mathematics)
査読
- ・メディア掲載: 「生命科学に貢献する 新たな数理モデルの創出へ」 早稲田大学 Featured Article 真理の探求に挑む研究者たち (早稲田大学PI飛躍プログラム 2023年度支援対象者の研究内容) 2023年 インターネットメディア
<https://www.waseda.jp/inst/research/news/74285>
- ・メディア掲載: 「応用数学の視点から、細胞系統の解析手法を構築」 早稲田大学 研究活動紹介パンフレット 2023年 会誌・広報誌
- ・国内の研究集会・セミナー開催実績多数

- ・アウトリーチ活動多数
- ・早稲田大学PI飛躍プログラム 採択, JST創発的研究支援事業 採択

清水泰隆：

日本アクチュアリー会 客員,
日本アクチュアリー会 評議員,
日本アクチュアリー会 学術委員会アドバイザー
日本年金保険リスク学会誌 編集委員

Japanese Journal of Statistics and Data Science編集委員,
Journal of the Japanese Association of Risk, Insurance and Pensions Editor

劉言：

学術誌編集：

Econometrics and Statistics, Guest Editor (国際誌) 2022—現在

谷口正信：自動車産業との“数理統計学活用による共同研究” 研究代表者：谷口正信、
早稲田大学総合理工学研究所、名誉研究員

学術誌編集：

- (i) Associate Editor, Statistica Sinica (国際誌) 2020—現在
- (ii) Editorial Board, Statistical Inference for Stochastic Processes (仏誌), 2007—現在。
- (iii) Editorial Board, Mathematica Japonicae (国内誌), 1998—現在。

学術団体役職：

- (i) Institute of Mathematical Statistics (USA), 1987— 米国数理統計学会 特別会員 (Fellow)
- (ii) International Statistical Society (国際統計協会), 1987— 正会員 (Elected Member)
- (iii) イタリア Bologna 大学の学位審査委員 2021—

また、基盤 (S) Grand-in-Aid No.18H05290 (谷口) に基づき、谷口は下記、ワークショップを欧州で主催した。

1. “ROME- WASEDA TIME SERIES SYMPOSIUM” October 5 – 7, 2022 in Villa Mondragone, Rome.
Local organizer : Tommaso Proietti (Università di Roma)

<https://www.taniguchi.sci.waseda.ac.jp/2022kaken/RomeWaseda.pdf>

2. “BOLOGNA-WASEDA TIME SERIES WORKSHOP” October 8 – 9, 2022 in Accademia delle Scienze, Bologna
Local organizer : Alessandra Luati (Univ. Bologna)
<https://www.taniguchi.sci.waseda.ac.jp/2022kaken/BolognaWaseda.pdf>
3. “Luxembourg-Waseda Conference on Modelling and Inference for Complex Data”
January 23 – 25, 2023.
Local organizer: Christophe Ley (Univ. Luxembourg)
<https://math.uni.lu/midas/conferences/waseda2023/schedule.html>
4. “Bergamo-Waseda Workshop on Inference for Stochastic Processes and Applications”
January 26-28, 2023, Bergamo , Italy
Local organizer : Ilia Negri (Univ. Bergamo)
<https://bw-wispa.unibg.it>
5. “NUS-Waseda Workshop 2023” March 14, 2023 at National University of Singapore (NUS),
Local organizer : Ying Chen (NUS)
<https://www.taniguchi.sci.waseda.ac.jp/2022kaken/nuswaseda2023.pdf>
6. “IMR-Waseda Workshop Advances in pragmatic computational methodologies for fish stock assessment, human impact, and environmental factor on marine ecosystems”
March 28-29, 2023, at Institute of Marine Research(IMR), Bergen
Local organizer : Solvang Hiroko Kato (IMR).
<https://www.taniguchi.sci.waseda.ac.jp/2022kaken/bergenwaseda2023.pdf>
7. “Bolzano-Waseda Workshop on Statistics & Time Series analysis” at Hotel Oswald, Italy,
April 1 – 3, 2023.
Local organizer : Greta Goracci (Univ. Bolzano)
<https://www.taniguchi.sci.waseda.ac.jp/2022kaken/bolzanowaseda2023.pdf>

5. 研究活動の課題と展望

本研究は TDA・因果推論・系統学の融合分野を開拓する萌芽的な試みである。細胞の系統解析に TDA の手法である Mapper を活用することで有望な解析結果を得ることに成功したため、因果推論との接点を今後も継続的に探求する。統計学的には本研究は時系列や生命科学データに限らず極めて広い対象に適用可能なフレームワークを構築するものであり、複数のシグマ加法族の離反度

から因果性を導入すれば、その適用範囲は通常の、金融、医学、環境データだけでなく、グラフ、遺伝子、トポロジカル な対象まで広汎に広がる。またこの離反度への最適推測論の構築も極めてチャレンジングである。さらに複数の観測を点群データに変換し、パーシステント・ランドスケープという位相グラフ化し、これらの一般化離反度から、因果性、隠れ因子を推定すれば、従来の金融・統計手法では捉えられなかった未来の、突然の大変化の事前因子、隠れ因子が、一部見出せてきつつある。これは、従前の統計解析ではできなかった革新的手法になりうると思われる。この解析を系統解析や保険数理に適用すれば、いままでなされなかった因果性・TDA・系統学、保険数理の4分野の協業による革新的なパラダイムが出現し、進化生物学や感染症医学を含む生命科学分野や保険、年金数理への波及効果も生み出せるであろう。