

令和 6 年 5 月 29 日現在

機関番号：14401

研究種目：基盤研究(C)（一般）

研究期間：2018～2023

課題番号：18K11192

研究課題名（和文）連続変数を含む相互情報量の推定と、グラフィカルモデルの構築への応用

研究課題名（英文）Estimation of Mutual Information for Continuous Variables and its Applications to Graphical Model Construction

研究代表者

鈴木 譲 (Suzuki, Joe)

大阪大学・大学院基礎工学研究科・教授

研究者番号：50216397

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究は、連続変数を含むデータの相互情報量推定とグラフィカルモデルの構築に焦点を当てている。2018年度には、相互情報量の一致性と独立性検出を達成し、国際会議で発表した。2019年度には、連続変数を含むデータの相互情報量推定法を確立し、2020年度には、交絡を考慮した因果探索を提案した。2021年度には、交絡の存在を許容する因果順序の識別方法を開発し、2022年度には条件付き相互情報量の概念を拡張した。2023年度には、提案手法の実用性を検証した。そうした成果は、IEEE trans. on Information Theoryを含むジャーナル11件、国際会議4件に掲載している。

研究成果の学術的意義や社会的意義

連続変数を含むデータの相互情報量の正確な推定とグラフィカルモデルの構築を実現し、データ解析や機械学習の分野での新たな手法を提供した。これにより、遺伝子ネットワーク解析や経済データの依存関係解析など、複雑なデータの構造解明が可能となった。また、医療データ解析や金融リスク評価などの実世界の問題解決に応用が期待される。さらに、提案手法の普及を通じて、様々な分野でのデータ活用が促進され、社会全体のデータリテラシー向上にも寄与している。

研究成果の概要（英文）：This research focuses on the estimation of mutual information in data involving continuous variables and the construction of graphical models. In 2018, we achieved consistency and independence detection in mutual information estimation and presented these results at an international conference. In 2019, we established a method for estimating mutual information in data with continuous variables. In 2020, we proposed a causal discovery method considering confounding factors. In 2021, we developed a method for identifying causal order allowing for the presence of confounders. In 2022, we extended the concept of conditional mutual information. In 2023, we validated the practicality of the proposed methods. These findings have been published in 11 journal articles, including the IEEE Transactions on Information Theory, and presented at four international conferences.

研究分野：機械学習の数理

キーワード：相互情報量 グラフィカルモデル ベイジアンネットワーク

1. 研究開始当初の背景

統計的学習理論の種々の側面で、2変量に関するデータから、それらが独立であることからどれだけ乖離しているかを示す、いわゆる相互情報量を推定する状況が頻繁に生じる。その場合、サンプル数が大きくなった場合に正しい値を推定できる(一致性)ようにすることと、独立であるときに0であると推定できるようにすること(独立性検出)の両方が求められる。2変量が離散であるか連続であるかを区別しないで、一致性と独立性検出の両方の性質を満足する方法を開発する。その基本的なアルゴリズムは、本課題の着手の前に完成していて、遺伝子ネットワークの構築、および独立性検定 HSIC との比較についての成果を発表している。

2. 研究の目的

本研究では、多変量のデータから、事後確率を最大とするグラフィカルモデルを構築する問題を検討する。そのために、連続値を含む、離散でも無限個の値を取るといった一般の2変数に関する相互情報量の推定方法を検討する。そして、その推定量が、サンプル数とともに真の相互情報量に収束することを証明する。次に、ゲノム解析の RNA シーケンスから得られたデータから、複数の遺伝子の発現量の間の因果関係を表現する森を構築することを検討する。最後に、その理論を条件付き相互情報量の推定の場合に一般化し、与えられたデータに対して事後確率を最大にするベイジアンネットワークを構築する方法を検討する。最後に、マイクロアレイから得られた森と RNA シーケンスで得られた森を比較する。さらに、既存のグラフィカルモデルを用いない遺伝子変異解析と比較した場合のメリットを確認する。

3. 研究の方法

4年の計画であったが、コロナの関係で延長を2回行った。期間を通して、当初の予定通り、下記の3課題について検討した。

課題1 連続変数を含むなど一般的な2変数に適用できる相互情報量の推定量 $J(n)$ が、サンプル数 n とともに $I(X, Y)$ に収束することを数学的に証明する: 研究代表者の数学のスキルに依存する部分が多い。真の分布が、ヒストグラムの幅を十分に小さくすることによって得られることを仮定して証明する。現時点では、証明は完成していないが、詳細な方針までできあがっている。

課題2 ノンパラメトリック的方法の汎用性のメリット、およびゲノム解析での有用性を主張するために、RNA シーケンスのカウントデータ(無限個数の値)もしくはFPKM(連続値)から森の構造を学習する方法を確立する。特にサンプル数 n が小さい場合の挙動を確認する: マイクロアレイから得られた発現量のデータ(連続値)から森を構築することは、すでに検討している。カウントデータの値の分布は、遺伝子によってまちまちで、有限個の値をとる離散変数として扱うことは困難であるとされている。

課題3 独立性および相互情報量の概念を拡張し、条件付独立性および条件付相互情報量についても、推定量 $J(n)$ が真の値 $I(X, Y | Z)$ に収束し、確率1で $J(n)=0$ と Z のもとで X, Y が条件付き独立であることが同値であることを数学的に証明するとともに、ゲノム解析の問題を通じて有効性を実証する: アルゴリズムは準備している。相互情報量 $I(X, Y)$ を推定する際に、 XY 平面をメッシュに区切り、量子化された相互情報量の推定量を求めたが、これをさらに Z 軸についても区切り、 $I(X, Y | Z)$ の量子化された条件付相互情報量の最大値を計算するということになる。また、 n が大きい場合に、その推定量が真の値 $I(X, Y | Z)$ に確率1で収束することは、課題1が解決すれば、証明できるものと思われる。そして、条件付き相互情報量の推定ができれば、その値から、事後確率を最大にするベイジアンネットワークを構築することが可能となる。

しかし、 n が小さい場合に、相互情報量を求める場合と比較して、メッシュの各区切りでサンプルが少なくなるため、推定が難しくなる。ただ、ゲノム解析でFPKM(連続値)ではなく、カウントデータ(無限個の値をとる)の場合、そのような問題が生じないものと思われる。また、前者も連続値であっても正規化しただけなので、離散データの場合と同様の挙動を示す可能性がある。

4. 研究成果

本研究は、連続変数を含むデータの相互情報量推定と、その結果を用いたグラフィカルモデルの構築に焦点を当てている。相互情報量は、二つの確率変数の依存関係を定量化する重要な指標であり、データ解析や機械学習において広く利用されている。

2018年度は、基本的なアルゴリズムの開発を進め、相互情報量の一貫性と独立性検出の両方を満たす手法を確立した。また、欠損データを含む場合の一般化も行い、IEEE trans. on

Information Theory などのジャーナルに掲載した。

2019 年度は、多変量データから変数間の依存関係を評価し、事後確率を最大化するグラフィカルモデルの構築を検討した。連続変数を含むデータに対して、相互情報量推定量がサンプル数とともに真の値に収束することを数学的に証明し、その成果を国際会議 (IEEE ISIT) で発表した。

2020 年度は、交絡の存在を許容する因果探索に関する研究を進めた。連続および離散変数に対して、交絡を考慮した因果探索アルゴリズムを開発した。また、ベイズ測度を求めるための事前確率に関する研究も行い、その成果を Weiley の「Wire's Computational Statistics」にレビュー論文として発表した。

2021 年度は、交絡の存在を考慮した因果順序の識別に関する研究を行い、「Causal Order Identification to Address Confounding: Binary Variables」を Behaviormetrika に発表した。この研究では、連続および離散変数の条件付き相互情報量の推定方法を提案し、実証している。

2022 年度は、条件付き相互情報量の概念を拡張し、その推定量が真の値に収束することを数学的に証明した。この成果により、信頼性の高い相互情報量の推定が可能となり、ベイジアンネットワークの構築手法が確立された (Yang & Suzuki, 2022)。

2023 年度は、実データへの適用と応用に注力し、提案手法の実用性と効果を検証しました。Tianle Yang と共著した論文が Behaviormetrika に採択され、研究の信頼性と国際的な評価を高めた (Yang & Suzuki, 2023)。

以上が、2018 年から 2023 年にかけての研究成果の概要である。本研究の成果は、多くのデータ解析分野での応用が期待され、今後の研究や実務において重要な基盤となることが期待される。なお、そうした成果は、

5. 主な発表論文等

〔雑誌論文〕 計18件（うち査読付論文 17件 / うち国際共著 7件 / うちオープンアクセス 5件）

1. 著者名 Shimmura Ryosuke, Suzuki Joe	4. 巻 5
2. 論文標題 Newton-Type Methods with the Proximal Gradient Step for Sparse Estimation	5. 発行年 2024年
3. 雑誌名 Operations Research Forum	6. 最初と最後の頁 1-1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s43069-024-00307-x	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yang Tian-Le, Suzuki Joe	4. 巻 7
2. 論文標題 Dropout drops double descent	5. 発行年 2024年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 1-1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-024-00242-5	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Islam Ashraful, Suzuki Joe	4. 巻 51
2. 論文標題 Forest construction of Gaussian and discrete variables with the application of Watanabe Bayesian Information Criterion	5. 発行年 2024年
3. 雑誌名 Behaviormetrika	6. 最初と最後の頁 1-1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41237-024-00227-4	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Yang Tian-Le, Lee Kuang-Yao, Zhang Kun, Suzuki Joe	4. 巻 51
2. 論文標題 Functional linear non-Gaussian acyclic model for causal discovery	5. 発行年 2024年
3. 雑誌名 Behaviormetrika	6. 最初と最後の頁 1-1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41237-024-00226-5	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Shimmura Ryosuke, Suzuki Joe	4. 巻 26
2. 論文標題 Estimation of a Simple Structure in a Multidimensional IRT Model Using Structure Regularization	5. 発行年 2023年
3. 雑誌名 Entropy	6. 最初と最後の頁 44 ~ 44
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/e26010044	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 鈴木 讓	4. 巻 53
2. 論文標題 スパース推定における選択的推論	5. 発行年 2023年
3. 雑誌名 日本統計学会誌	6. 最初と最後の頁 139 ~ 167
掲載論文のDOI (デジタルオブジェクト識別子) 10.11329/jjssj.53.139	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Zhang Bingyuan, Suzuki Joe	4. 巻 25
2. 論文標題 Extending Hilbert-Schmidt Independence Criterion for Testing Conditional Independence	5. 発行年 2023年
3. 雑誌名 Entropy	6. 最初と最後の頁 425 ~ 425
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/e25030425	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Shimmura Ryosuke, Suzuki Joe	4. 巻 5
2. 論文標題 Converting ADMM to a proximal gradient for efficient sparse estimation	5. 発行年 2022年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 725 ~ 745
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-022-00150-6	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tianle Yang, Joe Suzuki	4. 巻 186
2. 論文標題 The Functional LiNGAM	5. 発行年 2022年
3. 雑誌名 PMLR (Probabilistic Graphical Models)	6. 最初と最後の頁 25-36
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Suzuki Joe, Inaoka Yusuke	4. 巻 49
2. 論文標題 Causal order identification to address confounding: binary variables	5. 発行年 2021年
3. 雑誌名 Behaviormetrika	6. 最初と最後の頁 5~21
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41237-021-00149-5	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Chen Jie, Suzuki Joe	4. 巻 9
2. 論文標題 An Efficient Algorithm for Convex Biclustering	5. 発行年 2021年
3. 雑誌名 Mathematics	6. 最初と最後の頁 3021~3021
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/math9233021	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Chen Jie, Shimmura Ryosuke, Suzuki Joe	4. 巻 23
2. 論文標題 Efficient Proximal Gradient Algorithms for Joint Graphical Lasso	5. 発行年 2021年
3. 雑誌名 Entropy	6. 最初と最後の頁 1623~1623
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/e23121623	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Shimmura Ryosuke, Suzuki Joe	4. 巻 1
2. 論文標題 Converting ADMM to a proximal gradient for efficient sparse estimation	5. 発行年 2022年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-022-00150-6	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Suzuki Joe	4. 巻 1
2. 論文標題 Why BDeu? Regular Bayesian network structure learning with discrete and continuous variables	5. 発行年 2021年
3. 雑誌名 WIREs Computational Statistics	6. 最初と最後の頁 1-1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/wics.1554	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Suzuki Joe	4. 巻 1
2. 論文標題 Mutual Information Estimation: Independence Detection and Consistency	5. 発行年 2019年
3. 雑誌名 IEEE International Symposium on Information Theory	6. 最初と最後の頁 1-1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ISIT.2019.8849612	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 鈴木 譲	4. 巻 65
2. 論文標題 統計学の使い方より、本質を見抜く力 機械学習の数理 100問シリーズと凸最適化への期待	5. 発行年 2020年
3. 雑誌名 オペレーションズ・リサーチ	6. 最初と最後の頁 551-556
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Joe Suzuki	4. 巻 1
2. 論文標題 Branch and Bound for Continuous Bayesian Network Structure Learning	5. 発行年 2018年
3. 雑誌名 Probabilistic Graphical Models	6. 最初と最後の頁 1-1
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Suzuki, Joe	4. 巻 64
2. 論文標題 Forest Learning From Data and its Universal Coding	5. 発行年 2018年
3. 雑誌名 IEEE Transactions on Information Theory	6. 最初と最後の頁 7349 ~ 7358
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TIT.2018.2869215	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計14件 (うち招待講演 3件 / うち国際学会 3件)

1. 発表者名 Suzuki, Joe
2. 発表標題 Solve 100 Problems of Math and R/Python for Statistical Learning
3. 学会等名 IASC-ARS2022 (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 鈴木 讓
2. 発表標題 統計学の使い方より本質を見抜く力
3. 学会等名 統計関連学会連合大会 (招待講演)
4. 発表年 2020年

〔図書〕 計16件

1. 著者名 Joe Suzuki	4. 発行年 2022年
2. 出版社 Springer	5. 総ページ数 208
3. 書名 Kernel Methods for Machine Learning with Math and Python	

1. 著者名 Joe Suzuki	4. 発行年 2022年
2. 出版社 Springer	5. 総ページ数 196
3. 書名 Kernel Methods for Machine Learning with Math and R	

1. 著者名 Suzuki, Joe	4. 発行年 2021年
2. 出版社 Springer	5. 総ページ数 240
3. 書名 Sparse Estimation with Math and R: 100 Exercises for Building Logic	

1. 著者名 Suzuki, Joe	4. 発行年 2021年
2. 出版社 Springer	5. 総ページ数 250
3. 書名 Sparse Estimation with Math and Python: 100 Exercises for Building Logic	

1. 著者名 Suzuki, Joe	4. 発行年 2021年
2. 出版社 Springer	5. 総ページ数 220
3. 書名 Statistical Learning with Math and Python: 100 Exercises for Building Logic	

1. 著者名 鈴木 譲	4. 発行年 2021年
2. 出版社 共立出版	5. 総ページ数 200
3. 書名 機械学習のためのカーネル100問 with R	

1. 著者名 鈴木 譲	4. 発行年 2021年
2. 出版社 共立出版	5. 総ページ数 216
3. 書名 機械学習のためのカーネル100問 with Python	

1. 著者名 Joe Suzuki	4. 発行年 2020年
2. 出版社 Springer	5. 総ページ数 220
3. 書名 Statistical Learning with Math and R	

1. 著者名 鈴木 讓	4. 発行年 2020年
2. 出版社 共立出版	5. 総ページ数 272
3. 書名 統計的機械学習の数理100問 with Python	

1. 著者名 鈴木 讓	4. 発行年 2020年
2. 出版社 共立出版	5. 総ページ数 240
3. 書名 スパース推定100問 with R	

1. 著者名 鈴木 讓	4. 発行年 2021年
2. 出版社 共立出版	5. 総ページ数 260
3. 書名 スパース推定100問 with Python	

1. 著者名 鈴木 讓	4. 発行年 2020年
2. 出版社 共立出版	5. 総ページ数 256
3. 書名 統計的機械学習の数理100問 with R	

1. 著者名 Joe Suzuki	4. 発行年 2023年
2. 出版社 Springer	5. 総ページ数 239
3. 書名 WAIC and WBIC with Math and R Stan	

1. 著者名 Joe Suzuki	4. 発行年 2023年
2. 出版社 Springer	5. 総ページ数 254
3. 書名 WAIC and WBIC with Math and Python Stan	

1. 著者名 鈴木 讓	4. 発行年 2023年
2. 出版社 共立出版	5. 総ページ数 238
3. 書名 渡辺澄夫ベイズ理論 with R and Stan	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------

タイ	Mahidol University	Chulalongkorn University		
----	--------------------	--------------------------	--	--