

九州大学統計グループ  
2021年度 統計データサイエンス研究集会 講演概要

2021年12月18日[土] Zoomミーティングによるオンライン開催

開催趣旨

データサイエンスの根幹となっている統計学は、古くから生物学, 医学, 経済学, 社会科学等の分野に活用されてきました。最近では、環境, エネルギー, 食糧問題, 貧困対策など、持続可能な社会に向けた統計解析がますます重要となっています。本研究集会では、理論から実践まで幅広い分野の先生方にご講演いただく予定です。講演を通じて他領域への理解を深めることにより、これからのデータサイエンスの時代に、どのような統計解析手法および数理統計の理論を展開すべきかを考えることができればと思っております。

研究集会幹事 廣瀬 慧

アブストラクト (プログラム順)

間野 修平 統計数理研究所 数理・推論研究系

非交叉分割, 自由独立性とランダム行列

自由独立性の概念とその考察に非交叉分割が有効に使われることを説明し、ランダム行列を通して統計学や機械学習に応用されていることを紹介する。

藤森 洸 信州大学 経法学部応用経済学科

The variable selection by the Dantzig selector for Cox's proportional hazards model

本講演では高次元・スパースな設定におけるCox 比例ハザードモデルに焦点を当てる。回帰パラメータをDantzig selectorにより推定し、得られる推定量が一致性及び変数選択の一致性を適当な条件で満たすことを示す。さらに、この事実を用いて次元の削減を行い、回帰パラメータ及び累積ベースラインハザード関数の漸近正規推定量を構成する。また、正則化パラメータの選択方法を含む、比例ハザードモデルに対するDantzig selector 型推定量の数値計算方法についても言及する。

岡 檀 統計数理研究所 医療健康データ科学研究センター

COVID-19感染拡大による生活変化の把握と対策に資する質的/量的混合アプローチ

演者は、コミュニティ特性が住民の心身の健康に与える影響について質的/量的混合アプローチによる研究を行ってきた。COVID-19感染拡大による生活変化を把握し、対策に資する提言を行うことを目的に、これまでに行った一連の研究成果を報告する。

1. コロナ後の自殺率上昇の地域差及び性差—全国市区町村のパネルデータ解析から
2. 子どもの心の健康を促進/阻害する要因—コホートスタディによる探索
3. 町の空間構造特性と援助希求行動—「自然実験」をふまえた分析

Dou Xiaoling 早稲田大学 データ科学センター

### Nonlinear causality for CHARN models

We consider nonlinear Granger causality with a highly flexible nonlinear time series model, the conditional heteroscedastic autoregressive nonlinear (CHARN) model. We show that the causality of the CHARN models can be examined by a Portmanteau test based on a constrained maximum likelihood estimator of the parameters, and the test statistic has an approximate asymptotic Chi-square distribution. We describe the Chi-square asymptotics of the Portmanteau test for a CHARN model, provide calculations of the test statistic and investigate the performance of the Portmanteau test using a simulation. This idea is also illustrated using a real data set.

川口 淳 佐賀大学 医学部

### 医療データサイエンス

医学において診断や治療の客観的な判断材料としてデータを科学的に評価することは必要とされている。そのために統計学を中心とした方法論の開発が活発に行われている。そこで最も重要視されている点の一つとして医療に還元できる情報をデータから得ることがある。効果的な治療のための早期発見、迅速な社会還元のための早期判断、など予測という考え方が必要とされる場面とその解析例を紹介する。また、これらの事においても、そしてさらに探索的発見を行うために必要とされるビッグデータについて、予測精度の高いことに加え、医療において説明性を備えた機械学習が必要という事を強調し、実用性を重んじた解析例を紹介する。

太田 真理 九州大学 大学院人文科学研究院

### 神経科学者から見た脳活動データ解析のこれまでとこれから

近年の脳機能計測技術の急速な発展により、大量の脳活動データが容易に収集できるようになった。その一方で、データ解析ではさまざまな手法が乱立しており、標準的な解析手法は確立していない。この原因は、脳活動データの持つ、時間的・空間的な依存性、多次元性、信号/ノイズ比の低さ、などの特性にあると考えられる。本発表では、機能的磁気共鳴画像法 (fMRI) と脳波・脳磁図 (M/EEG) という、代表的な脳機能計測手法を対象に、データ解析の非専門家 (神経科学者) から見た現状と課題について概観する。さらに近年利用機会が増えている、機械学習を用いた手法や、ベイズ推定を用いた手法についても簡単に紹介する。

松井 秀俊 滋賀大学 データサイエンス学部

### 関数回帰モデルに対するドメイン選択と収量データ分析への応用

各観測が経時的に測定されたデータを関数として表現したデータは、関数データとよばれている。本報告では、関数データとして与えられた説明変数と目的変数との関係を表す関数回帰モデルの1つである変化係数関数線形モデルにおいて、目的変数と関連する説明変数の定義域を選択するドメイン選択の方法について紹介する。提案手法を農業データの分析に適用することで、農作物の収量にどの時点での環境要因が関連しているかを選択することを試みる。

### 逐次データ同化技術と変分推論の協働によるモデル推論

観測時系列データからデータ駆動でモデルを推論する研究は、制御・統計・機械学習の分野で発展してきている。本発表では、近年の深層学習の潮流から生まれた Sequential VAE (SVAE) という手法群に焦点を当てる。SVAE は、潜在変数の遷移構造を VAE に取り入れることで、時系列生成モデルの推論を可能にしており、音楽・言語・生物等の分野で使われている。この SVAE に逐次データ同化技術を導入することで推論精度を理論的・実験的に向上させる先行研究を紹介し、先行研究の欠点を克服した発表者らの提案枠組みについても解説する。

### 多変量線形ガウス状態空間モデルを使った当日予測とその高速化

本発表では、多変量線形ガウス状態空間モデルの1つである、一見無関係な時系列方程式モデル(SUTSE model) を用いた時系列データの予測について考える。例えば、バスの混雑度データなどは、1日ごとに、時刻表の数の次元を持つ多変量混雑度データが得られていると考えられ、このモデルを適用することで、カルマンフィルタを用いた1日先の混雑度の予測を行うことができる。一方で、予測を行いたい日のうち、いくつかの時刻表の混雑度データが既に得られているとき、これらと1日先予測との誤差を用いて、残りの時刻表の混雑度の1日先予測を修正するような予測が考えられる。この予測の修正項は、1日先予測誤差の共分散をもとに導出され、我々は当日予測と呼んでいる。しかし、これらの予測には、特に高次元である際に、計算コストが大きという問題点がある。この問題は、カルマンフィルタに出てくる行列計算に起因する。そこで我々は、高速化法として、多変量時系列データの相関構造を一旦無視し、各時刻バラバラに解析を行なった。こうすることで、カルマンフィルタの行列計算のコストが削減できる。さらに、我々は、この高速化法における1日先予測誤差の期待値が、適当な正則条件のもとで、実際は真のモデルが相関を持っていたとしても0に収束し、共分散も、ある行列に収束することを証明した。また、高速化法においても、当日予測の修正項の導出には、1日先予測誤差の共分散が必要となる。一方で、高速化法において、この値は陽に得られない。高速化法では相関を全く無視して解析を行なっているため、このような不具合が生じる。そこで、1日先予測誤差の共分散の収束先を推定し、その推定量を使うことで、高速化法においても当日予測を精度良く行う方法を提案した。本発表では、当日予測について説明したのち、高速化手法の研究について発表を行う。

### 無限分散回帰モデルの推定

非対称安定分布を誤差とする線形回帰モデルの推定を考える。ここでは Nolan (1998) により提案された、全パラメータ全てに関して連続となるような安定分布のパラメトリゼーションを用いる。推定方法としては最尤法も考えられるが、数値的な計算負荷が大きという欠点があり、高次元・大標本の場合には実用的とはいえない。そこで本発表では、モーメント法と最小絶対偏差推定法に基づいた計算容易な段階的推定方式を提案し、関連する理論的性質および数値実験結果について報告する。