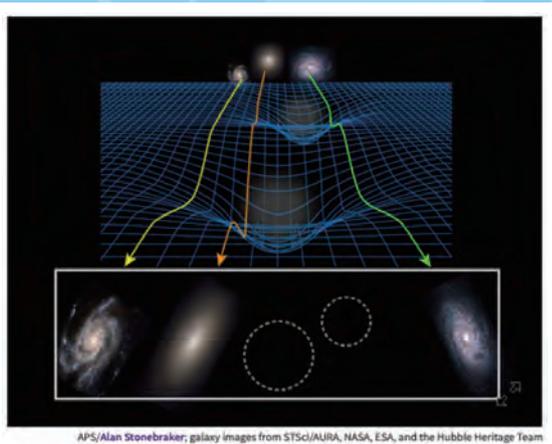


## ▼ CONTENTS

- 02 響き合う人とデーター 統数研プロジェクト紹介  
第31回「データ科学と天文学の境界領域の進展を目指して」
- 06 研究教育活動  
新入教員紹介  
2022年7月-9月の公開講座実施状況  
統計数理セミナー実施報告(2022年9月~10月)  
統計数理研究所夏期大学院「データ同化夏の学校」の開催  
連続最適化および関連分野に関する夏季学校2022実施報告  
「探索的ビッグデータ解析と再現可能研究」実施報告  
「統計サマーセミナー 2022」実施報告
- 09 統数研トピックス  
第4回赤池メモリアルレクチャー賞受賞記念講演開催  
第2回「統計エキスパート育成システムの構築」に向けたワークショップを開催  
「問題解決プロセス」に着目した、社会に活かす統計の考え方ワークショップの開催  
田中未来准教授が日本オペレーションズ・リサーチ学会第12回研究奨励賞を受賞  
兵庫高校研修オンライン講義を実施
- 12 総合研究大学院大学複合科学研究科統計科学専攻関係  
SOKENDAI賞・研究科長賞授賞式
- 13 共同利用  
2022年度共同利用公募追加課題
- 13 外部資金・研究員等の受入れ  
受託研究・受託事業等の受入れ  
外来研究員の受入れ  
寄附金の受入れ
- 14 人事
- 15 刊行物  
Research Memorandum (2022.8 ~ 2022.10)  
統計数理研究所調査研究レポート  
研究教育活動報告  
Annals of the Institute of Statistical Mathematics
- 16 コラム

大学共同利用機関法人 情報・システム研究機構  
 統計数理研究所 ニュース



APS/Alan Stonebraker, galaxy images from STScI/AURA, NASA, ESA, and the Hubble Heritage Team

# 天文学のビッグデータ 統計手法を応用



▲白崎正人助教



▲服部公平助教

## 銀河の“ひしゃげ”を推定し ダークマターの宇宙地図を描く

2020年秋、国立天文台の新進気鋭の二人の若手研究者、白崎正人助教と服部公平助教が統計数理研究所統計思考院に兼務で着任した。統計思考院は、データに基づく推論である統計思考力を鍛えるための開かれた「道場」。統計学のシニア教員や異分野の研究者との交流によって、各自の研究テーマに援用できる統計手法のアイデアやヒントを掴む場とすることがねらいだ。

白崎が手掛けているのは、人工知能(AI)を使った深層学習による宇宙の「大規模構造」の解明だ。天文学に深層学習を応用した研究は世界で急速に進んでおり、2015年には10本程度だった論文

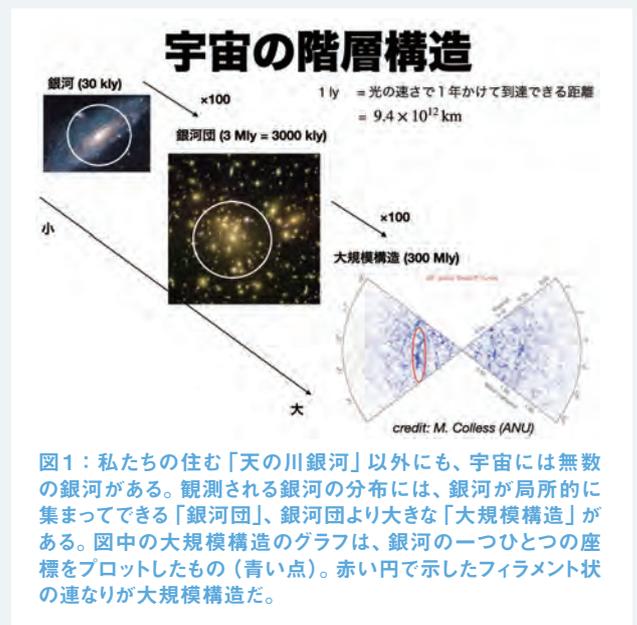
投稿数が2021年には1000本近くと、6年で100倍になっているという。白崎は以前から統数研の池田思朗教授らとこのテーマで共同研究を進めており、その研究成果は『英国王立天文学会誌』の2021年6月版に掲載された。概要は次のとおりだ。

太陽のように自ら光を発する「恒星」の集まりを「銀河」と呼ぶ。銀河が1000個ほど集まった状態が「銀河団」で、その大きさは銀河の100倍程度。端から

端まで、光の速さで進んでも300万年かかる「300万光年」の規模だ。「宇宙の『大規模構造』は、銀河が連なって節状や泡状の構造を作る状態を指します。典型的な大きさは、銀河団の100倍になります」と白崎は説明する(図1)。

大規模構造は宇宙誕生後137億年をかけて、重力によって成長してできたと考えられる。当初はある範囲内に物質の集まりのわずかな濃淡があり、そのうちの高密度領域に物質が引き寄せられて大規模構造を形成していった。

この物質を引き寄せる過程で重要な役割を果たしたのが「ダークマター(暗黒物質)」だ。だが、その正体は未だ解明され



# 解析に

天文学や宇宙物理学の分野では、観測機器の進化とともにデータ量が爆発的に増えつつある。今や宇宙の解明には、ビッグデータを扱うデータサイエンスの知見が不可欠となった。統計学の手法を天文学に応用して研究を進化させることを目指し、現在、国立天文台から二人の若手研究者が統計数理研究所統計思考院へ出向している。

ていない。「大規模構造の形成に関わるダークマターをなんとか捕まえたい。私が追いかけているのはこの課題です」と白崎は話す。

ダークマターの正体を解明するために

は、宇宙のどこにどれくらいのダークマターがあるか調べ、「地図」を作成する必要がある。ダークマターはその性質によって、分布の仕方が変わると理論的に予測されているからだ。

約1億天体のビッグデータ。データ規模が大きいと、解析には高度な技術が必要になる。

「やっかいなのは、銀河の形は一定ではなく、もともとひしゃげていることです」と白崎は言う。元が完全な球形をした銀河なら、ひしゃげた形から推測したダークマターも明確な分布を描く。しかし、銀河の形が分からないと、推定されたダークマターはノイズが入り、迷彩模様のようにはっきりしない分布になってしまう(図3)。

この状態から「ノイズだけを取り除きたい」というニーズに対応できるのが、深層学習だ。まずシミュレーションによって大規模構造をいくつも作成する。次に、コンピュータ上に形成された模擬宇宙に、重力レンズ現象をシミュレーションしておく。そうして得られた「ノイズあり疑似データ」と「ノイズなし疑似データ」の組み合わせを3万セットほどAIに学習させ、「ノイズあ

## 重力レンズ効果

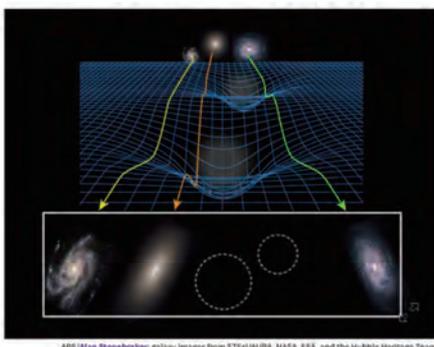


図2：遠方にある銀河は、手前にあるダークマターによって、地球からはひしゃげて見える。この「遠方銀河の弱重力レンズ効果」と呼ばれる性質を利用して、ダークマターの分布を推定する。

ダークマターの地図を作成するために近年注目されているのが、「重力レンズ効果」と呼ばれる現象を利用する方法だ。ダークマターはその重力によって時空を歪め、レンズのような効果をもたらす。そのせいで、地球からは遠方の銀河の形がひしゃげて見える(図2)。この性質を利用して、銀河の手前にあるダークマターを推測する。

白崎のチームが使用したのは、国立天文台がハワイで運用する「すばる望遠鏡」で300晩にわたり撮影、公開されている

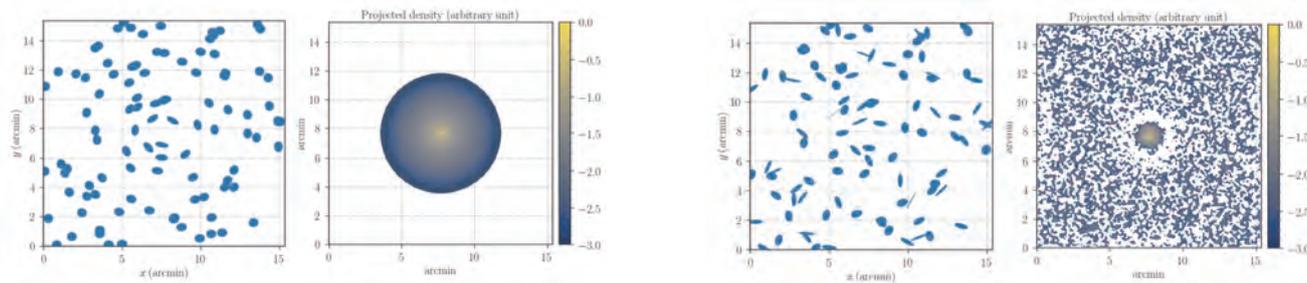


図3：左図は銀河が完全な球形だった場合で、ひしゃげた形から推測したダークマターも明確な分布を描く。しかし、銀河の元の形が分からないと、推定されたダークマターはノイズが入り、右図のようになってしまう。

り]を「ノイズなし」に変換するAIを構築する。「このAIに実際の観測データを適用し、ほぼ期待どおりのノイズ除去性能が達成できました」と白崎は話す(図4)。

論文の査読段階では何度もフィードバックがあり、その対応に約2年を費やしたものの、そこで得られた知見もあった。「局所的なノイズ除去は得意な半面、大域的になると苦手であるなど、AIの癖が分かってきました。すべてのダークマターに適応させるのは難しいかもしれませんが、他の手法では不可能な場所への適用が増えれば、大規模構造の解明に一步近づきましょう」と白崎は期待する。

## 楽観的に、貪欲に。観測誤差を乗り越えて離散した兄弟星を探せ!

一方、服部の研究対象は、太陽系が属している銀河、すなわち「天の川銀河(銀河系)」の歴史だ。「宇宙全体の歴史を世界史に例えると、私の手掛けている研究は日本史のようなものです」と服部は説明する。

天の川銀河は、恒星が1000億個ほども集まった巨大な円盤状の集合体だ。太陽はその中を秒速240kmで円運動している。円の軌道を1周するのにかかる時間は2億年と聞けば、「日本史」といってもそのスケールの大きさが伺えよう。

なぜ、天の川銀河はこれほど巨大なのか。「今、最も支持されているのは、小さな銀河が次々にぶつかって一体化し、大きくなったというシナリオです」と服部は言う。天の川銀河の周囲には、大マゼラン銀河や小マゼラン銀河など、子分のような矮小銀河がたくさんある。天の川銀河も当初はこうした小さな銀河だったが、そこに他の銀河がぶつかり巨大化したと考えられる(図5)。その歴史を解明しようと

というのが、服部の研究テーマだ。

「天の川銀河は私たちの住む銀河なので、他の銀河に比べて圧倒的に情報が入手しやすい。今は天文学史上、革命的といえるほどのビッグデータが得られる時代になっています。天の川銀河は無数にある銀河の一つですから、これを詳細に調べることで、宇宙全体の構造や歴史の解明につなげられるのです」(服部)。

例えば、恒星の化学組成を調べることで、その星が生まれた場所(星雲)の環境が分かる。天の川銀河に属する恒星の化学組成については世界中で大型望遠鏡を駆使した観測が進み、すでに恒星100万個分ほどのデータが集まっている。

また、観測用の人工衛星「Gaia」のデータからは恒星の位置と速度が分かる。現在、天の川銀河の1%にあたる恒星10億個分のデータが公開されている。

宇宙の年齢は137億歳といわれる。一方で、現存する恒星の中には100億歳を超えるものも多くあり、しかもその化学組成や軌道は誕生時からほぼ変わらない。つまり、現在の恒星を調べれば、その星が生まれた星雲の化学組成や軌道が解明できるわけだ。「いうなれば、星は『記憶力のよい歴史の生き証人』なのです」と服部は話す。

データを解析することで、天の川銀河への衝突によって離散した恒星同士が、

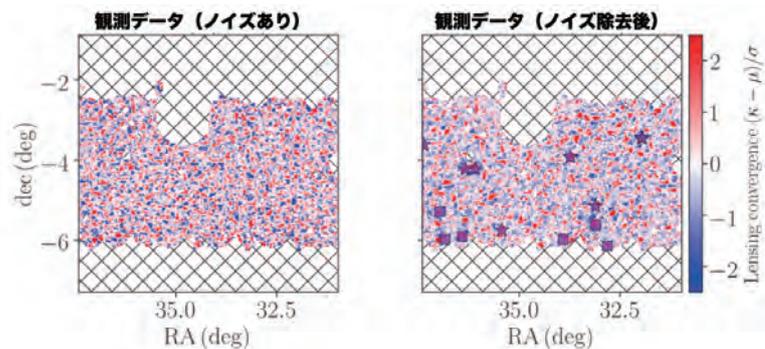


図4: 深層学習によって構築したAIで、観測データを「ノイズあり」を「ノイズなし」に変換し、ノイズを除去した。



図5: 当初は小さな銀河だった「天の川銀河」に他の銀河や星団が次々にぶつかった結果、現在の大きさに成長したと考えられている。

元はどの矮小銀河に所属していたかも突き止められる。恒星の軌道形状は、位置と速度から算出する「角運動量」で表すことができ、この角運動量が同じであれば、同じ銀河を母体とする兄弟星だったことになる(図6)。

服部が統数研で奥野彰文助教と共に取り組んでいるのは、Gaiaのビッグデータを使ったクラスタリング解析だ。観測データには必ず誤差があることから、角運動量の正しい推定は難しく、そのままではうまくグルーピングできない。服部らの開発した手法は、誤差があっても、データの構造を利用して兄弟星のグループを見つけることができるものだ(図7)。

こうした手法は、必ずしもクラスターがあることを前提とせず、「もしクラスターがあるとしたら、どこにあるか」という問いの立て方をすることから「楽観的なクラスタリング」と呼ばれる。服部は「膨大なデータをどう扱えばいいかわからないが、とにかくグループを探したい、という天文学のモチベーションに合った方法」だと感じているという。服部と奥野は、この新たな手法を「Greedy Optimistic Clustering (貪欲楽観クラスタリング)」と名付けた。

### 天文学と統計学のマッチングで宇宙の解明は新たな展開へ

白崎と服部が着任して感じたのは、「統数研は自然科学の分野でも、もっと存在感があっていいはず」ということだ。

「統数研ではクラシカルな統計手法の応用は進んでいる一方、AIに関しては数学的な理論の構築が中心で、応用の広がりはこちらから感じる。私たち天文学者が気づいていない手法がまだまだたくさんあると思うので、天文学に活用できるものを出向期間中に見つけたい」と白崎は言う。

例えば、大規模構造の研究ではよく「乱数」の壁にぶち当たる。観測データから宇宙を小さなグリッドで区切ったとき、その中にある銀河の数は規則性のない乱数だ。そこで、乱数の現れる頻度を統計的に推定する「カーネル密度推定」などが応用できないかと考えている。「サイコロを一度に100個振ったときの値の頻度分布を求めるような高次元のデータの頻度推定などに強い方は、ぜひご連絡いただきたい」と白崎は呼びかける。

統計思考院は、自由な雰囲気の中で創発が生まれることを期待し、広いフロアにリラックスできるスペースが用意されている。服部と奥野の共同研究も、たまたま通りかかった奥野が服部の話に興味を示したことがきっかけで始まった。しかし、

コロナ禍の影響で在宅勤務の研究者が多く、以前ほど偶発的な出会いの機会は多くないのが実情だ。

「統計手法の引き出しが増えれば、Gaiaのビッグデータを使ってこんなことができるのでは、というアイデアも生まれやすくなるはず」と服部は期待する。「例えば、『異常値推定』や『極値統計』などの手法を使えば、観測された膨大な天体の中から、レアな性質を持つ天体を抽出することができるのではないかと。また、ベイズ統計で尤度を近似でも高速で計算する手法も知りたい」と(服部)。

こうした天文学の問題に関連のありそうな知見を持つ研究者との出会いを、二人は待ち望んでいるところだ。

(広報室)

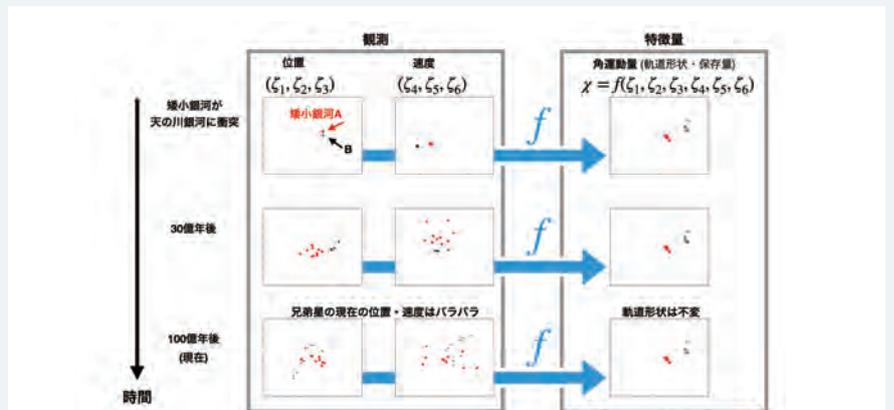


図6：矮小銀河が天の川銀河に衝突して破壊される様子。別々の位置に離散しても、軌道形状は変わらないため、データを解析することで兄弟星を探し当てることができる。

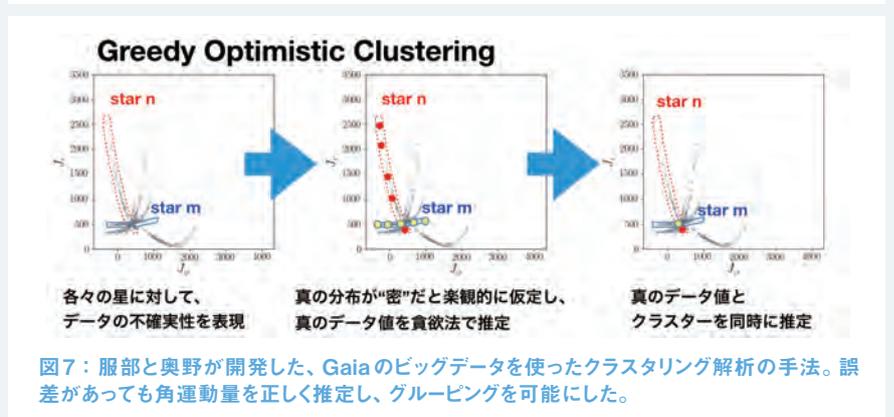


図7：服部と奥野が開発した、Gaiaのビッグデータを使ったクラスタリング解析の手法。誤差があっても角運動量を正しく推定し、グルーピングを可能にした。

新型コロナウイルス感染拡大防止のため、本インタビューはオンラインで行われました。



## Tam Le

数理・推論研究系 助教

新入 1  
教員紹介

I have officially been an assistant professor at The Institute of Statistical Mathematics (ISM) since September 1st, 2022. I am privileged to become a faculty member at ISM. My research interests are to develop machine learning algorithmic approaches which can learn from complex data (e.g., non-Euclidean, large-scale, high-dimensional, non-linear, corruption) and take into account social demands (e.g., fairness) and properties of applications (e.g., under covariate perturbations, low-sample-size) to output insightful recommendations. I have recently focused on optimal transport geometry for probability measures by leveraging local structures, e.g., tree structure, or graph structure.

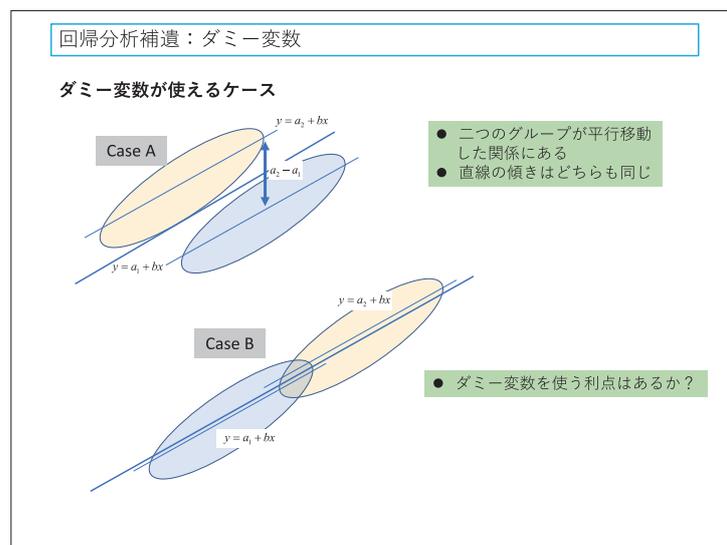
Overall, ISM provides a great research environment for me to pursuit high-level research in machine learning and statistics. I will try to contribute to the research visibility and development of ISM.

### 2022年7月-9月の公開講座実施状況

8月4日(木)から9日(火)まで、「多変量解析法」がオンラインにて配信されました。多摩大学の今泉忠講師と、当研究所の馬場康維講師、清水信夫講師が、多変量解析の基礎的・標準的な手法の解説をしました。また、クラスター分析、共分散構造分析などの解説もしました。質問も多数寄せられ、口頭での回答だけでは時間が足りなくなったため、テキストでの回答も併用しました。

9月15日(木)から9月16日(金)まで、当研究所の伊庭

幸人講師、二宮嘉行講師、日野英逸講師、矢野恵佑講師、大久保祐作講師、慶應義塾大学医学部医療システムイノベーション寄附講座所属で当研究所客員の立森久照講師による、リーディングDAT講座「L-A.現代統計学の基礎」がオンラインにて配信されました。「データを扱うための基礎事項からはじめて、早い段階で推測統計の考え方を導入し、現代的な統計モデリングへと進む基礎を固める」ことを主眼とした入門講座です。(情報資源室)



公開講座「多変量解析法」資料より

## 統計数理セミナー実施報告(2022年9月~10月)

毎週水曜16時から所内研究教育職員および外部の方が1人40分ずつ、1日に2人の講演を行っています。2022年9月~10月のセミナーは下記の通り行われました。新型コロナウイルス感染拡大防止のため、統計数理セミナーは現在オンラインで開催しています。

日程	氏名	タイトル
2022年 9月21日	池田 思朗	いて座A*に関するEvent Horizon Telescope の成果について
9月21日	松井 知子	Tukey g-and-h 分布を利用した都市部の熱波の時空間解析
9月28日	庄 建倉	Estimation, diagnostics, and extensions of nonparametric Hawkes processes
9月28日	南 和宏	公的マイクロデータに対する k-匿名化加工の検討
10月 5日	日野 英逸	双対ガンマ冪ダイバージェンスによるロバストな分布混合と能動学習への応用
10月 5日	川崎 能典	GARCH-UGH 法によるバリュー・アット・リスクの推定
10月12日	野場 啓	Lévy 過程に対する、古典的反射戦略および Poisson 的反射戦略の最適性
10月12日	小山 慎介	多変量感染症時系列モデル
10月19日	間野 修平	グラフ上の測度値拡散、衝突粒子系、それらの応用
10月19日	島谷 健一郎	非定常で繰り返し可能なクラスター点過程
10月26日	青木 祐太	マルチタスク学習による高分子/溶媒系の相溶性予測モデルの構築

セミナーの開催予定はホームページにてご案内しています。 <https://www.ism.ac.jp/>

(メディア開発室)

## 統計数理研究所夏期大学院「データ同化夏の学校」の開催

2022年8月8日から10日までの3日間、統計思考院公募型人材育成事業の助成を受け、統計数理研究所夏期大学院として2022年度データ同化夏の学校を開催しました。データ同化夏の学校は、データ同化に関わる若手研究者の育成と実践的な技術の普及を目的として毎夏開催されており、今年で26回目となります。夏期大学院の冠がついた2020年から2年続けてオンライン開催を余儀なくされましたが、今年は青森県むつ市のプラザホテルむつで開催し、46名が現地に集まった他、20名余りがオンラインで参加しました。

今年は、1日目に基礎講義、研究発表、データ同化に関する研究相談を行い、2日目にほぼ1日かけてデータ同化の演習を実施しました。最終日の成果発表会は熱心に演習に取り組んだことがうかがえる優れた内容ばかりで、盛況のうちに終わりました。オンライン参加者への対応も含めて多大なご助力、ご支援をいただいた日本海洋科学振興財団、感染予防策として広い会場の確保にご支援いただいたデータサイエンス共同利用基盤施設、および演習のチューターや講師としてご協力いただいた諸先生方に感謝申し上げます。(中野慎也)



## 連続最適化および関連分野に関する夏季学校2022実施報告

2022年8月8日から10日にかけて、統計思考院公募型人材育成事業として、連続最適化および関連分野に関する夏季学校が開催されました。本夏季学校は、連続最適化とその関連分野における基本的な事項から最先端の動向までを整理・理解し、学生を含む若手研究者の基礎力の養成および新たな研究テーマの発見を目的として2021年から開催しているものです。今年は講師として名古屋産業科学研究所の畔上秀幸先生と中央大学の後藤順哉先生をお招きして、それぞれ偏微分方程式制約つき最適化問題とその応用、ロバスト最適化モデリング入門と題した講義と演習をしていただきました。昨年はZoomのみによる完全遠隔開催でしたが、今年は現地と遠隔の併用開催となりました。現地会場でのポスターセッションでは参加者同士の議論が盛んに行なわれました。やはり現地参加は遠隔参

加とは比較にならない効果があるように思います。一方で、遠隔参加は気軽に参加できるという長所もあります。現地と遠隔を合わせて122名と多くの方にご参加いただくことができました。(田中未来)



## 「探索的ビッグデータ解析と再現可能研究」実施報告

統計数理研究所統計思考院2022年度公募型人材育成事業(ワークショップ)に採択された「探索的ビッグデータ解析と再現可能研究」が、2022年8月28日(日)にオンラインで開催され、126名の参加者を得ました。

昨年度が財務ビッグデータにフォーカスした半日開催であったのに対し、今年度は時空間ビッグデータの処理を念頭に「データの可視化と地図表現」で瓜生真也先生(徳島大学)、「金融ティックデータ前処理」をテーマに小池祐太先生(東京大学)を講師に加え、一日開催となりました。ワークショップ代表者の地道正行先生(関西学院大学)か

らは「R+tidyverseパッケージ速習」と題したチュートリアルの講演を、阪智香先生(関西学院大学)からは所謂ESGの視点からの財務ビッグデータ解析の含意に関する講演を、宮本大輔先生(東京大学)からは、データ駆動科学にフォーカスした仮想化環境mdxにおけるビッグデータ解析環境の構築事例をご紹介いただきました。

質疑応答では、聴衆から講師に具体的な別解法の提案が行われるなど、昨年度以上にインタラクティブなワークショップとなりました。(川崎能典)



瓜生真也先生の講演スライドから

## 「統計サマーセミナー2022」実施報告

統計思考院公募型人材育成事業の一環として開催された統計サマーセミナー 2022は2022年8月6日から8月9日まで、宮崎フェニックス・シーガイア・リゾートとオンラインのハイブリッドで行われました。

本セミナーで重要とすることは、研究発表・討論を通じた若手同士の交流です。多くの研究者・学生が、何らかの個別科学の学部学科等に所属しながら統計科学周辺の研究をしている現在の環境下において、研究早期の段階からいろいろな個別科学での統計の使われ方に触れ、視野を広げることは極めて重要であり、そのひとつの機会を与えるのがこのセミナーの役割となっています。

統計サマーセミナー 2022は2名の招待講演を含む21件の講演がありました。参加者の年齢が近いため、遠慮のない質問・討論が行われるなど活発な交流ができ、若手セミナーの特色が活かされた場となりました。セッション終了後のオンライン懇親会でも議論は夜遅くまで続きました。招待講演では特に学会講演では聴く機会の少ない研究初期段階での問題点やこれまでの研究過程の紹介なども講演者に紹介していただきました。総じて、すべての参加者にとって、実質的に得るところが大きいセミナーとなりました。

(鹿児島大学 吉田拓真)

## 統数研トピックス

### 第4回赤池メモリアルレクチャー賞受賞記念講演開催

第4回赤池メモリアルレクチャー賞受賞者アーボ・ヒバリネン教授（ヘルシンキ大学）を講演者に迎えた記念講演が2022年9月5日（月）に2022年度統計関連学会連合大会のプレナリーセッション「第4回赤池メモリアルレクチャー」として開催されました。

新型コロナウイルス感染症（COVID-19）対策として、今回の統計関連学会連合大会はハイブリッド形式で行われましたが、赤池メモリアルレクチャーでは、ヘルシンキ在住のヒバリネン教授がオンラインで、司会、座長、指定討論者は現地で参加しました。

統計関連学会連合理事長・日本統計学会長の樋口知之教授（中央大学）が司会を務め、椿広計統計数理研究所長が赤池メモリアルレクチャー賞の説明とヒバリネン教授の紹介を行った後、ヒバリネン教授により赤池メモリアルレクチャーが開始されました。

1時間の講演の後は、2名の若手研究者、松田孟留博士（理化学研究所脳神経科学研究センター）と森岡博史博士（理化学研究所革新知能統合研究センター）による指定討論とヒバリネン教授のリジョインダーが行われました。

その後、残りの時間を使って、参加者の質問を受け付けるフロアディスカッションの機会が設けられ、活発な質疑応答が展開され、盛況の内に第4回赤池メモリアルレクチャーが閉幕しました。

現地では64名、オンラインでは231名の総計295名の参加があり、ハイブリッド形式ということからも過去最大の参加者に上りました。未だCOVID-19の影響も強く、前回に引き続き今回も受賞者を日本に招へいすることが叶いませんでしたが、状況が落ち着き次第、直接対面での学术交流の機会があること、赤池メモリアルレクチャーが、今後も国内外の統計科学研究者の交流や若手研究者および統計科学分野のさらなる発展の契機となることを関係者一同、心から祈念します。次回の赤池メモリアルレクチャーは2024年度に開催予定です。（国際連携推進室）



受賞者アーボ・ヒバリネン教授

## 第2回「統計エキスパート育成システムの構築」に向けたワークショップを開催

統計数理研究所が中核機関となる「統計エキスパート人材育成コンソーシアム」では、海外や国内の大学・大学院教育における統計エキスパート育成システム構築に関する情報を共有するため、2022年8月31日（水）に「第2回『統計エキスパート育成システムの構築』に向けたワークショップ～なぜ今、統計エキスパート育成か?～」を開催しました。

ワークショップには、国・地方公共団体や大学・研究機関・民間企業などから合わせて135名が参加し、千野雅人大学統計教員育成センター長の進行のもと、椿広計 統計数理研究所長による「なぜ今、統計エキスパート育成か?」のオリエンテーションにより幕を開けました。

その後、広島大学の小林信一高等教育研究開発センター長による「英国における統計教育をめぐる動向」に続いて、統計数理研究所の国友直人特任教授、長崎大学の西井龍映情報データ科学部長、早稲田大学データ科学センターの堀井俊佑准教授から、取組状況報告・課題提起等

が行われました。

報告後のディスカッションでは、椿広計所長の進行のもと、4人の報告者を中心に熱心な議論が行われ、参加の方々のご支援とご協力をお願いして閉会となりました。

（大学統計教員育成センター）



ディスカッションを進行させる椿所長とワークショップの登壇者

## 「問題解決プロセス」に着目した、社会に活かす統計の考え方ワークショップの開催

2022年9月3日に、統計数理研究所は「『問題解決プロセス』に着目した、社会に活かす統計の考え方ワークショップ」（統計数理研究所、国立大学法人電気通信大学、国立大学法人東京学芸大学、総務省統計局、独立行政法人統計センター共催）を開催しました。椿広計所長によるオープニング「生徒に気づきを与える探究的教材を」の後、東

京学芸大学大学院 西村圭一教授による「Bowland Maths. やCODAPを用いて」、東京学芸大学先端教育人材育成推進機構 山下雅代准教授による「感性を用いた身近な教材」、電気通信大学 鈴木和幸特任教授・名誉教授による「問題解決プロセスの重要性と将来像」のワークショップが行われました。その後、総務省統計局統計情報利用推進課 長藤洋明統計利用推進研究官、統計センター統計技術・提供部 榎田直木部長による「公的統計からの社会課題へのアプローチ」と題した講演が行われました。

最後に、椿所長がモデレータを務めて登壇者によるまとめと質疑応答があり、初中等教育における探求的教材や問題解決プロセスから公的統計まで幅広い視点から活発な意見が交わされました。本ワークショップはオンラインで開催し、253名の事前参加登録と、当日は151名の視聴があり、大変盛況なワークショップでした。なお、本ワークショップは「社会に活かす統計の考え方」シンポジウム（2022年6月16日）のフォローアップとして開催しました。（URAステーション）



## 田中未来准教授が日本オペレーションズ・リサーチ学会第12回研究賞奨励賞を受賞

田中未来准教授が日本オペレーションズ・リサーチ学会第12回研究賞奨励賞を受賞しました。研究賞奨励賞は、独創性と将来性に富み、ORの発展に寄与する研究業績を挙げている若手研究者個人を対象とする賞です。今回の受賞では、数理最適化の理論と応用の両面から独創性の高い研究成果を取めたことが受賞理由とされました。日本オペレーションズ・リサーチ学会2022年秋季研究発表会での表彰式において賞状とメダルが授与されました。

(田中未来)



## 兵庫高校研修オンライン講義を実施

2022年8月1日(月)、兵庫県立兵庫高等学校の2年生24名を対象に統計数理研究所オンライン講義を実施しました。

最初に兵庫高等学校長 福浦潤先生によるご挨拶と、本研究所 山下智志副所長による研究所の概要紹介が行われました。次に兵庫高等学校の生徒より課題意識が発表さ

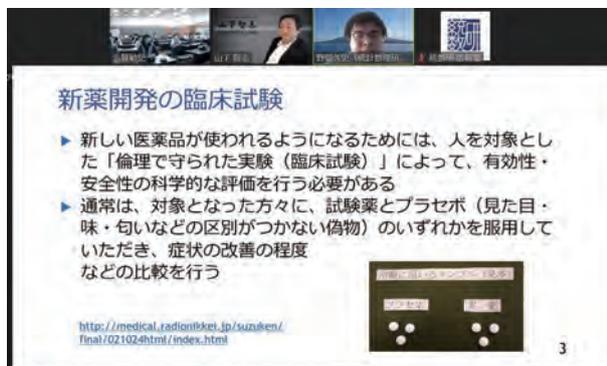
れ、続いてデータ科学研究系 野間久史准教授が統計学の原理・方法と実科学への応用について講義を行いました。新薬開発の臨床試験やWomen's Health Initiative試験を例に仮説検定について講義形式で説明され、講義後は質疑応答が交わされました。

また、生徒からは事前に「大量のデータを扱う中で気を付けていることは」「こういう問題は統計学を用いて問題解決や問題改善が難しいというものがあるのか」等の質問が寄せられており、山下副所長と野間准教授が回答しました。

講義後、生徒からは統計学が日常に関わっていると感じた、自分の研究に今日学んだことを活かせるようにしたい、という感想があり、統計学に触れる貴重な機会となりました。

本講義はコロナ禍により研究所への訪問は行わず、令和4年度東京未来フロンティアツアーの代替研修としてオンラインで開催されました。

(広報室)



オンライン講義の画面

## 入学者選抜試験結果

※5年一貫制は出願なし

## 【3年次編入学】

試験年月日	合格者数
令和4年8月17日(水)~18日(木)	令和4年10月入学(第2回) 8名

## 専攻修了式

令和4年9月21日(水)に、オンライン(Zoom)にて、専攻修了式が行われ、4名が本専攻を修了しました。

## 秋季学位記授与式

令和4年9月28日(水)に、総合研究大学院大学葉山キャンパスにおいて秋季学位記授与式が行われ、本専攻から4名が学位記を授与されました。



## 学位取得者

令和4年9月学位取得者は次のとおりです。

### 【課程博士】

氏名	論文題目
濱口 雄太	ベイズ流メタアナリシスにおける予測区間の評価と影響力解析
貝淵 響	Quantitative Risk Management Using Extreme Value Theory(統計的極値理論に基づく定量的リスク管理)
Wu Pengzhou	Treatment Effect Estimation and Bivariate Causal Discovery via Nonlinear ICA (非線形ICAによる治療効果推定と二変量因果探索)
村瀬 博典	Anomaly Detection by Generating Pseudo Anomalous Data(擬似異常データ生成による異常検知)

## 秋季入学式

令和4年10月4日(火)に、オンライン(Zoom)にて秋季入学式が行われ、本専攻へは8名が入学しました。

(総務課・大学院係)

## SOKENDAI賞・研究科長賞授賞式

令和4年9月28日(水)に、総合研究大学院大学葉山キャンパスにおいてSOKENDAI賞・研究科長賞授賞式が行われ、本専攻からWu Pengzhouさんが第9回研究科長賞を受賞されました。

(総務課・大学院係)



# 共同利用

## 2022年度共同利用公募追加課題

### 【共同利用登録】1件

分野	研究課題名	研究代表者(所属)
d5	主変数選択法による抗酸化物質の構造的特徴づけ	吉田 悠夏(横浜国立大学・学部生)

### 【一般研究1】2件

分野	研究課題名	研究代表者(所属)
h2	項目反応理論による言語芸術の潜在評価軸の推定	持橋 大地(統計数理研究所・准教授)
h2	ノンパラメトリックベイズ法による無限木構造トピックモデルの研究	持橋 大地(統計数理研究所・准教授)

(研究推進課・共同利用係)

# 外部資金・研究員等の受入れ

## 受託研究・受託事業等の受入れ

委託者の名称	研究題目	研究期間	研究経費(円)	受入担当研究教育職員
国立大学法人京都大学 学長 湊 長博 代理人 医学・病院構内共通 事務部長 河野矢 英成	こころの健康の保持増進のための超個別化AIプロジェクト:完全要因ランダム化試験からlivingRCTプラットフォームに至る開発研究	R4.4.1~ R5.3.31	650,000	データ科学研究系 野間 久史 准教授
国立大学法人京都大学 学長 湊 長博 代理人 医学・病院構内共通 事務部長 河野矢 英成	京都大学大学院における臨床統計家育成推進のための大学院・卒後一貫したプログラム構築に関する研究開発	R4.4.1~ R5.3.31	6,784,700	椿 広計 名誉教授

(研究推進課・研究推進係)

## 外来研究員の受入れ

氏名	職名	研究題目	研究期間	受入担当研究教育職員	称号付与
Ying Chen	National University of Singapore, Department of Mathematics, Associate Professor	学術論文における科学的権威指標のオープンデータ化に関する研究	2022.9.26~ 2022.9.28	川崎 能典 教授	
Fangyuan Zhang	EURECOM, Postdoctoral researcher	機械学習による年金システムのリスク解析	2022.9.26~ 2022.9.30	栗木 哲 教授	
高橋 武則	慶應義塾大学大学院健康マネジメント研究科・客員教授	初中等教育におけるデータにもとづく問題解決の実践的方法論とその指導	2022.9.16~ 2023.3.31	川崎 能典 教授	
金川 元信	EURECOM, Data Science Department, Assistant Professor	ガウス過程と再生核ヒルベルト空間の関係について	2022.9.26~ 2022.9.30	福水 健次 センター長	
Ziqing Dong	Université de Neuchâtel, Institut de Statistique, Doctoral student	層別サンプリングデータに基づくボンフェロニ曲線の信頼区間構成	2022.10.1~ 2022.11.29	栗木 哲 教授	
Qi Zhang	統計数理研究所ものづくりデータ科学研究センター・元特任研究員	機械学習による分子設計と合成経路予測	2022.9.1~ 2023.3.31	吉田 亮 センター長	

(研究推進課・共同利用係)

## 寄附金の受入れ

受入決定年月日	寄附者	寄附金額	担当教員	寄附目的
R4.8.5	株式会社ブリヂストン 取締役 代表執行役 Global CEO 石橋 秀一	500,000	松井 知子	「機械学習及びモデル化技術の基礎研究」の研究助成

(研究推進課・研究推進係)

## 人事

### 令和4年9月1日転入者(研究教育職員)

異動内容	氏名	新職名等	旧職名等
採用	Le Thanh Tam	数理・推論研究系学習推論グループ 助教	理化学研究所革新知能研究センター 博士研究員

### 令和4年10月1日転入者(技術職員)

異動内容	氏名	新職名等	旧職名等
採用	佐藤 純子	統計科学技術センター情報資源室 技術職員	—

(総務課・人事係)

## 外国人客員紹介



### ● Gareth William Peters 客員教授

GARETH W. PETERS is currently working as a Janet and Ian Duncan Endowed Chair Professor in actuarial science and a Chair Professor in statistics for risk and insurance with the Department of Applied Probability and Statistics, University of California at Santa Barbara. He also holds positions as an Honorary Professor of statistics with University College London, since 2018, an Affiliated Professor of statistics with the University of New South Wales, Australia, since 2015, and an Affiliate Member of the Systemic Risk Center, London School of Economics, since 2014, and Oxford Man Institute, University of Oxford, since 2013. He has also been an Honorary Professor of statistics with The University of Sydney, Australia, since 2018, and Macquarie University, Australia, since 2018. He has also performed work for a variety of international organizations, including ISM. He has published more than 150 peer-reviewed articles on risk and insurance modeling, two research textbooks on operational risk and insurance as well as being an editor and a contributor for three edited textbooks on spatial statistics and Monte Carlo methods. He is an Elected Member of the Young Academy of Scotland in the Royal Society of Edinburgh and an Elected Fellow of the Institute of Operational Risk.

(総務課・人事係)

Report

**Research Memorandum (2022.8~2022.10)**

No.1218: 中西 正, 混合時系列回帰モデルの推定に精度ついて—Rパッケージ ‘midasr’ によるモンテカルロ実験—  
(メディア開発室)

Report

**統計数理研究所調査研究レポート**

No.121: 芝井 清久, 核軍縮問題に関する国際世論調査 —日本, 広島・長崎, アメリカ2022 調査報告書— (2022.6)  
(メディア開発室)

Report

**研究教育活動報告**

No.53: 統計数理研究所, 総合研究大学院大学 複合科学研究科 統計科学専攻, 2022年 統計数理研究所オープンハウス  
ポスター発表 及び 統計科学専攻学生研究発表会 (2022.8)  
(メディア開発室)

Report

**Annals of the Institute of Statistical Mathematics  
Volume 74, Number 5 (October 2022)**

Giorgos Bakoyannis and Dipankar Bandyopadhyay  
Nonparametric tests for multistate processes with clustered data..... 837

Fengrui Di and Lei Wang  
Multi-round smoothed composite quantile regression for distributed data .....869

Yuichi Akaoka, Kazuki Okamura and Yoshiki Otobe  
Bahadur efficiency of the maximum likelihood estimator and one-step estimator for quasi-arithmetic means of the  
Cauchy distribution .....895

Vlad Stefan Barbu, Slim Beltaief and Serguei Pergamenchtchikov  
Adaptive efficient estimation for generalized semi-Markov big data models .....925

Tiandong Wang and Panpan Zhang  
Directed hybrid random networks mixing preferential attachment with uniform attachment mechanisms .....957

Lu Li, Niwen Zhou and Lixing Zhu  
Outcome regression-based estimation of conditional average treatment effect .....987  
(メディア開発室)



# EBPM(エビデンス・ベースト・ポリシー・メイキング) に向けた歴史的転換点

南 和宏

データ科学研究系

コロナ禍の生活は長くなり、研究所でも対面で人に会う機会は大きく減っている。我々の社会に大変な試練を与えていることは間違いないが、平時であれば困難な社会の変革が大きく進み、リモートワーク、オンライン会議は完全に日常の一部となった。その中でもう一つの大きな変革は、データ活用の社会への普及ではないだろうか。毎日のニュースで新規陽性者数の推移を示す棒グラフを見て一喜一憂してきたが、これほどデータが全面に出てくるニュースの話題はなかったように思う。感染数はもちろん、実効再生産数、病床使用率、ワクチン接種率、さらには数理モデルによる感染予測シミュレーションの結果等々、様々なデータがメディアに取り上げられ、議論されてきた。東京都の「新型コロナウイルス感染症対策サイト」では、都内の最新感染動向を様々なモニタリング指標で示すとともに、感染者単位のマイクロデータをオープンデータとして公開している。私が非常勤講師として中央大学で担当する「データサイエンス」の入門コースでは、そのマイクロデータを利用して、感染者数を日別に集計した棒グラフをR言語で作成する演習を行っている。

感染症の防止の部分に関しては、感染者数、病床使用率といったように目的の指標に数値化されたため、自ずとそれに対する政策の立案、有効性の検証も多くの専門家を巻き込む形でデータに基づいて議論されてきたように思う。統計数理研究所でも「新型コロナウイルス対応プロジェクト」が立ち上げられ、多くの研究成果を発信している。これは歴史の長期的視点で見ると、我が国のEBPM(証拠に基づく政策立案)への大きな潮流の節目かもしれない。EBPMはEvidence-Based Policy Makingの略語であり、国が行う政策をこれまでのような経験則ではなくデータに基づき正当評価をしようという取り組みである。近年、我が国において様々

な省庁、自治体等の行政組織で積極的に推進されてきているが、データの重要性が社会的に広く認識された今日、そうした取組が飛躍的に進展することが期待される。

ただし、現状の日本においてEBPMを推進する体制はまだ十分とは言えない。それはEBPMに必要なデータが存在しないという理由ではない。総務省統計局のホームページを見ると、国勢調査、家計調査等、多種多様な統計調査が数多く実施されていることに驚かされる。問題は、これらの「官製ビッグデータ」を活用できる専門性をもつ人材が現在の行政組織に著しく不足している点にある。官庁データサイエンティスト育成の取り組みも始まっているがかなり時間を要すると予想される。この人材不足を補う有望な解決策として考えられるのは、統計調査のマイクロデータを外部の研究者に二次的に提供し、様々な学術研究から得られる知見を通してデータの価値を実証し、社会的な事象間の因果関係を導出する方法論を整備していく道である。政策立案と学術研究の方向性は必ずしも一致するわけではないが、新型コロナの感染症対策では、多くの研究者が学術研究の成果をもとに具体的な提言を行った。同様のことは、経済、教育、社会学の分野でも可能であろう。

近年、このような公的統計のマイクロデータを外部の研究者が利用するための制度は整備されてきており、データ分析を行うオンサイト施設が日本各地の大学、研究機関に整備されつつある。情報・システム研究機構のデータサイエンス共同利用基盤施設(統計数理研究所と同じ敷地)においてもオンサイト施設が設置されており、外部の研究者の方々にも赤池ゲストハウスの利用と合わせて、集中的な分析活動に活用いただいているようで幸いである。

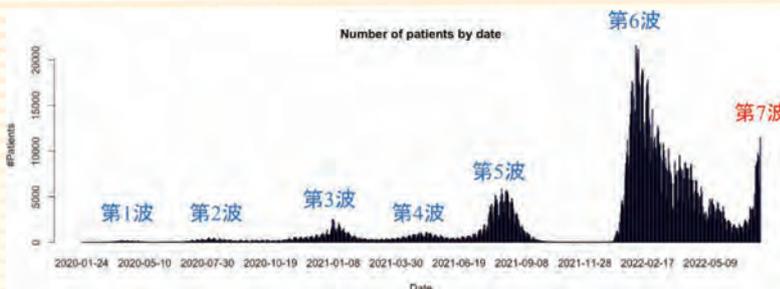


図1：東京都の「新型コロナウイルス感染症対策サイト」のマイクロデータを用い、R言語で作成した新型コロナ陽性者の日別による推移を示すグラフ



図2：データサイエンス共同利用基盤施設のオンサイト室