

▼ CONTENTS

- 02 響き合う人とデーター 統数研プロジェクト紹介
第19回「基礎医学と社会医学をつなぐ離散幾何学的モデリング」
- 06 シンポジウム報告
「Application of R Software for Statistical and Econometric Analysis」開催報告
「Training on Statistical Analysis in "R" for Forest Resource Management Level 4」共同開催報告
- 07 研究教育活動
2019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（2019年9月～10月）
2019年度統計数理研究所 夏期大学院の開講
「統計サマーセミナー 2019」実施報告
- 08 統数研トピックス
医療健康データ科学研究センター公開講座「最近の臨床試験方法論：ベイズ流アプローチ」共催
岡檀特任准教授の活動が、共同通信社の取材により配信
日野英逸准教授が日本神経回路学会「論文賞」を受賞
今泉允聡助教が「第1回 ISI 東京大会記念奨励賞」を受賞
宮城県古川黎明中学校・高等学校の訪問
兵庫高校の訪問
- 10 総合研究大学院大学複合科学研究科統計科学専攻関係
大学院説明会
- 11 共同利用
2019年度共同利用公募追加採択課題
- 11 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
- 13 人事
- 14 刊行物
統計数理／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

統計数理研究所二ユース



離散数学で解き明かす 「細胞分化の木構造」



▲早水桃子助教

距離データに潜む グラフ構造を描出する 「離散幾何学的モデリング」

早水助教が提案する「離散幾何学的モデリング」という枠組みは、有限個の点同士の非類似度が「距離」として数値化されていて、その非類似度のデータを「グラフ」（点と線からなる図）で記述したいという状況を扱うためのものだ。「複雑なデータや現象の全体像をシンプルなグラフで描出するという課題は色々な文脈で現れますが、特に生物学の研究では重要なものです」と早水。

例えば、形態学的な特徴やDNAの塩基配列などに基づく「距離」から進化の系統樹を再構築するというデータ解析は、離散幾何学的モデリングの古典的な例だ（図1）。距離の情報から系統樹というグラフ構造を知るための理論や

方法は深く研究されてきた歴史があるが、生物学の世界には既存の数学的な概念や統計学的手法だけでは太刀打ちできない課題が多数ある。例えば、生物の進化と同様に、細胞の分化もまた木構造で記述される現象だが、進化の道筋を推定する課題とはデータの種類や医学・生物学的なゴールが異なるため、細胞分化の木構造を解明する課題ならではの新しい方法論が必要だ。一方、農作物の品種改良のように、異なる種の掛け合わせで新しい種が生まれる「交雑」と呼ばれる現象は、系統樹のように単純な木構造では記述できないため、「系統ネットワーク」と呼ばれる系統樹よりも表現力の高い一般的なグラフ構造が必要になる。

早水のプロジェクトは、現代の生命科学のニーズに応じて、多様なデータや現象をモデル化するために、さまざま

なグラフと距離の関係を明らかにしよう、という挑戦だ。「複雑なデータに潜むシンプルな構造を描出することは、データの可視化だけでなく、本質的に重要な生物学的発見にもつながります」と早水は語る。

現代の細胞生物学における データ解析上の大きな課題

「動物の進化と、細胞の分化は、木で記述される現象という意味では似ていますが、数学的には大きな違いがあります」と早水は指摘する。図1の例では、①から⑥で示した現存種同士の距離がデータとして与えられ、それらの6点の他に●印の祖先を含むような系統樹が構築されているが、言い換えれば、6つの現存種の多様性を説明するために、●印の絶滅種が存在したと考えて木を描いたということだ。しかし、このよ

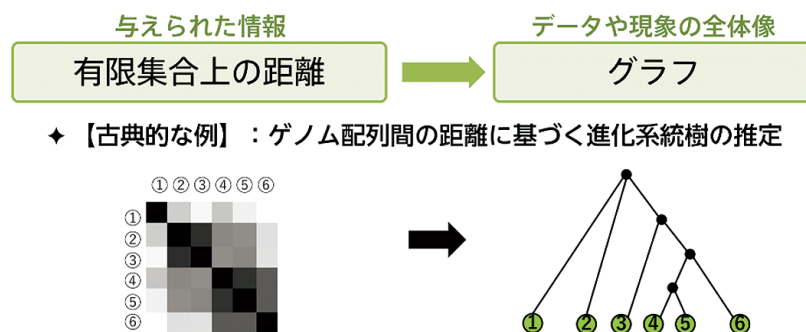


図1：「離散幾何学的モデリング」という枠組みの概要と、その典型的なデータ解析の例。①から⑥の点は現存する種、●の点は想像上の祖先を表す。

第19回 「基礎医学と社会医学をつなぐ 離散幾何学的モデリング」

未来のイノベーションの芽を育む個人型研究「JSTさきがけ」。統計数理研究所の早水桃子助教の「基礎医学と社会医学をつなぐ離散幾何学的モデリング」は、国が推進するこのプログラムに採択された研究提案の一つだ。コンピューターサイエンスの基礎である離散数学という分野の研究からオープンソースのデータ解析ソフトウェアの開発までを手がけ、幹細胞工学など社会的に重要な領域にもつながる研究の一端を紹介する。

うな想像上の点を含めずに、「6種の関係性だけを描きたい」という状況もある。「細胞同士の関係性の全体像を木で説明したい」という問題意識は、まさにその一例だ。

細胞の分化とは、細胞の形態や機能が変わり、別の種類の細胞に変化する不可逆的な運命決定プロセスを意味する。いったん変化すると決して元に戻らないので、生物学者たちは、細胞分化の全体像は合流のない木構造で記述できると考えている(図2)。このような細胞分化の木モデルは、細胞生物学者たちが試行錯誤を繰り返して長い歴史の中で築き上げてきたもので、人類がこれまでの実験から得た知見の集大成といえる。

ところが、近年の実験計測技術の進歩はめざましく、今日の細胞生物学は大きな転換期を迎えている。かつての技術水準では、細胞の集団の遺伝子発現の状態を大雑把に調べることでしかできなかったが、「1細胞技術」と呼ばれる現代のテクノロジーを使えば、個々の細胞における各遺伝子の発現量を、細やかに、しかも網羅的に計測できる。このような新技術がもたらしたデータをうまく活用できれば、これまでの細胞生物学の常識を覆すような新発見ができるかもしれない。実際、1細胞の遺伝子発現データを活用して細胞分化の木構造の全貌を解き明かすことは、今や世界中の細胞生物学者たちが抱く一つ

の大きな夢になった。そこで脚光を浴びているのが、個々の細胞同士の距離を説明する木グラフを構築するという、ま

さに「離散幾何学的モデリング」の考え方なのである(図3)。

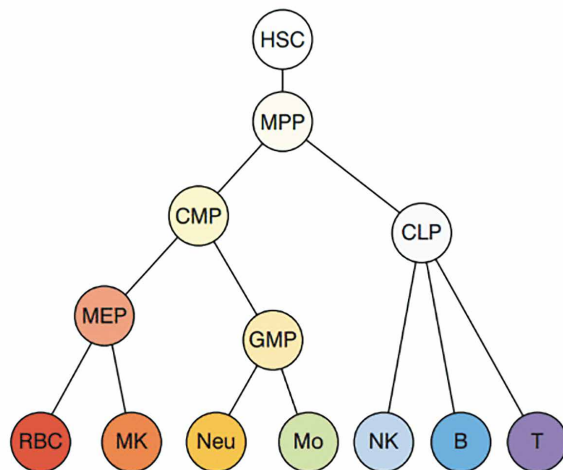


図2：血液に関する細胞の分化の全体像を描いた木グラフ。造血幹細胞(HSC)は、どの種類の血液細胞にもなれるポテンシャルを持つが、最終的には赤血球(RBC)やTリンパ球(T)などの何らかの細胞に分化する。

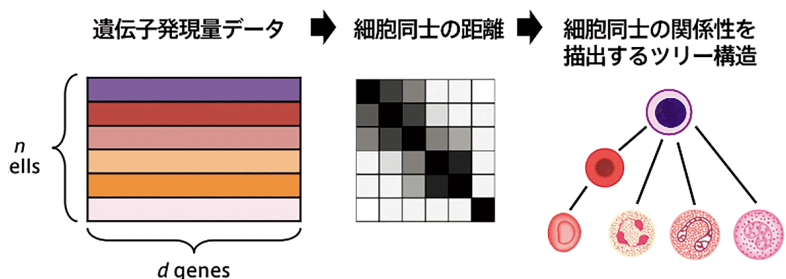


図3：細胞生物学者のニーズは、「個々の細胞から得られた詳細な遺伝子発現データを活用して、細胞分化の真の木構造を解明したい」というもの。

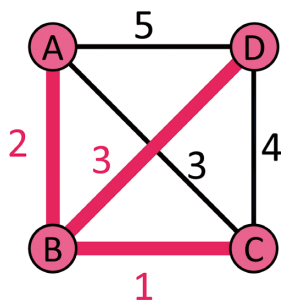


図4：4つの点からなる距離データを表す完全グラフと、その中の最小全域木。最小全域木を求めるアルゴリズムは、距離から細胞の木を構築する“経験則”として普及し始めている。

細胞分化の全体像を描き出す科学的に信頼できる方法論の開発

細胞同士の距離データから木グラフを抽出するための“経験則”として定着し始めたのは、離散数学の基本である「最小全域木問題」を解くアルゴリズムだ。これは、途中でサイクル（輪っか）ができないように距離が近いペアから順に結んでいき、すべての点を通る木が得られたら終了するという単純明快な方法だ。例えば、A～Dという4つの細胞間の距離が得られているとしよう（図4）。まずは一番短い辺BCを選び、その次に二番目に短い辺ABを選ぶ。三番目に短い辺は二つあるが、辺ACを結んでしまうと、ABCというサイクルが形成されてしまうので、辺BDのほうを選ぶ。以上で、A～Dのすべての点を通る木が得られたことになる。色を付けて強調されているのが、最小全域木だ。

「似た細胞同士を結んで木を作るという方法は直観的に自然ですが、最小全域木を求めるアルゴリズムで正しい細胞分化の木構造が得られるという理論的な保証はありません。科学的に信

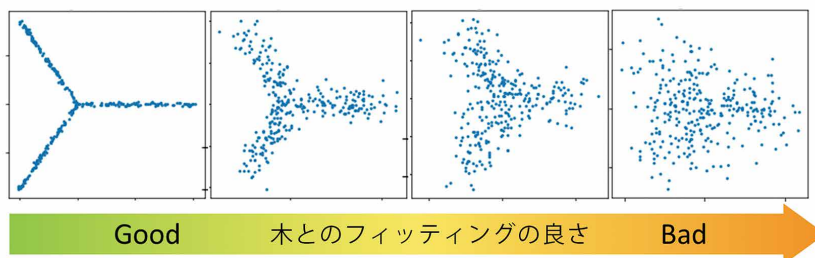


図5：最小全域木は元の距離情報をよく表わしていることもあるが、まったくフィットしていないこともある。

細胞分化の全体像の描出

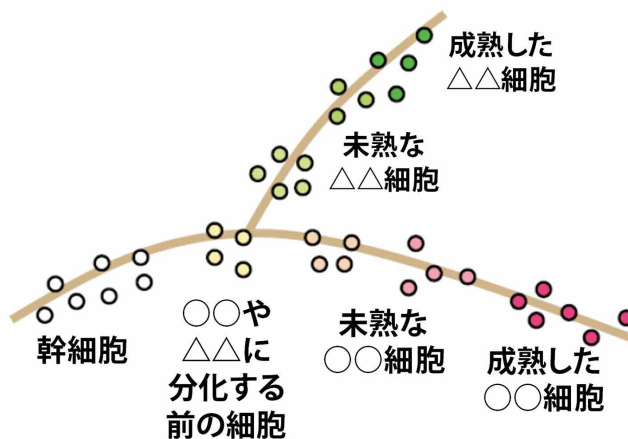


図6：プロジェクトのゴールのイメージ。細胞分化の全体像を推定して描出するような遺伝子発現データ解析ソフトウェアを開発する。

頼できる真に有用なデータ解析手法を創るためには、データと木モデルのフィッティングの良さを定量的に評価する尺度を確立させることが重要だと考えています」と早水は説明する（図5）。

このプロジェクトの最終的な目標は、与えられた距離情報に潜む本質的な木構造を抽出するための確かな方法論を構築し、それを実装したソフトウェアを開発し、世界中の生物学者が利用できる形で公開することだ（図6）。早水は「この研究を着想したのは、iPS細胞の分

化メカニズムを研究している医学系の研究者の方からデータ解析上の相談をいただいたのがきっかけです。数学や情報の研究者だけでなく、医学を含む幅広い分野の研究者たちが使いこなせるものを作り、生命科学の研究に貢献したいと思っています」と語る。

生命科学の根底にある課題を数理科学の力で解決したい

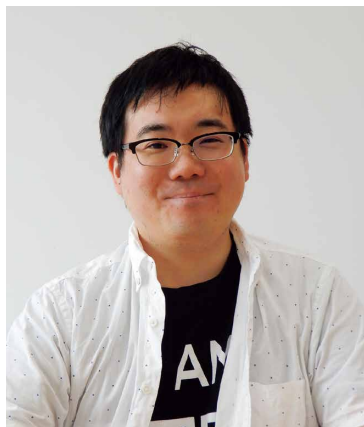
早水はもともと医学部出身で、数理の研究者としては異色の経歴を持つ。



▲須藤功平氏（株式会社クリアコード代表取締役）

統計数理研究所で研究を始める前は、脳機能や代謝の程度を画像検査で調べる核医学分野が専門の放射線科医だった。「あまり聞き慣れない診療科かもしれませんが、放射性同位体を使って初期の認知症やがんなどの進行度合いを評価するという重要な分野で、統計学や物理学の知識は必須でした。目に見えないものをどうやって定量的に可視化するかという問題意識は、今の研究とも通じる部分があります」と早水は言う。

生命科学の知識と臨床経験に加えて、統計科学・数理学・情報科学にわたる幅広い視野を持つ早水は、医学の研究者にとっても、数学の研究者にとっても、非常に心強い存在だ。とはいえ、異分野から転向した当初は、様々な分野との壁を感じる事がしばしばあったという。数理の世界で医学部出身者はあまりにも珍しいうえに、医学の世界でも数学者とのコラボレーションに対して懐疑的な向きがあった。また、離散数学という分野自体も、統計の一部の研究者たちにとっては、あまり馴染みがないものに映ったようだ。「最近はやや多くの方々に興味を



▲石井大海外来研究員（DeepFlow株式会社）

持っていただけるようになりましたが、はじめのうちは、とにかく早く結果を出して、自分の研究の魅力や価値を伝えたいという気持ちがありました。大学院在学中に少し背伸びをして『さきがけ』に応募したのも、そういう背景があります」と早水は振り返る。

さきがけの研究タイトルに掲げたのは『基礎医学と社会医学をつなぐ』という言葉。早水は「幹細胞生物学が良い例ですが、医学の基礎研究のなかには、臨床医学や社会医学に対する影響力が大きいテーマがたくさんあります。そのような研究の突破口を、数理の力で開きたいのです」と思いを語る。

異業種連携でオープンソースのデータ解析ソフトウェアを開発

このプロジェクト紹介の取材には、早水が受け入れている統数研の石井大海外来研究員と、ソフトウェア開発を手掛ける民間企業、クリアコードの須藤功平代表が参加した。

石井は民間企業の研究開発者だが、もともとは数理論理学や計算代数が専門だった。「プログラミングの過程で離散数学やグラフ理論を使うことはあ

りましたが、アルゴリズムを適用することの妥当性を定量的に評価するような研究は初めてなので新鮮です」と感想を述べる。専門が異なるからこそ、有意義なディスカッションができる部分があると早水も言う。

一方、クリアコードはフリーソフトウェアを開発していた須藤氏が立ち上げた会社だ。自分たちで書いたコードを独占するのではなく、誰でも自由に使える形で公開することで対価を得る開発スタイルを貫いている。

アメリカでは研究者とエンジニアが連携してソフトウェアを開発し、論文を発表するというのが珍しくないが、日本ではまだあまり一般的ではない。論文の査読者や読者が内容を検証できるように、開発したデータ解析ソフトウェアを無償で公開しようと考えていた早水にとって、オープンソース・ソフトウェアを開発する企業の理念はうってつけだった。

須藤氏は「ソフトウェアを開発する技術に加えて、公開して普及させるのが、当社の得意とするところです。この強みをプロジェクトで生かしたいと思っています」と話す。

早水は「大学や研究所での研究プロジェクトに民間の意欲的なエンジニアを巻き込むことは、研究者側が積極的に動かないと難しい。私のプロジェクトが良い成功事例として広まり、日本でもこのような異業種連携がもっと一般的になればいいと思っています」と話す。医学と統計数理をつなぐ研究内容も先駆的なら、その進め方も新しい。さまざまな意味で「さきがけ」と言えるプロジェクトだ。

（広報室）

「Application of R Software for Statistical and Econometric Analysis」開催報告

令和元年7月24日から26日にかけて、インドネシア・ボゴールのBogor Agricultural University (IPB) のIPB International Convention Centerにおいて、Indonesian Society of Agricultural Economics (PERHEPI) の創立50周年記念国際学会「International conference of 50 years PERHEPI」が開催され、その初日(24日)にプレ・カンファレンス・ワークショップ「Application of R Software for Statistical and Econometric Analysis」を開催しました。これまではインドネシアのMOU締結研究機関ランブン大学においてワークショップを展開してきましたが、今回はランブン大学のArfin教授が同学会の実行委員であることから、彼の要請に対応したものです。これまでのワークショップ同様に、統計解析ソフトRを用いた実践的解析について講義・実習を行いました。過去のワークショップへ参加した受講者もいましたが、大半は他の研究機関からの参加であったため、初歩レベルから開始し、データハンドリング、グラフィックス、そして線形回帰分析、成長データを用いた非線形(成長関数)回帰分析を経て、最後は受講者側からの要

望に応え時系列分析(ARモデルとARIMAモデル)の講義を行いました。今後は、受講者が直面している統計的な問題について、Q&A形式によるアドバイスの講義の展開も検討していく必要がありそうでした。今回、初めてRを利用した受講者にも、普段からRを積極的に利活用し慣れ親しむことの重要性を再確認してワークショップを終了しました。ワークショップの運営は、ランブン大学のArifin教授らが中心に行いました。今回は学会の一環ということもあり、ドロップイン参加者も多く、合計70名の参加でした。(吉本 敦)



「Training on Statistical Analysis in “R” for Forest Resource Management Level 4」共同開催報告

カンボジアMOU締結研究機関Institute of Forest and Wildlife Research and Developmentにおいて、令和元年9月3~4日の二日間に渡り、ワークショップ「Training on Statistical Analysis in “R” for Forest Resource Management Level 4」を開催しました。本ワークショップは、統計解析ソフトRを用いて、データハンドリング・可視化から森林データ分析に主眼を置いた様々な統計分析法の実践に取り組んできました。今回が初参加の受講者が複数いたため、

まず、これまで扱ってきたLevel 3までの内容の自己学習用の事前配布資料を現地語で参加者同士教え合うことで習熟度の差異の解消に取り組みました。これにより、初日午後からLevel 4の内容にスムーズに入ることができました。Level 4では、連続値をとる測定データに対する統計モデルとして、線形回帰モデルおよび成長関数に基づく非線形回帰モデルを扱い、推定法、結果の解釈、データへ適合度の評価、モデル選択など、森林データ分析を想定した実践的な演習に加えて、その背景にある理論的な側面についても概説しました。2日間を通じて積極的な質疑応答が繰り返されました。今回のワークショップは、実践的データ解析に主眼を置き、その背景にある理論的な内容は概説のみで展開しましたが、理論的な側面のより詳しい内容についても講義して欲しいと、引き続きの開催要望がありました。ワークショップの運営は、Institute of Forest and Wildlife Research and DevelopmentのSokh Heng氏らが中心に行いました。尚、今回の参加者は30名でした。(吉本 敦)



2019年7月－9月の公開講座実施状況

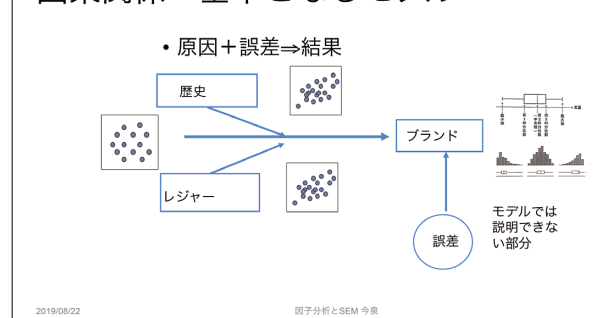
7月17日（水）に、電気通信大学の川野秀一講師による「スパース推定」が開講されました。正則化法からはじめ、lasso法を軸としたスパース推定法、スパース推定の推定値を得るための計算アルゴリズム、構築したスパースモデルの評価方法等の講義を行いました。この講座は昨年度も実施しておりますが、今年も多数の申込みがあり、抽選により

受講者を決定しました。

8月20日（火）から23日（金）まで、4日間の日程で「多変量解析法」がエッサム神田ホール2号館で開講されました。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、多変量解析の古典的・標準的また発展的な手法の解説を講義しました。連日、休み時間や講義終了後に受講生が講師に質問する場面が往々にして見られ、この講座への関心の高さが窺えました。

9月17日（火）から19日（木）まで、当研究所の伊庭幸人講師、川崎能典講師、二宮嘉行講師、日野秀逸講師、坂田綾香講師、国立精神・神経医療研究センター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. データサイエンスの基礎」が開講されました。データを扱うための基礎から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、基礎を固めるための講義を行いました。（情報資源室）

因果関係 基本となるモデル



「多変量解析法」資料より

統計数理セミナー実施報告（2019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。2019年9月～10月のセミナーは下記の通り行われました。

日 程	氏 名	タイトル
2019年 9月25日	宮里 義彦	入力非線形特性を有するマルチエージェント系の制御 ～入力非線形特性のモデリングと制御～
9月25日	齋藤 翔	曲面上のベクトル束のモジュライ空間のコンパクト化
10月 2日	川崎 能典	動的トピックモデルを使ったボラティリティ予測
10月 2日	後藤 振一郎	離散幾何学によるマスター方程式からの離散拡散方程式の導出
10月 9日	横山 雅之	統計手法を用いた核融合プラズマの熱輸送モデリング
10月 9日	島谷 健一郎	統計学と統計を使う生態学の科学哲学
10月16日	三分一 史和	ニューロン間の因果的結合性とネットワーク構造の推定および実験的検証
10月16日	齋藤 正也	血清データおよび機構モデルにもとづくインフルエンザ定点報告率の推定
10月23日	前田 忠彦	現代日本における生活満足感の規定因の安定性について
10月23日	Chen Ye	Stein point Markov chain Monte Carlo
10月30日	逸見 昌之	A worst-case bound for publication bias in meta-analysis of diagnostic studies
10月30日	Guo Yicun	Modeling and forecasting aftershocks can be improved by incorporating rupture geometry in the ETAS model

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）

2019年度統計数理研究所 夏期大学院の開講

統計思考院公募型人材育成事業として、2019年8月1日から10日間連続で感染症数理モデル短期コース（正式「入門：感染症数理モデルによる流行データ分析と問題解決」）を開催しました。今年度は第6回目の開催で、受講者総数

は62人、うち外国人が29人という国際色豊かなコースとなりました。統計学、数理科学、情報学、医学、獣医学はもちろん、多様な背景の方に系統的な入門講義と実践的なグループワーク研究を経験してもらいました。6年連続で本夏期

大学院で短期コースを無料開講させていただき、数理モデルの定式化から統計学的推定や予測の実装までを体系的に学ぶ機会を提供してきましたが、研究と教育の状況は開始前と隔世の感があります。来年の東京五輪等も勘案して本コースは今年度で最後とさせていただきますが、確実に次世代の芽が出つつありますので、今後もこの歩みは止めずに進化させた短期コースを計画・提供していきたいと思えます。

(北海道大学大学院医学研究院 西浦 博)



「統計サマーセミナー2019」実施報告

統計数理研究所公募型人材育成事業採択課題の「統計サマーセミナー 2019」は2019年8月5日から8月8日まで、国民宿舎ひびきで開催されました。

本セミナーの目的は、若手に研究発表の機会を与え、将来統計学の発展を担う学生・研究者、あるいは実社会で企業人として統計学を使いこなせるような人材を育成することです。多くの学生・研究者が、何らかの個別科学の学部学科等に所属し統計学周りの研究をしている現在の環境において、研究早期の段階から色々な分野での統計の使われ方に触れ、自身の研究テーマに絞られがちな視野を大きく広げることが本

セミナーの大きな狙いです。また本セミナーは、合宿形式で行ったため、講演以外の時間でも心おきなく討論が可能という利点があります。実際にセッション終了後も、毎晩遅くまで積極的な議論が繰り広げられ、各自が疑問点を述べ、様々な主張をし、各分野間で活発な交流がされていました。

また、広島大学 柳原氏、電気通信大学 川野氏による招待講演を設け、特に学会講演では聴く機会が少ないこれまでの研究者生活の過程なども聴けました。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。

(中京大学 永井 勇)

統数研トピックス

医療健康データ科学研究センター公開講座「最近の臨床試験方法論:ベイズ流アプローチ」共催

2019年9月3日、統計数理研究所医療健康データ科学研究センターは公開講座として、「最近の臨床試験方法論:ベイズ流アプローチ」を共催いたしました。開催にあたっては、



主催がスタートコム株式会社、東京大学大学院医学系研究科生物統計情報学講座と統計数理研究所医療健康データ科学研究センターが共催いたしました。講演演題としては、中央大学の大橋靖雄教授がイントロダクションを話され、続いて京都府立医科大学の手良向聡教授による「医療現場で必要とされているベイズ流意思決定」、東京大学の平川晃弘特任准教授による「バスケット型臨床試験におけるベイズ流アプローチ」、ケンブリッジ大学のデビット・シュピーゲルハルター教授による「Bayesian Approach to Clinical Trials: Historical Perspective and Future」が話されました。総参加者数は96名となり、非常に活況でした。

(伊藤陽一)

岡檀特任准教授の活動が、共同通信社の取材により配信

医療健康データ科学研究センターの岡檀特任准教授の活動が、共同通信社の取材によって「14歳の君へーわたし

たちの授業」という記事となり、全国に配信されました。

全国の中学生へ向けて、さまざまな分野の人が“授業”

という形式で語りかける連載記事で、この回は数学の授業でした。自身の研究成果になぞらえながら、「統計的思考」の定義、その重要性などを平易な言葉で説明しました。

(医療健康データ科学研究センター)



日野英逸准教授が日本神経回路学会「論文賞」を受賞



日野英逸准教授が日本神経回路学会において2019年度「論文賞」を受賞しました。

「Neural Networks, Volume 108 (2018)」に掲載された論文『Estimation of neural connections from partially observed neural spikes』での受賞になります。

授賞式は9月6日(金)に日本神経回路学会第29回全国大会(会場 東京工業大学蔵前会館くらまえホール)にて行われました。(広報室)

今泉允聡助教が「第1回 ISI東京大会記念奨励賞」を受賞

今泉允聡助教が「第1回 ISI東京大会記念奨励賞」を受賞し、統計関連学会連合大会企画セッション「日本統計学会会長講演、各賞授賞式および受賞者記念講演」にて表彰されました。ISIはInternational Statistical Institute(国際統計協会)の略称で、将来活躍が期待できる若手の育成に寄与するため日本統計学会が「ISI東京大会記念奨励賞」を制定しました。(広報室)



宮城県古川黎明中学校・高等学校の訪問

2019年9月4日(水)、宮城県古川黎明中学校・高等学校の中学生を含む生徒20名と教諭2名が本研究所を訪れました。同校はSSH(スーパーサイエンスハイスクール)の指定校であり、サイエンス研修の一環として、研究施設を訪問し研究内容に触れることで、科学技術系人材育成へつなげることを目的としています。

当日は北村上席URAの研究所紹介から始まり、南和宏准教授による「ビッグデータとプライバシー保護」の研究紹介が行われました。検索・キーワードやドライブレコーダーの



データがどのように活用されているか、またプライバシー保護のデータ加工について身近なSNSを例に挙げて解説が行われました。そして上野玄太教授の施設概要説明ののち、計算機室、計算機展示室を見学しました。生徒達からは「プ

ライバシー保護は自らが率先して行うべきだと改めて実感した」「スーパーコンピュータに圧倒された」などの感想が寄せられ、関心の高さが伺えました。この訪問が将来の進路選択の幅を広げる一助となることを願っています。（広報室）

兵庫高校の訪問

2019年7月31日（水）、兵庫県立兵庫高等学校の生徒22名と教諭2名が研究所を訪れました。これは同校がキャリア教育の一環で「東京みらいフロンティアツアー」と称して4研究所を訪問し、見識を広め将来の進路実現を図る目的で実施しているものです。

当日は北村上席URAの司会のもと、田村名誉教授の挨拶



で始まり、研究紹介と施設見学が行われました。

各研究所の研究紹介は、統計数理研究所 島谷准教授による「データから推定できること、できないこと：観察されなかった種と種数」、国立極地研究所 岡田准教授による「南極観測から探る地球環境」、国文学研究資料館 入口教授による「本のかたちとジャンル」、国立国語研究所 小木曾教授による「コーパスでさぐる日本語の歴史」と題して行われました。それぞれの講義で質問の手が挙がり、興味の深さが伺えました。

その後、各研究所の施設を見学し、統数研では川崎教授の解説による「4K3D可視化システム」の体験、本多主任URAによる計算機展示室とスパコン見学を行いました。写真を撮る姿も多くみられ、生徒達から「視野が広がった」「色々な分野を学びたい」等の感想があり、将来を考える良い機会となりました。（広報室）

総合研究大学院大学複合科学研究科統計科学専攻関係

入学者選抜試験結果

【5年一貫制】

試験年月日	合格者数	
2019年8月20日（火）、8月22日（木）	2019年10月入学（第2回）	0名
	2020年4月入学（第1回）	—

【3年次編入学】

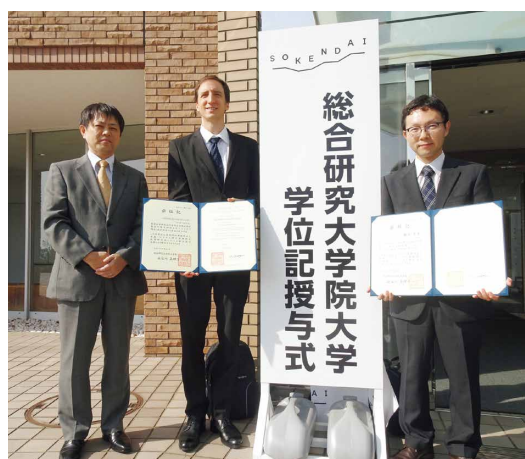
試験年月日	合格者数	
2019年8月21日（水）～8月22日（木）	2019年10月入学（第2回）	6名
	2020年4月入学（第1回）	2名

専攻修了式

2019年9月25日（水）に、セミナー室2（D304）において、専攻修了式が行われ、2名が本専攻を修了しました。

秋季学位記授与式

2019年9月27日（金）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2名が学位記を授与されました。



学位記授与式（左から藤澤専攻長、Andradeさん、右京さん）

学位取得者

2019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
右京 芳文	欠測を伴う経時的測定データに対するMMRM (Mixed-effects Model Repeated Measures) における小標本下での推測手法
Andrade Silva Daniel Georg	Bayesian Model Selection under Noise - From Statistical to Practical Significance (ノイズを考慮したベイズモデル選択 - 実用的なモデル選択へ)

秋季入学式

2019年10月8日(火)に、総合研究大学院大学葉山キャンパスにおいて、秋季入学式が行われ、本専攻への入学者6名を含む39名の新入生が迎えられました。

(研究推進課・大学院係)



入学式(左から沈さん、藤澤専攻長、岩山さん、村瀬さん、榑所長)

大学院説明会

2019年6月5日(水)に、学術総合センターにおいて、2019年度第1回大学院説明会を開催しました。

今回は75周年記念式典の一部としての開催だったため、教員との面談は行わず全体説明1時間だけの開催でしたが、44名の参加者があり、質疑応答も活発に行われました。いつもどおり参加者のほとんど(今回は80%)が社会人で、仕事と学業を両立させる方法についての質問もありました。今後もますます社会人学生が増えていきそうです。

(研究推進課・大学院係)



共同利用

2019年度共同利用公募追加採択課題

【共同利用登録】2件

分野	研究課題名	研究代表者(所属)
j9	臨床試験における外部情報を活用した試験デザインと解析手法の構築	大東 智洋(筑波大学大学院・大学院生)
e3	ネット生存を比較するための正確な検定法の開発	五所 正彦(筑波大学・教授)

【一般研究1】1件

分野	研究課題名	研究代表者(所属)
c4	地磁気永年変化予測のためのデータ同化技術の開発	中野 慎也(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
国立研究開発法人国立長寿医療研究センター 理事長 新井 秀典	高齢者2型糖尿病における認知症予防のための多因子介入研究 ―パイロット研究―	H31.4.1～ R2.3.31	500,000	データ科学研究系 野間 久史 准教授

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
国立研究開発法人新エネルギー・産業技術総合開発機構 理事長 石塚 博昭	深層確率コンピューティングに適した計算アルゴリズムの開発	H31.4.1～R2.3.31	12,000,000	数理・推論研究系 福水 健次 教授
学校法人順天堂 順天堂大学 学長 新井 一	研究データ品質管理に関する教育プログラムの作成(データ管理の立場からデータ保存について)	H31.4.1～R2.3.31	130,000	データ科学研究系 伊藤 陽一 教授
国立大学法人北海道大学 大学院医学研究院長 吉岡 充弘	新興再興感染症のリアルタイム研究、予防接種研究	H31.4.1～R2.3.31	0	モデリング研究系 齋藤 正也 特任准教授
国立研究開発法人国立がん研究センター 理事長 中釜 斉	「統合された科学的根拠に基づく日本人のためのがんリスク評価モデルの開発とその革新的改善に資する疫学研究の推進」	H31.4.1～R2.3.31	3,900,000	データ科学研究系 野間 久史 准教授
国立研究開発法人国立がん研究センター 理事長 中釜 斉	血液バイオマーカーを用いた効率的な膵がん検診の実用化	H31.4.1～R2.3.31	520,000	医療健康データ科学研究センター 長島 健悟 特任准教授
国立研究開発法人国立がん研究センター 理事長 中釜 斉	タンパク質・ペプチド修飾解析による早期がん・リスク疾患診断のための血液バイオマーカーの開発	H31.4.1～R2.3.31	390,000	医療健康データ科学研究センター 長島 健悟 特任准教授

(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
Can Burcu	Hacettepe University, Department of Computer Engineering, Assistant Professor	統計的自然言語処理における教師なし学習について	2019.8.13～2019.8.26	持橋 大地 准教授
金川 元信	University of Tuebingen, Research Scientist	ガウス過程と再生核ヒルベルト空間の関係について	2019.7.4～2019.7.21	福水 健次 センター長
Minh Le Nguyen	Japan Advanced Institute of Science and Technology, Associate Professor	法律文書からの重要情報抽出に関する研究	2019.7.4～2020.3.31	松井 知子 教授
柳 松	University of Bristol, Lecturer	物質科学への統計的機械学習の応用	2019.7.21～2019.8.4	福水 健次 センター長
磯村 哲	株式会社三菱ケミカルホールディングス	分子設計のためのケモインフォマティクス技術の開発	2019.7.4～2020.3.31	吉田 亮 センター長
Patrick Thomas Georges Rubin-Delanchy	University of Bristol, Lecturer	位相的データ解析の機械学習的研究	2019.7.13～2019.7.27	福水 健次 センター長
Alexander Donald Modell	University of Bristol, PhD Student	位相的データ解析の機械学習的研究	2019.7.20～2019.7.27	福水 健次 センター長
熊 子瑤	Peking University, Department of Geophysics, PhD Student	地震ハザード解析における統計的モデリング	2019.9.1～2019.9.24	庄 建倉 准教授
陳 石	China Earthquake Administration, Institute of Geophysics, Research Professor	重力データの3次元高解像度Bayesインバージョン	2019.8.16～2019.9.16	庄 建倉 准教授
張 貝	China Earthquake Administration, Institute of Geophysics, Assistant Professor	重力データの3次元高解像度Bayesインバージョン	2019.9.1～2019.9.24	庄 建倉 准教授
李 紅蕾	China Earthquake Administration, Institute of Geophysics, Assistant Professor	重力データの3次元高解像度Bayesインバージョン	2019.9.1～2019.9.24	庄 建倉 准教授
Ilaria Spassiani	Istituto Nazionale di Geofisica e Vulcanologia (INGV), PostDoc Researcher	短期地震予報モデルの改善	2019.8.16～2019.9.22	庄 建倉 准教授
Katerina Orfanogiannaki	Institute of Geodynamics, National Observatory of Athens, Research Associate	地震発生のETAS型隠れマルコフモデル	2019.8.16～2019.8.31	庄 建倉 准教授
周 仕勇	Peking University, Department of Geophysics, Professor	地震シミュレーションと大災害債券設計	2019.8.8～2019.8.31	庄 建倉 准教授
小林 和博	青山学院大学理工学部・准教授	航路計画最適化における数値最適化とそのデータ解析	2019.9.1～2020.3.31	田中 未来 助教

氏 名	職 名	研究題目	研究期間	受入担当研究教育職員
池端 久貴	旭化成株式会社 研究・開発本部インフォマティクス推進センターデータサイエンス部・エキスパート	化学分析から得られる各種関数データ間の関係理解のための統計的手法の開発	2019.9.3～2020.3.31	吉田 亮 センター長
Saeed Hayati	University of Isfahan, PhD Student	カーネル法を用いた関数データ解析の理論的研究	2019.9.3～2019.11.1	福水 健次 センター長
李 澤瀟	School of Geosciences, Institute of Disaster Prevention Science and Technology, Master Student	地震の発震機構に基づく構造応力場の反転	2019.9.17～2019.11.17	庄 建倉 准教授
高橋 宏典	株式会社ACCESS	Group Fused Lassoを用いた実務的制約付きクラス料率算定法	2019.9.3～2020.3.31	野村 俊一 助教

(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(円)	担当教員	寄附目的
H31.7.10	統計数理研究所創設75周年記念事業実行委員会委員長 椿 広計	85,240	椿 広計	統計数理の研究助成
H31.7.10	ムサシノ機器株式会社代表取締役 杉山 直樹	1,000,000	瀧澤 由美	電磁波による測位に関する研究助成
H31.9.6	川崎 能典	1,000,000	川崎 能典	「市場制度変更が金融取引に及ぼす影響の高頻度データに基づく分析」への研究助成

(財務課・予算・決算第二係)

人 事

令和元年8月1日機構内異動(研究教育職員)

異動内容	氏 名	新職名等	旧職名等
配置換	前田 忠彦	データサイエンス共同利用基盤施設社会データ構造化センター准教授	データ科学研究系調査科学グループ准教授

令和元年9月30日転出者(研究教育職員)

異動内容	氏 名	新職名等	旧職名等
早期退職	吉野 諒三	同志社大学文化情報学部特別客員教授	データサイエンス共同利用基盤施設社会データ構造化センター教授

令和元年7月1日役職者の異動

異動内容	氏 名	職 名	任 期
兼 務	上野 玄太	図書室長	令和2年3月31日まで
兼務免	川崎 能典	図書室長	

令和元年10月1日役職者の異動

異動内容	氏 名	職 名	任 期
兼 務	船渡川 伊久子	医療健康データ科学研究センター副センター長	令和3年3月31日まで
兼務免	野間 久史	医療健康データ科学研究センター副センター長	

外国人研究員(客員)

氏 名	現 職	所 属	職名	研究課題	期 間	受入教員
ノーディン アザオイ Nourddine Azzaoui	准教授	Clermont Auvergne University	客員准教授	時空間データの特徴と表現	令和元.6.25～ 令和元.7.26	松井 知子 教授
ピーターズ ギャレス ウィリアム Peters Gareth William	教授	Heriot-Watt University	客員教授	ガウシアンプロセスによる時空間推定	令和元.6.17～ 令和元.7.29	松井 知子 教授
ミルボルトル アンドレ Myrvoll Tor Andre	研究員	ノルウェー工業技術研究所	客員教授	統計的機械学習によるマルチモーダルデータ処理	令和元.7.8～ 令和元.8.2	松井 知子 教授
セプタイヤー フランシス ジャン ミッシェル Septier Francois Jean Michel	教授	Université Bretagne Sud	客員教授	人間行動の統計的解析のための基盤技術の開発	令和元.7.16～ 令和元.8.15	松井 知子 教授

(総務課・人事係)

外国人客員紹介



●Gareth William Peters 客員教授

This year I have visited Prof. Tomoko Matsui. We have had an interesting research engagement on a computer vision challenge known in Machine Learning as Disentanglement, which involves looking for statistical methods to extract features from an image in order to capture the independent components of the frames in a video sequence. In addition, we have been completing our latest research papers, we have submitted 2 and one is in preparation for submission. The discussion and research collaboration is very interesting. I have enjoyed to visit ISM very much and thank you for your kind hospitality during my research visit.



●Azzaoui Nourddine 客員准教授

I am very grateful to Prof. Tomoko Matsui for hosting me in the Institute of Statistical Mathematics for one month in June-July 2019. The main purpose of my visit was to progress on our joint projects with Prof Matsui and her team about spatiotemporal modelling in the context of impulsive and non-Gaussian context. We are developing theoretical methods to characterise some heavy tailed non-stationary spatial processes that can be applied for instance to meteorological data. In this setting, we are also investigating, the hot topic of property testing of Markov chains and Machine Learning theory to detect the behaviour changes in this kind of data. I will be very happy to continue our collaboration in this very important research area which can have a decisive impact on our societies. I really enjoyed experience of working environment, culture and everyday life in Japan and looking forward to visiting ISM again.



●Tor Andre Myrvoll 客員教授

I am happy to be back for yet another summer visit at ISM, and I want to thank my host, Professor Tomoko Matsui, for once again inviting me. This year we have continued our work on the use of reinforcement learning techniques in climate modeling. A simple climate model including stochasticity has been proposed, and an optimal policy on how to apply mitigative procedures to curb carbon emission is sought. This year we have also studied the well know DICE model by Nordhaus, and plan to extend the model to include stochasticity and apply our reinforcement learning procedures to solve the policy optimization problem.



●Francois Jean Michel Septier 客員教授

I am very glad to be visiting ISM once again this year for one month in July-August 2019. During this visit, I am more specifically working with Prof. Tomoko Matsui on the development of novel spatio-temporal random fields in order to more accurately capture the (possibly extreme) variation of physical phenomena such as, for example, temperature, pollutant levels, etc. We are also conducting some research activities on the development of novel statistical techniques for Bayesian optimal filtering in high-dimensional problems. I greatly appreciate the research environment at ISM including the Akaike guest house. As always, ISM gives me the opportunity to interact with Japanese researchers and other visiting researchers to make progress on exciting research projects.

刊行物

統計数理 第67巻 第1号

特集「農林業の生態系サービスの経済・統計分析」

「特集 農林業の生態系サービスの経済・統計分析」について

田中 勝也、吉本 敦1

日本における森林生態系サービスの経済評価 [原著論文]

佐藤 真行、栗山 浩一、藤井 秀道、馬奈木 俊介3

農林業の生態系サービスと自然資本のグローバル分析 [研究ノート]

池田 真也、山口 臨太郎、馬奈木 俊介	21
外来種駆除の生物多様性保全効果：保全優先地域と脅威動態の関係 [原著論文]	
楠本 聞太郎、南木 大祐、久保田 康裕	39
インターネット調査におけるモニター情報の利用による非回答バイアスの補正—国内草資源を利用した放牧飼養牛肉に対する消費者評価への適用— [研究ノート]	
楠戸 建、後藤 貴文、高橋 義文、矢部 光保	51
ミツバチの送粉サービスと景観構造との関係解析 —宮崎県綾町における日向夏生産の事例— [研究ノート]	
光田 靖、湯村 昂広、平田 令子、伊藤 哲	63
関数データに基づく統計的モデリング [総合報告]	
松井 秀俊	73
農地と森林の生態系サービスの経済評価手法 [研究詳解]	
柘植 隆宏	97

個人企業向けクレジットスコアリングモデルにおける業歴の有効性 [原著論文]

尾木 研三、内海 裕一、枇々木 規雄	121
--------------------------	-----

(メディア開発室)

Annals of the Institute of Statistical Mathematics

Volume 71, Number 5 (October 2019)

Xiaolin Chen, Yi Liu and Qihua Wang

Joint feature screening for ultra-high-dimensional sparse additive hazards model by the sparsity-restricted pseudo-score estimator	1007
--	------

Youngmi Lee and Sangyeol Lee

CUSUM test for general nonlinear integer-valued GARCH models: comparison study	1033
--	------

Haejune Oh and Sangyeol Lee

Modified residual CUSUM test for location-scale time series models with heteroscedasticity	1059
--	------

Xianyi Wu and Xian Zhou

On Hodges' superefficiency and merits of oracle property in model selection	1093
---	------

Jungjun Choi and In Choi

Maximum likelihood estimation of autoregressive models with a near unit root and Cauchy errors	1121
--	------

Xuejun Wang, Yi Wu and Shuhe Hu

The Berry-Esseen bounds of the weighted estimator in a nonparametric regression model	1143
---	------

Jing Lv and Chaohui Guo

Quantile estimations via modified Cholesky decomposition for longitudinal single-index models	1163
---	------

Zhuoer Sun and Suojin Wang

Semiparametric estimation in regression with missing covariates using single-index models	1201
---	------

Matthias Hansmann, Michael Kohler and Harro Walk

On the strong universal consistency of local averaging regression estimates	1233
---	------

Matthias Hansmann, Michael Kohler and Harro Walk

Correction to: On the strong universal consistency of local averaging regression estimates	1265
--	------

Frederic Paik Schoenberg, Marc Hoffmann and Ryan J. Harrigan

A recursive point process model for infectious diseases	1271
---	------

Avijit Maji, Abhik Ghosh, Ayanendranath Basu and Leandro Pardo

Robust statistical inference based on the C -divergence family	1289
--	------

(メディア開発室)



Law of Large Numbers and Winners Take All

志村 隆彰

数理・推論研究系

大数の法則はかの広辞苑にも載っている立派な定理である。コインを何度も繰り返し投げたときの表の出る割合が1/2位になるようなことがその例で、視聴率、選挙予測などの社会調査をはじめ、統計が用いられる裏で大活躍している。その数学的定式化は、次のようになる。

大数の(弱)法則：

平均 μ 、分散 σ^2 の分布 F に互いに独立に従う確率変数 $X_1, X_2, \dots$ と、任意の $\epsilon > 0$ に対して、

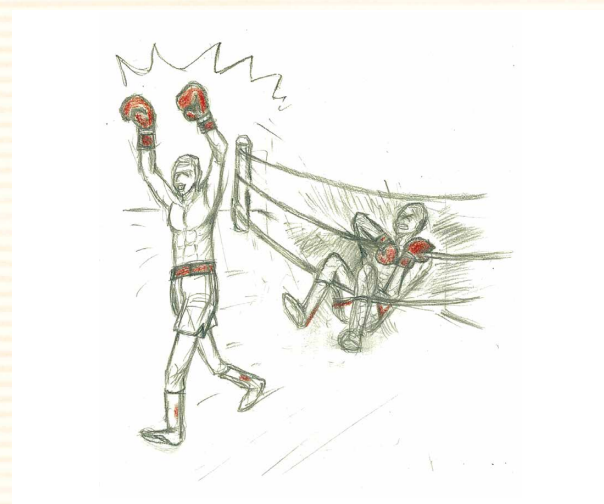
$$\lim_{n \rightarrow \infty} P(|(X_1 + X_2 + \dots + X_n)/n - \mu| < \epsilon) = 1$$

確率変数はランダムな現象を表すもので、コインの例では、 X_k が k 回目にコインを振った結果、表なら1、裏なら0とし、表裏(0と1)が同じように出るとすれば、それぞれの確率は1/2となる。 X_k の分布 F の平均 μ は出る数値とその確率をかけたものを足し合わせて、 $1 \times 1/2 + 0 \times 1/2 = 1/2$ と計算される。コインを一回投げたとき表が出る回数の平均は1/2という意味である(分布から決まる定数)。一方、 $(X_1 + X_2 + \dots + X_n)/n$ は一般に標本平均と呼ばれ、コインの例では n 回コインを投げたときの表が出た割合で、試行(確率的実験)の結果だから、ランダムである(試行のたびに違うことが普通)。つまり、大数の法則は、 n (標本数)が大きくなるにつれて、ランダムなものがそうでない定数(分布の平均)に近づいていくことを意味している。社会調査のサンプル調査では、コインの例と違って分布の平均はわからないから、調査結果から真の値(分布の平均)を推測する形で用いられる。より正確に定式化の意味を説明するには、 P (事象の確率を表す。Probabilityの頭文字)や ϵ (いくらでも小さくてもいい定数)などにも触る必要があるが、コイン投げで永久に表が出続け、1の割合がずっと1のままで、1/2には近づかない確率も0ではないためという程度に留めて、以下では、この法則が成り立つ条件「平均 μ 、分散 σ^2 」の意味を考えたい。

この条件はさっと書いてあるので、確率変数の分布にはいつでも平均があるのだらうと思うかもしれない(ばらつきを表す分散も同じだが、以下、平均のみ扱う)。だが、標本平均が常にある(試行の結果として計算できる)のに対して、分布の平均は常にあるとは限らないのである。0か1かのコイン投げとは違い、いくらでも大きい値を取る確率変数で、非常に大きい値がある程度大きな確率で起こるような分布では、分布の(有限な)平均を定

めることが出来ず、そのような分布に従う標本平均は標本数が大きくなっても、なんらかの定数に落ち着きはないのである(いつまでもランダムさが残る)。たとえば、 $X_1, X_2, \dots, X_n$ と続いてきて、次の X_{n+1} がそれまでのどんぐりの背比べとは違い、それ以前の合計 $X_1 + X_2 + \dots + X_n$ よりもはるかに大きい値を取ったら、 $n+1$ までの標本平均は、 X_{n+1} たったひとつに左右されて、 n までは大きく違ってしまふ。これでは大数の法則も形無しである。つまり、大数の法則は極端なことが余り起こらない場合に成り立つ定理で、「平均 μ 、分散 σ^2 」という仮定の意味はそこにある。

平均を持たない、極端な値を比較的良好な分布は理論的には古くから知られていたが、近年、その現実的な存在意義が増してきている。大数の法則が成り立たない状況は、平穏な日々が続く中で、突然想定外の出来事が起こる場面のモデル化とみなすことができる。想定外の出来事とは、ときに大災害を起こす豪雨であり、ときに金融危機をもたらす株価の大暴落と、近年の社会の大問題にも例を見出すことができる。そこでは、大多数の普通のことよりも、極少数の極端なことが支配的になる。このような極端な出来事の研究は、リスク管理の数理をはじめ、様々な場面で用いられ、かつての“想定外”を明日の“想定内”に捉えることを目指している。



Winners Take All:

所得や資産は対数正規分布やパレート分布に従うといわれ、少数の大金持ちが平均を引き上げて、多数が平均以下になってしまふ。2019年現在、全人類の所得下位半分38億人の総資産と同じ額を上位26人がもつという。超格差社会では、平均を定められないのかもしれない。