

会員ニュース

青嶋誠先生の令和2年度文部科学大臣表彰 科学技術賞受賞に寄せて

筑波大学数理物質系
矢田 和善

青嶋誠先生（筑波大学数理物質系教授）が、令和2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）を受賞されました。受賞の対象となった研究業績は、「高次元統計解析による高次元現象の解明に関する研究」です。御受賞を心よりお慶び申し上げます。青嶋先生のこれまでの御功績について、青嶋先生の弟子のひとりとして私が存じ上げる範囲で、概略を述べたいと思います。

青嶋先生の膨大な御研究の中でも、特筆されるべきものは、高次元データに代表される大規模複雑データを高速かつ高精度に処理するための統計数理の構築です。その理論と方法論は体系化され、高次元統計解析という新しい統計学が創生されました。これは、従来の統計学や機械学習の常識を覆す極めて独創的かつ先駆的なもので、世界の統計科学とその周辺領域に革新をもたらすきっかけとなりました。青嶋先生は2005年頃からこの研究に取り組み、新しい高次元漸近理論の開拓、高次元データの幾何学的表現の発見、高次元PCAの考案、高次元ノイズの確率解析、高次元ノイズを除去するデータ変換法の考案、高次元データを高速かつ高精度に解析する非スパースモデリングの開発など、数多くの顕著な業績を積み上げられ、高次元統計解析に結実されました。私は大学院生の頃から、これまで10年以上にわたり、青嶋先生と共同研究をさせて頂いております。青嶋先生の研究に対する真摯な姿勢と、謙虚でありながらも併せもたれた卓越した発想力と絶大な研究遂行力には、ただただ感服するばかりでした。その御功績は、このたびの文部科学大臣表彰科学技術賞（研究部門）の受賞や、2017年の日本統計学会賞の受賞、2012年のAbraham Wald Prizeの受賞など、国内外での数々の受賞や国際会議での多くの基調講演へと結実されています。

ここからは、高次元データ解析の歴史に触れながら、高次元統計解析のパイオニアである青嶋先生の御研究が如何に独創的かつ先駆的なものであるかを説明したいと思います。「高次元漸近理論と高次元PCA」、「高次元ノイズの確率解析とデータ変換法」、「非スパースモデリング」という3つのテーマに絞ってお話します。

まずは、青嶋先生が開拓された基礎理論である「高次元漸近理論と高次元PCA」についてお話します。1990年代後半より情報化の進展に伴い、ゲノム科学・情報

工学・金融工学などの方面から、様々な高次元データが出現しました。そのデータの一つの特徴は、数十程度の標本数 n に対してデータの次元数 d が優に数万を超えるというもので、しばしば高次元小標本 ($d \gg n$) データとよばれます。従来の統計学は、 $n > d$ を前提としているため、高次元データに対しては全く歯が立ちません。2000年代に入り、高次元データ解析の理論研究が徐々に始まりました。その一つの潮流が、ランダム行列理論です。ランダム行列理論の世界的大家であるスタンフォード大学の Johnstone 教授は、 d と n がどちらも十分に大きな高次元大標本 ($d/n \rightarrow c$) の設定で、すべての固有値に有界性を仮定して、Tracy-Widom 分布に基づく標本固有値の極限分布を導出しました (Johnstone, 2001)。これ以降、多くの研究者がこの結果に基づいて高次元における PCA の研究を進めていきました。しかし、青嶋先生は、この潮流に当時から否定的な見解をおもちでした。その根拠として、実際の高次元データの固有値は次元数が大きくなるにつれて増加し、特に最初の数個の固有値が次元数の冪乗で発散することを示されました。そして、ランダム行列理論が成立しないこのような状況において、しかも実際の高次元データ解析と同じ小標本の設定で漸近理論を展開するために、新たに高次元小標本 ($d/n \rightarrow \infty$) 漸近理論を開拓されました。この新しい漸近理論は、青嶋先生が発見された高次元データの幾何学的表現に依拠しています。青嶋先生は、高次元において巨大化するノイズの確率変動を幾何的な構造として特定し、高次元データに内在する次元の呪いを理論的に解明されたのです。例えば、高次元小標本における高次元ノイズは次元数が大きくなるにつれて幾何的な構造を顕わにさせ、膨張し、その結果、標本固有値や標本固有ベクトルが高次元小標本において不一致性を起こしてしまいます。青嶋先生は、こういった次元の呪いを解決するために、膨張する高次元ノイズの確率的挙動を精確に捉えそれを潜在空間から掃き出すことで、高次元小標本でも一致性をもつ高次元 PCA を考案されました。これは、ノイズ掃き出し法とよばれています。ノイズ掃き出し法は汎用性が高く、例えば、高次元データの方法論における世界的権威であるプリンストン大学の Fan 教授は、ノイズ掃き出し法を応用して、S-POET とよばれる高精度な高次元共分散行列の推定法を開発しています (Wang and Fan, 2017)。青嶋先生は、ノイズ掃き出し法だけでなく、幾何学的表現に基づく高次元統計解析の様々な方法論を分類的に考案されています。詳細については、青嶋・矢田による論説 (数学 65, 2013) をご参照ください。

次に、「高次元ノイズの確率解析とデータ変換法」についてお話しします。青嶋先生は、高次元ノイズの確率解析の研究を進められ、高次元ノイズが潜在情報の増加よりも速く発散するという高次元現象の解明に着手されました。そして、高次元統計量のリンデベルグ条件を精密に評価することで、高次元統計量のマルチンゲール中心極限定理による漸近正規性の成立は固有値の指数 0.5 が境界になることを明らかにさ

れ、高次元ノイズを弱スパイク (NSSE) と強スパイク (SSE) とに分類されました。この分類が、高次元統計解析に精度保証を与えるための鍵となるのです。NSSE は高次元ノイズのスパース性に対応します。SSE の場合、高次元ノイズの非スパース性が漸近正規性を壊し、統計的推測に精度が保証されません。例えば、著名な理論統計家であるアイオワ州立大学・北京大学の Chen 教授が提案した高次元二標本検定法 (Chen and Qin, 2010) は広く知られていますが、実は SSE において、この検定法は第一種の過誤さえ制御できず間違っ了解析結果を与えることが証明できます。困ったことに、SSE の状況は実際の高次元データ解析でごく普通に起こります。青嶋先生は、この問題を解決するために、高次元ノイズを SSE から NSSE に変換するという斬新な発想を提唱され、高次元ノイズが非スパースであっても高次元統計量の漸近正規性を成立させるデータ変換法を考案されました。先の高次元二標本検定も、データ変換法を使えば、SSE において第一種の過誤を制御することができ、さらに、高い検出力を保証することもできます。青嶋先生は複数の企業や他分野とも共同研究を推進されています。先端材料技術の開発、次世代シーケンサーによるゲノム解析、天体物理学・素粒子物理学など多くの研究領域で、大規模複雑データの高次元ノイズが SSE になること、そして、それをデータ変換法で除去することの必要性と重要性が認識されてきています。今後、データ変換法を用いた高次元統計解析が、大規模複雑データの解析においてスタンダードとなることが予想されます。

最後に、「非スパースモデリング」についてお話しします。1990 年代後半、高次元データを解析するための方法論として、機械学習が目覚ましく発展してきました。例えば、高次元小標本からの情報抽出技術のひとつに、スタンフォード大学の Donoho 教授や Tibshirani 教授らが発展させたスパースモデリングがあります。スパースモデリングは、高次元データにスパース性を仮定して、高次元データの特徴を少数の変数や標本で捉えようとする方法論です。代表的な技法に、2018 年にガウス賞を受賞した Donoho 教授が提唱した圧縮センシングがあります (Donoho, 2006)。いまや、スパースモデリングは流行りといえるでしょう。この風潮に対して、青嶋先生の発想は全く異なるものでした。高次元データの潜在情報は非スパースであり、高次元ノイズも非スパースであり、それゆえ、非スパース性と向き合うことで高次元データをありのままに生かすことができる。青嶋先生のこの発想によって、高次元データの非スパース性が幾何的な構造として理論的に浮き彫りになることが証明され、統計的推測の精度向上と計算コストの大幅削減という技術的な革新に繋がりました。この一連の方法論は、非スパースモデリングとよばれています。例えば、サポートベクターマシン (SVM) という分類器は広く普及したスパースモデリングのひとつですが、標本数不均衡の場合には、解に強不一致性を起こすことが証明できます。これを解決するた

めにはある特殊な補正が必要になるのですが、実は、補正後の SVM が非スパースモデリングと漸近同等になります。青嶋先生が開発したこの非スパースモデリングは、数百万次元の超高次元データをノートパソコンでもたった数分で高精度に解析できる革新的な技術で、次世代シーケンサーによるゲノムデータの識別に成功するなど、多くの重要な実績を上げられています。

ここまでは、青嶋先生の研究面での御功績をお話ししました。最後に、管理・運営、社会貢献、人材育成の面から、青嶋先生の御功績をお話ししたいと思います。

青嶋先生は、日本数学会の統計数学分科会評議員を務められ、受賞候補推薦委員や解析学賞委員会委員を歴任されました。日本統計学会では理事と代議員を長年務められ、学会誌の編集長を和文誌と欧文誌とで歴任されました。統計関連 6 学会連合からも国際学術誌の初代編集長として招かれ、Springer から「Japanese Journal of Statistics and Data Science」を創刊されました。現在、統計学界のトップジャーナルである「Journal of the American Statistical Association」をはじめ、「Journal of Multivariate Analysis」や「Statistica Sinica」など、統計学界の第一級の国際学術誌ばかり計 6 誌の編集委員を務められています。また、国際統計協会の日本代表を務められ、700 名を超える規模の環太平洋地域会議を開催するなど、我が国の国際的プレゼンスの向上に多大な貢献をされています。日本学術振興会では科研費委員会の幹事を務められるなど中心的な役割を担われ、日本学術振興会科研費審査委員表彰を 2 度、受賞されています。現在は、日本学術会議の連携会員を務められ、我が国の科学界全体の舵を取られています。また、国立研究開発法人科学技術振興機構 (JST) など国や企業の諮問委員も多く務められています。筑波大学では、数理科学研究コアを発足され、文部科学省委託事業「数学アドバンストイノベーションプラットフォーム (AIMaP)」の拠点代表として、産学独との連携を推進されています。

一方で、青嶋先生は、人材育成にも大変な熱意をおもちです。私が存じ上げるだけでも、高次元統計解析に関連して科研費基盤研究(A)や(S)などの外部資金を受けられ、10 年以上も継続して、毎年 5 回ほど科研費シンポジウムを日本全国で開催されています。国内のみならず海外からも一流研究者を招聘し、若手研究者の発表を奨励して、分け隔てなく交流の場を提供されています。また、学生からの青嶋研究室の人気は絶大で、毎年、個性の強い、向上心溢れる学生が多く集まっています。私自身も教えを受けたひとりですが、青嶋先生は、研究指導だけにとどまらず、プロとしての仕事の取り組み方やプレゼンテーションの心得なども、熱心に指導されます。大学院のゼミでは、世界最先端の研究テーマを大学院生に惜しみなく与えられ、研究プロジェクトの一員として信頼し迎え入れることで、彼らをプロの研究者に育て上げられています。日本数学会や日本統計学会などのコンペティションにおいて、青嶋研究室の大学院生

が毎年受賞していることも、それほど不思議なことではないかもしれません。

青嶋先生の御功績は、研究だけにとどまらず、管理・運営、社会貢献、人材育成においても傑出され、多方面で絶大なる影響力を発揮されています。常に時代の先陣を切って新しい発想で私たちに光明を与えてくださる青嶋先生こそ、まさに稀代のパイオニアと言えるのではないでしょうか。これからもその影響力を存分に発揮され、我が国のみならず世界の科学界を牽引してくださることを、心から願っております。