

会員ニュース

加藤賢悟さんの第16回日本学術振興会賞および 日本学士院学術奨励賞の受賞を祝して

一橋大学大学院経済学研究科

本田 敏雄

加藤賢悟さんが、「ビッグデータと高次元データ解析における統計理論研究」により第16回（令和元年度）の日本学術振興会賞および日本学士院学術奨励賞を受賞されました。日本学術振興会賞は、大変優れた業績を挙げた若手研究者に贈られる荣誉ある賞です。加藤さんはその賞に加えて、日本学術振興会賞受賞者から特に優れた6名以内の方を選んで授賞される日本学士院学術奨励賞も受賞されました。東京大学の久保川達也先生から同じくご指導いただいたものとして、心よりお祝い申し上げます。

まず加藤さんの経歴から簡単にご紹介させていただきたいと思います。2005年に東京大学経済学部を卒業され、2007年には同修士課程を修了し、2009年には同博士課程を早期修了されています。加藤さんの優秀さは大学院時代から有名でしたが、私が親しくお話ししたのは2007年の11月の一橋大学での科研費集会が最初だったように思います。大学院修了後は、2009年春から4年間広島大学数学科で助教をされ、1年間のMITでの在外研究も経験されました。その間マイアミビーチの学会で一緒に食事をしたのを覚えています。その後は東京大学大学院経済学研究科のテニュアトラック講師、准教授を経て、2018年夏よりコーネル大学統計データサイエンス学部の准教授として活躍されています。加藤さんのご研究は相互に密接に関連しながらも大変多岐にわたっています。そのなかから、これらの賞の授賞理由にある高次元統計解析、分位点回帰分析、ノンパラメトリック推定について、私なりに以下に解説させていただきます。

まず国際的評価の非常に高い、高次元統計解析に関するご研究ですが、これは近年様々な分野で重要性が強く言われているビッグデータの一つの形態である高次元データに関するものです。医学における遺伝子データなどがその典型例ですが、線形回帰モデルを言えば、標本数 n に対して説明変数の数 p が非常に大きく ($n \ll p$)、従来の統計学の手法が全く適用できないようなデータの事です。加藤さんは、共同研究者であるMITの Chernozhukov 教授、UCLAの Chetverikov 准教授らと、高次元データ解析の基礎となる中心極限定理(正規近似)に関して大変重要な貢献をされています。科学技術とコンピュータの飛躍的進歩に伴って高次元データに関する研究自体が始まってからは、このような高次元データの場合には、まず適当な方法で説明変数に関する変数選択(モデル選択)を行い、その後に従来の統計学の手法あるいはその拡張を用いるという解析手法が主

流でした。しかしながらそのような手法には、モデル選択後の統計的推測であるという問題があり、高次元データの構造を維持したままの統計的推測の必要性が広く認識されていました。また変数選択の方法に関する理論にも、確率的な評価ができていないなどの不十分な点が多くあります。これらの問題を解決するためのいくつかの試みがありますが、すべての基礎となるものはやはり高次元の確率変数に対する中心極限定理(正規近似)であり、加藤さんらの貢献は今後の数多くの研究の基礎となるものです。以下その貢献の最初のものとなる2013年のAnnals of Statistics誌の結果を簡単に紹介させていただき、ついでその後の発展についてふれて行きたいと思います。

まず X_1, \dots, X_n を、各成分の母平均が0である同一分布に従う p 次元 ($n \ll p$) 独立確率変数列とし、 $S_n^X = (S_{n1}^X, \dots, S_{np}^X)^T = \sum_{i=1}^n X_i / \sqrt{n}$ とおきます。検定を含む様々な統計的推測を行う場合には、 $T_0 = \max_{1 \leq j \leq p} S_{nj}^X$ が重要であり、その T_0 の分布を考えるために、 X_1, \dots, X_n と同じ平均と分散の構造を持ち、互いに独立に正規分布に従う Y_1, \dots, Y_n を考え、 $S_n^Y = (S_{n1}^Y, \dots, S_{np}^Y)^T = \sum_{i=1}^n Y_i / \sqrt{n}$ および $Z_0 = \max_{1 \leq j \leq p} S_{nj}^Y$ と定義します。加藤さんらは、Slepian 補間、Stein の leave-one-out 法、最大値関数の微分可能な関数での近似、正規過程の非集中性などに関する結果を駆使して、 T_0 の分布は Z_0 分布で近似できることを示しています。しかしながら高次元データの場合においては、 Z_0 の適当な基準化の下での極限分布というものを考えることができないので、この正規近似の結果だけでは統計学には応用ができません。そこで加藤さんらは、multiplier bootstrap と empirical bootstrap という二つの実用的な手法により Z_0 の分布が近似的に得られることも示しています。そういう意味では、一つの論文で完全に問題を解き切っていると言えます。極限分布のない場合の統計的推測の問題を鮮やかに扱うというのも、加藤さんたちの研究の特長です。

その後この論文の結果は、加藤さんとその共同研究者によって、より一般的な形や高次元 U 統計量に関する結果へと拡張されています。特に2017年のAnnals of Probability誌の結果では、最大値から矩形のクラス(正確にはより一般的なクラスまで) A と結果を拡張し、 $\sup_{A \in \mathcal{A}} |P(S_n^X \in A) - P(S_n^Y \in A)|$ の最適にかなり近いと思われる上限を与えています。先の最大値は、 $A = I\{x_j \leq a, 1 \leq j \leq p\}$ の場合となっています。

次に分位点回帰に関する結果を紹介させていただきます。分位点回帰は L_1 (メディアン) 回帰の一般化で、リスク管理への応用、平均回帰のみではデータに関する情報が不十分であること、計算機の進歩で推定量が計算可能になったことなどから、近年統計学、計量経済学における重要な研究分野の一つになっています。加藤さんはアリゾナ大学のGalvao教授らの共同研究者とパネルデータにおける分位点回帰の研究を行い、ここでも国際的に顕著な貢献をしています。計量経済学で言えばパネルデータ、統計学で言えば経時測定データですが、計量経済学と統計学での扱いの大きな違いは、個別効果を入れるか入れないかです。計量経済学では、特定の個人あるいは企業などから継続的かつ定期的

にデータをとるということを強く意識しているためか必ず個別効果を入れますが、この個別効果のパラメータ数は観測の対象数と同じため大変大きくなります。平均回帰のときにはこの個別効果は簡単に消去できるのですが、分位点回帰のときには何らかの形で推定するか代替的な回帰を考えるしかなく、非常に扱いにくくなっています。大変著名な研究者ですら誤った結果しか出せないなか、加藤さんたちはこの問題に対しても興味深くかつ重要な結果を得ています。このパネルデータに関する貢献以外に、関数線形モデルにおける分位点回帰に関する単独研究があります。関数データの解析では推定量の計算を行うのが難しく、加藤さんの計算能力の高さにも驚かされます。またデューク大学の Belloni 教授、Chernozhukov 教授との高次元線形分位点回帰の論文では、Neyman 直交性の概念を用いて、高次元線形分位点回帰モデルにおいての特定の係数の信頼区間と検定の問題を大変うまく扱っています。

最後にノンパラメトリック関数推定の一様信頼区間の構成に関する結果を紹介させていただきます。密度関数、回帰関数などをノンパラメトリック推定し、それらの関数の引数に関する一様信頼区間を構成するという問題は、ノンパラメトリック関数推定では古くからの重要な問題です。これまでは、ノンパラメトリック推定量の広義の分布収束、そしてその分布収束先である正規確率過程の引数に関する最大値の極限分布に基づいて一様信頼区間が構成されていました。加藤さんたちは、ご自身たちの高次元中心極限定理に関する結果からの類推でしょうか、推定量の分布収束を考えることなく、正規近似と正規過程の非集中性のみによる一様信頼区間の構成法を考案しました。ここでの信頼区間の幅は multiplier bootstrap などで求めています。そしてこの斬新な手法により、いくつかの未解決問題が解決されました。以上に紹介した研究以外に、加藤さんにはここでは紹介しきれない、正規過程の非集中性自体の研究、正規過程間の比較定理などの多くの顕著な業績があります。

加藤さんは 2018 年夏より活躍の場をアメリカに移されました。我々周囲の人間も、いずれはアメリカの大学に移籍されるだろうなどは思っておりましたので、大変残念には思いましたが、そう意外な印象はうけませんでした。2017 年 12 月のロンドンでの CMStatistics2017 の招待セッションで、オーガナイザーの UC Davis の J.-L. Wang 教授が講演者の紹介の際に、加藤さんが大変若いので驚いた、と話されたことがとても印象に残っています。加藤さんはまだ大変若く、統計学も今後も高次元データ解析のような困難かつ重要な課題に出会うことでしょう。そのときの加藤さんのご活躍を楽しみにしています。