

## 第36回小川研究奨励賞受賞のことば

石井 晶

この度は小川研究奨励賞という名誉ある賞を賜りまして、大変光栄に存じます。ご推薦頂きました先生方、選考委員会の先生方、東京理科大学でお世話になっております先生方に感謝申し上げます。特に、学部3年次より熱心にご指導を頂きました青嶋誠先生、共著者である矢田和善先生、大学院生時代から様々なご助言を頂きました赤平昌文先生に、心より御礼申し上げます。

私が受賞論文の評価にある高次元統計解析に出会うきっかけとなったのは、学部3年次に履修した、青嶋先生が担当される数理統計学の授業でした。数理統計学の面白さが最大限に伝わる授業展開はもちろん、高次元統計解析の研究について、大変楽しそうにお話されているお姿が、特に印象的でした。当時、私は卒業研究や大学院進学先の研究室選別に悩んでおりましたが、青嶋先生の初回の授業を受けて、青嶋研究室で高次元統計解析の研究に取り組みたいと、強く希望することとなりました。

高次元統計解析は、ゲノム等の非スパース性に特徴をもつ高次元小標本データに対し、新たな理論・方法論を構築するものです。「高精度・高速計算・汎用性」という3つの特徴を持ちます。高次元統計解析では、従来の多変量解析では太刀打ちできない高次元小標本データに対し、たった数十の標本で高い精度保証を与え、モバイルPCでさえ数分で計算できるほど計算コストが低く、実際の高次元小標本データに対して、非現実的な仮定はおきません。高次元統計解析では、高次元小標本データを潜在空間とノイズ空間に分けて捉え、ノイズ空間の精密な評価によって高い精度保証を与えます。高次元データ解析が困難である理由のひとつは、高次元小標本データが、潜在空間を覆い隠すほどの巨大なノイズを含み、その扱いが非常に厄介であるからです。その巨大なノイズは、Aoshima and Yata (2018, Sinica)で提案された弱スパイク固有値モデルと強スパイク固有値モデルという2つのモデルで記述されます。2つのモデルの大きな違いは、精度保証を与える際に、高次元の中心極限定理が成立するか否かです。

私はこれまで、高次元の中心極限定理が成立しない、強スパイク固有値モデルに対する高次元統計解析の研究を進めて参りました。ゲノム等の高次元小標本データを解析する際、遺伝子間の強い相関や異常値の混入等によって、強スパイク固有値モデルになることがしばしば見受けられます。Journal of Multivariate Analysisの50周年記念特別号に招待論文として掲載されたIshii, Yata and Aoshima (2022)では、高次元小標本データの幾何学的表現と強スパイクするノイズを除去するデータ変換の技術を用いることで、高次元小標本データを高精度に識別する判別方式を与えています。論文はオープンアクセスとなっておりますので、ご興味のある方はご覧頂ければ幸いです。

授賞式の数か月前に第1子を出産した私にとって、本年度に小川研究奨励賞を受賞させて頂けたことは、今後の研究活動にとって大変励みになります。今後は研究と育児の両立

が大変にはなりますが、今回の受賞を糧に、周りの先生方にご助言を頂きながら、研究活動に精進して参ります。今後とも、ご指導ご鞭撻のほど宜しくお願い申し上げます。