

AIP 加速課題

2021 年度採択研究代表者

2021 年度 年次報告書
------------------

津田 宏治

東京大学 大学院新領域創成科学研究科  
教授

信頼できるデータ駆動科学のための統計学の深化

## § 1. 研究成果の概要

生物学・医学分野では、さらに高次元化が進み、それに伴って複雑な機械学習アルゴリズムの使用が増えているため、偽陽性が起きる危険性は、これまでになく高まっている。多数の仮説から少数を選択する際の操作に条件づけて、有効な検定値（P 値）を計算する**選択的推論法**は、高次元データに対しても有効な手法として、注目を集めている。本研究課題では、選択的推論法を中心に優れた統計検定法を提案し、それを生命科学分野に応用することによって、信頼できる統計科学の構築を目指している。

津田 G と竹内 G では、選択的推論アルゴリズムの開発を行っている。今年度は、ホモトピー法を用いた選択的推論法を開発し、従来の選択的推論法よりも検出力が高まることを確かめた<sup>1)</sup>。また、高次相互作用の検定に関して新たな方法を提案した<sup>2)</sup>。

門松 G では、Th-MYCN マウスモデルから取得した一細胞遺伝子発現データについて解析を進めた結果、神経芽腫細胞の運命分岐(がん化の促進か自然退縮)に関わると考えられる細胞クラスターを見出した。これらの解析結果については 2022 年中に論文化する予定で、並行して実験的な検証を進める。

瀬々 G では、大規模な健康診断情報(検診データ)への LAMP 法の応用を見据えて、年齢や疾病の種類や要因、健康状態のベースラインが異なることが、交絡因子となることを防ぐため、年齢や疾病の種類による層別化、また、発症年を基準として「発症 N 年前」の検診データとなるようにデータを変換するなどの対策を講じた。

山田 G では、遺伝学に選択的推論を適用する際の諸問題について考察を行った。特に、ゲノムワイドアソシエーションスタディ(GWAS)における要素ごとの帰無仮説棄却検定の多重検定補正を選択操作とみなしたときの、要素ごとの効果量区間推定における選択的推論について数理的な整理を行った。統計検定は頻度論的な統計科学の手法であるが、選択的推論を適用するとベイズ主義を加味した解釈が自然であることが示された。

## § 2. 研究実施体制

### (1) 津田グループ

- ① 研究代表者: 津田 宏治 (東京大学大学院新領域創成科学研究科教授)
- ② 研究項目
  - ・アルゴリズム開発と量子計算機応用

### (2) 門松グループ

- ① 研究代表者: 門松 健治 (名古屋大学大学院医学系研究科教授)
- ② 研究項目
  - ・癌科学への展開

### (3) 瀬々グループ

- ① 研究代表者: 瀬々 潤 (株式会社ヒューマノーム研究所代表取締役社長)
- ② 研究項目
  - ・健康科学への展開

### (4) 竹内グループ

- ① 研究代表者: 竹内 一郎 (名古屋工業大学大学院工学研究科教授)
- ② 研究項目
  - ・汎用的な選択的推論手法の構築

### (5) 山田グループ

- ① 研究代表者: 山田 亮 (京都大学大学院医学研究科教授)
- ② 研究項目
  - ・数理的基礎

### 【代表的な原著論文情報】

- 1) K. Sugiyama, V.N.L. Duy and I. Takeuchi, More Powerful and General Selective Inference for Stepwise Feature Selection Using Homotopy Method, Proceedings of the 38 th International Conference on Machine Learning, 2021.
- 2) D. Das, V.N.L. Duy, H. Hanada, K. Tsuda, I. Takeuchi. Fast and More Powerful Selective Inference for Sparse High-Order Interaction Model. 36th AAAI Conference on Artificial Intelligence, 2022.