

# 計量生物学との交流からもたらされる数理統計研究の広がり・深化

統計数理研究所 数理・推論研究系 二宮 嘉行

本発表では、九州大学数理学研究院のスタッフとして培われた経験、スタンフォード大学での留学の際の経験をもとに、計量生物学に足を入れたい数理統計学者が今後どのように思っていくべきか、また計量生物学の方々にどう思っていたらいいか、について考えてみたい。以下は、その考えのもとでスタートした二つの研究の概要である。

## 1. 疑似相関を用いた多重性調整

グループ間の多次元データの平均を比較するという基本的な多重検定問題において、多次元データの相関が未知である設定を考える。このとき、必ずしも一致推定量にはならない統計量、具体的にはある種の標本疑似相関を相関に代入して用いることにより、多重検定のサイズを漸近的にコントロールしつつ検出力を上げる、という方法を開発した。また、標本疑似相関のあるクラスの中で、ある種の最適性を満たすようなものを与え、それを実際に代入して用いるときに生じる数値計算上の問題を克服した。また、検出力を上げる既存手法であるステップダウン法と組み合わせ、さらなる検出力向上を達成した。数値実験では既存手法と比較し、常に検出力を上げることともに、設定によってはステップダウン法によるゲインよりはるかに大きなゲインがあることを示した。実データ解析では、コントロールマウスとその染色体の一本を別のマウスのもので入れ替えたコンソミックマウスに対し、身長や体重などの特徴量を比較する多重検定問題を扱った。

## 2. 因果推論におけるセミパラメトリックアプローチのための情報量規準の開発

最も単純な周辺構造モデルのモデル選択問題に対し、傾向スコアを用いたセミパラメトリック推定法に対処できる情報量規準を開発した。周辺構造モデルとは因果推論において基本となるモデルであり、ある種のデータが欠測している潜在結果変数モデルとみなせるものである。その欠測メカニズムと結果変数に相関があるにもかかわらずナイーブに推定をおこなうと、推定値がバイアスをもつことになるため、傾向スコアを用いた逆確率重み付け推定や二重頑健推定がよく用いられる。しかし、この推定法に対処できるモデル選択法はこれまで十分に開発されておらず、数理的に妥当ではない情報量規準が用いられていた。そこで、情報量規準元来の考え方に基づいてそれを導き、それは既存の情報量規準とは大きく異なることを示した。そして、数値実験により新しく導いた情報量規準と既存の情報量規準を比較し、真の構造を含むモデルあるいはそれに近いモデルを選択する確率について、いずれのシミュレーション設定においても前者は後者を上回ることを確認した。

## 参考文献

- [1] Ninomiya, Y., Kuriki, S., Shiroishi, T. and Takada, T. (2016). Use of spurious correlation for multiplicity adjustment. *arXiv preprint*, 1612.06029.
- [2] Baba, T., Kanemori, T. and Ninomiya, Y. (2017). A  $C_p$  criterion for semiparametric causal inference. *Biometrika*, 104, 845–861.