

計量生物学研究の面白さ

菅澤 翔之助 (東京大学 空間情報科学研究センター)

私は東京大学空間情報科学研究センターに所属しており、普段はベイズ統計学や階層モデル、時空間統計学に関する方法論を中心に研究をしています。私が計量生物学の研究に取り組み始めたのは、統計数理研究所で特任研究員をしていた際に、野間久史先生のもとで研究活動を行なったことが大きなきっかけです。着任前は階層モデルや混合効果モデルの研究を行っていたため、漠然と計量生物学との関連性は意識していたのですが、具体的な応用事例や研究テーマには触れる機会はありませんでした。特任研究員時代には、医療ビッグデータ解析に関する CREST プロジェクトへ参画し、本格的な計量生物学の研究を経験させていただきました。今回の記事を書かせていただくきっかけとなった奨励賞は、特任研究員時代に野間先生と共同で行なった研究が受賞対象となっております。

現職に就いてからは、必ずしも本格的な計量生物学の研究活動を行えているわけではないのですが、計量生物分野における研究動向は可能な範囲でフォローするようにしています。その最大の理由としては、他の統計学関連分野に比べても、計量生物学には面白い解析例や分析のモチベーションが非常に多く転がっていると感じるからです。これは、統計学を研究する人間にとって、面白い研究テーマを見つけることができるネタの宝庫であると思っています。実際に、計量生物学における分析上のリサーチクエスチョンから派生した統計的問題に関して、これまでにいくつかの研究を行うことができました。最近では、統計学が専門の周囲の研究者仲間や学生に、計量生物学周辺の論文も読んでみることを勧めています。

計量生物学における方法論研究の(個人的に思う)面白さの例として、分析に全く同じ統計モデルを用いているにも関わらず、分析データの種類によって、興味のあるモデルの側面や性質が大きく異なる点があります。例えば、メタアナリシスやマイクロアレイデータの解析で用いられている単純な正規階層モデル(観測データが正規分布に従う階層モデル)を考えてみます。両者ともに同じモデルを使った解析が行われるのにも関わらず、注意が向けられるモデルの特徴は互いに大きく異なります。現実のメタアナリシスでは、サンプルサイズが小さいことが多い故に、モデルの仮定は許容した上で平均パラメータの正確な推論(正確な統計的誤差の評価)が求められることが多いと思います。一方で、マイクロアレイデータの解析では、一般的に多くのセル上のデータが取れるため、統計的誤差はメタアナリシスほど気にする必要がなく、モデルの誤特定やより柔軟なモデルの利用に注意が向けられることが多いと思います。このような違いは、2つの事例における標準的なサンプルサイズの違いによって引き起こされる差異であると捉えることができます。このように、同じモデルに対しても、解析データの実際の特徴に応じて重視すべき統計的な性質が大きく異なってくる点は、計量生物学における非常に興味深い側面の1つではないかと考えています。また、計量生物学において何か新しいタイプのデータが登場したときに、仮に既存のモデルが使えるようなデータだとしても、既存の統計理論がカバーできない側面に分析の興味が向けられる可能性があることも示唆していると思います。このような新たな問題は、実応用から直接的に上がってくる価値の高い(かつ面白い)統計的問題として研究され、計量生物分野のみならず統計学分野の発展にも寄与していくのではないかと思います。私としても、このような面白い問題に出会えるように日々アンテナを張っていければと思っています。

最後になりますが、この度は奨励賞を授与いただきまして、ありがとうございました。今後とも、計量生物学における魅力的な分析課題に多く触れ、統計学研究者として何らかの有効な解を与えられるような研究を進めていき、分野の発展に少しでも貢献していければと考えております。