

近づきたいよ 君の理想に
林 賢一 (慶應義塾大学 理工学部)

「計量生物学の未来に向けて」という題目で何かを語れるほどの見識をもたないまま、この原稿を依頼される程度には統計学の世界に居続けてしまいました。未来に向けて何をなすべきか、と問われると反射的には「好きなことをやればよいのでは」としか答えられません。なので、過去の点を線で繋ぎ、未来に外挿してそれらしいことを絞り出そうと思います。

私は、心理学に漠然とした憧れを抱いて、大阪大学人間科学部に入学しました。授業を受けると、漠然と想い描いていた心理学像と実際は大きくかけ離れていて、自分には教養として吸収する以上のことはできないと感じるようになりました。その傍らで、心理の調査データを分散分析するか、因子分析にかけるとかリサーチアシスタントの大学院生が「なんか魔法っぽいもの」を駆使してデータの読み方を教えてくれる機会がありました。そして、この魔法の仕組みを理解したいと、統計学に興味を持ちました。狩野裕先生が担当された必修科目の統計学も衝撃でした。事物が抽象化されて議論に取り込まれ、それらがまた具体的な事象と関連をもつ感じが、数学が（不得意ながら）好きだった自分に合っていました。その後狩野先生の行動計量学研究室に入り、よりきちんと統計学の理論を勉強すべく、同大学大学院基礎工学研究科に進学しました。大学院では、機械学習の理論面について興味を赴くまま論文を読んでいました。ここまで、計量生物学との接点はありません。

転機は、最初の就職が同大学大学院医学系研究科だったことです。研究員として、観察研究のデータ解析を担当することになりました。まともにデータを解析した経験もなく「ロジスティック回帰？教師あり学習でしょ」と結果の出力はできるものの、医学研究者が見たい数字とその意味について満足に理解していない、それ以前にコミュニケーションが成立しない、ということを経験する日々でした。数学的命題の証明とは別種の難しさに、無力感から、今まで学んできたことが机上の空論、無用の長物にすら感じることもありました。幸いに、統計解析に強い興味を持った博士課程の医学研究者がおり、その方とデータ解析の仕方についてやりとりするうちに、医学研究の機微が理解できるようになっていきました。あるとき、その方の解析結果が、指標 net reclassification improvement (NRI; Pencina et al., 2011, Stat Med) によると、あるバイオマーカーがモデルの予測性能を有意に改善すると読める場面に遭遇しました。ちょうど、NRI などの指標が有意差を出しすぎてしまう問題 (Hilden and Gerds, 2014, Stat Med; Pepe et al., 2015, Stat Biosci など) に興味を持っていた私は、データを慎重に確認し、いま正にその「出しすぎ」の状況が起きていることを突き止めました。ある意味、解析結果に水を差すような指摘でしたが、状況と問題を丁寧に説明することで、議論は平和裡に収束しました。このときの経験が Hayashi and Eguchi (2019, Stat Med) に結実し、実験や反例により批判された NRI などがなぜ問題なのかを統計的性質と結び付けて示すことができました。この記事をご覧になる「王道」を往く方々は、このような学びは早い段階で獲得されると思いますが、邪道の私には一定の年月を費やす必要がありました。

ここまで書いて、計量生物学を背景とし、医学・医療の発展の一翼を担うためには具体と抽象の世界を軽快に行き来できる力が必要なのではないか、と思ひ至ることができました。そして、二つの世界を繋ぐための「近似」をうまく与える必要があると思います。何を言っているのかわかりにくいような気もしま

すが、チェビシェフの不等式で有名なチェビシェフのラテン語表記には Chebychev や Tchebycheff などあり、日本語でもチビシヨフやチェブイシェフなどの表記があります (今となつては、ありません、かもしれません)。要は、ある世界の「ことば」を別の世界の「ことば」に移すときには近似が必要で、またそれが一意ではないということです。医学研究における現象を抽象化し、統計的な問題に落とし込む。そしてその結果を、本質を損なわず、理解しやすい形で提示し、議論に至る。これらの行動すべてに「近似」が伴います。それゆえ、研究者の「近似」の仕方が研究の個性や良し悪しに繋がっていると考えられます。研究はひとつのデータ解析、一篇の論文で完結するものではなく、様々な研究者と協働して近似のサイクルを多く回すことによって真実に迫る活動ではないかと思えます。統計モデル自体が現象の近似であることは、いうまでもありません。それゆえ、統計科学に携わる者も統計モデルの妥当性、仮定を含めた適用妥当性、結果の解釈可能性、一般化可能性などの検証が常に求められます。これらの「近似」を弛みなく続けていく努力が、計量生物学の未来を照らすように思われます。私は現在、理工学部数理科学科に在籍しており、応用数学の立場から学生の教育にも従事しています。数学を武器とする学生に、統計科学を礎とした計量生物学に広がる楽しさと、この分野へ貢献できる喜びを伝えられるよう努力して参ります。