

まえがき

忘れもしない 2015 年 9 月に岡山大学で開催された 2015 年度統計関連学会連合大会にて中央大学の鎌倉稔成教授より、この「統計学 One Point」シリーズの計画をうかがった。筆者は勉強不足のため、このようなシリーズが他の分野では多数発行されていることを初めて認識し、自分には怖れ多いことと直感的に感じたのだが、鎌倉先生は「これは良い機会だと思って、ぜひ引き受けてほしい」と上手に筆者の退路を断たれた。

逆にお引き受けすると決めてかかると、幸いにも統計関連学会連合大会中ということもあり、統計家が大挙して参加している訳である。久留米大学バイオ統計センターに赴任してまだ間もなく、大変お忙しいとは思われたのだが、植木優夫先生に会うことができ、取り組むならぜひ植木先生と一緒にと考え「やらないか」と鎌倉先生を真似て、上手に誘導したところ奏功し、共著で進めることが決まった。誠に申し訳ない気持ちであったが、快く引き受けて下さり、順風満帆なスタートを切ることができた。

いささか冗談のような言い回しで前書きを始めたが、おおよそ状況はこの通りであったと回顧する。テーマはゲノムデータ解析と想定したため、筆者一人で全てを執筆するよりも、より最近のゲノムワイド関連解析 (GWAS) や次世代シーケンサによる超大量データ、それらに付随する統計的な問題などを熟知された植木先生と一緒に取り組めることはとても心強いことであった。実際、植木先生にお任せした範疇は、出来上がったものを見ていただくとわかるように、とても読み応えのあるものである。おかげで比較的最近の話題をいくつか簡単に紹介することもできたので、今回の話をお引き受けして取り組んで良かったとしみじみ感じている。

植木先生にご担当いただいた第 3 章は、GWAS を想定された関連研究を中心として、サンプル QC から始まる。ヘテロ接合度や近交係数・親縁係数も挙げられ、遺伝学の理論から説明していただいている。また、イ

ンピューテーションについても紹介していただくことを筆者からお願いしており、その理論的なアプローチから説明されているので、すでにインピューテーションを使ったことのある研究者にとっても改めて一読に値するだろう。ごく最近の話題であるレアバリエーション解析について、SKAT やその亜種の方法論まで言及いただいた。

筆者が担当したのはそれ以外の章である。第1章ではGWASを含めた、比較的最近に筆者が関わってきた遺伝統計解析のそれぞれを簡単に紹介させていただいた。データの検証において、マーカーQCに始まり、民族集団からの構造化問題、genome-wide significant level、マンハッタン・プロット、Haploviewなどを紹介した。統計的な問題として検出力やサンプルサイズ算出、false discovery rate (FDR)なども挙げている。第2章においては特にSNPデータ (SNPでなくともハプロタイプは推定できるが)の解析において、ハプロタイプや連鎖不平衡 (LD) にフォーカスをおいている。次世代シーケンサを用いた解析が進む現在ではすでにやや古いテーマかもしれないが、GWASや、そこからさらに細かく検討する際に、連鎖不平衡、連鎖不平衡係数、ハプロタイプ推定などを理解することは決して無駄な知識ではないと考えて紹介した。ややマニアックな領域にまで踏み込んでしまったかと思わないでもないが、満足できるレベルまで網羅できた。

このように前半と後半で別々に執筆しているため、文調は異なるが取って一致させず、それぞれのやり方を尊重した形で残した。それぞれのパートを著者同士で確認し合っているが、特に第3章は、東北大学 東北メディカル・メガバンク機構 ゲノム解析部門 リスク統計解析室長の田宮元教授にも原稿の査読に加えて、参考文献紹介、文章構成についての有用な助言をいただいた。また、久留米大学バイオ統計センター院生の中倉章祥氏には、第3章の細かい文章の誤りを指摘していただいた。さらに編集委員の閲読者の先生や共立出版編集部にも多くのご賢察・ご高配を賜った。この場をお借りして厚く感謝の意を表したい。

2016年7月

富田 誠