

まえがき

計算代数統計 (computational algebraic statistics) とは、代数学の理論を利用して統計学の諸問題に取り組んだ、比較的新しい研究分野をいう。もちろん、統計学には以前にも、代数学の理論にもとづくさまざまな研究があった。代表的なものには、統計的決定理論における変換群論とハール測度の応用や、実験計画法における有限体の理論、多変量分布論における連続群の表現論などが挙げられる。一方、計算代数統計は、代数幾何、可換代数、組合せ幾何などの理論が互いに関連して登場するという、多面的な特色をもつ。これらの理論を繋ぐキーワードのひとつが、グレブナー基底の理論である。

統計学 One Point シリーズの読者は、大学や研究所などの教育研究機関、あるいは、製造や流通などのさまざまな現場において、統計学に接している方が多いと思う。一方で、グレブナー基底などという耳慣れない単語をいきなり目にして、戸惑いを感じている方もいるだろう。グレブナー基底は純粋数学の概念のひとつであるが、比較的近年になってその応用分野が開拓され、純粋数学に限らない他領域の研究者からも注目されるようになった。最先端のトピックのひとつである。そこでまず、グレブナー基底が誕生してから統計学と結びつくまでの歴史的な背景を、本書でも随所で引用しているグレブナー道場 ([15]) の第 1 章の説明から、抜粋して概説する。

グレブナー基底の概念が最初に登場したのは 1960 年代であり、廣中平祐と Bruno Buchberger によりほぼ同時期に独立に導入された。グレブナー基底 (Gröbner basis) という用語は、Buchberger が自身の学位論文の指導教員である Wolfgang Gröbner に敬意を表して名付けたものである。グレブナー基底は、登場からおよそ 20 年後の 1980 年代の後半に、可換代数と代数幾何の計算ソフト Macaulay が開発されたことが転機とな

り、注目を集めた。これがグレブナー基底の理論展開の第一のブレイクスルーである。さらに90年代に入り、凸多面体の三角形分割の理論と可換代数におけるグレブナー基底の理論との繋がりが発見されたことが、第二のブレイクスルーといわれる。そして90年代に、後述する二つの論文によって統計学におけるグレブナー基底の応用例が発見され、計算代数統計とよばれる分野が誕生したことは、グレブナー基底の理論展開における第三のブレイクスルーと位置づけられる。

計算代数統計は、わずか20年の間に、統計学、代数学の両分野の研究者によって精力的に研究が進められ、現在も急速に進展している、分野融合的な特色をもつ研究分野である。その進展の背景には、グレブナー基底の研究と代数計算ソフトウェアの急速な発展がある。計算代数統計では、統計モデルを代数的に、つまり代数方程式系の零点の集合として特徴づけ、それを代数計算ソフトウェアで計算する（このことが、分野名に「計算」が付いている所以である）。ここでいう計算とは、例えば本書の1.6節で紹介するBuchbergerアルゴリズムによるグレブナー基底の計算などを指す。ところがこのBuchbergerアルゴリズムは、入力の変数の数に対して二重指数(!)の計算オーダーをもち、与えられた問題のグレブナー基底が現実的な時間で得られるかどうかは、問題のサイズに大きく依存する。実際、計算代数統計の発端となった上記の論文が出版された90年代の初めの頃は、応用統計学に現れるような、変数の数が数十を越えるような問題のグレブナー基底は、計算できないのがむしろ当たり前であったように思う。例えば、当時大学院生だった筆者は何度か、(データ解析として見れば)ごく小さいサイズの問題のグレブナー基底の計算を試みて、数時間から数日を要しても計算が終了しないという経験をした。当時の筆者が、計算代数の勉強を始めたばかりの駆け出しであったことは事実だが、それを差し引いても、その頃はまだこの分野には、「理論は面白くても実際には使えない」という側面があったように思う。しかしその後、グレブナー基底の理論と代数計算ソフトウェアの急速な発展とともに、計算できる問題のサイズは徐々に大きくなっていった。上述の「数日経っても終了しなかった計算」の中には、現在では、手元の標準的なノートパソコンで

1秒未満で終了するものもある。そのような問題であっても、さらにサイズをわずかに大きくしただけで、「計算できない問題」に戻ってしまうこともあるのだが（そしてその問題も、数年後には一瞬で計算が終了することになるのかもしれないが）、現在は、標準的な計算機があれば、勉強をしながら誰でも手軽に計算を実行し、出力を確認することができる状況にある、といっても間違いではないだろう。このことは、計算代数統計を学ぶ上での最大の魅力であると思う。

本書は、計算代数統計を初めて学ぶ学部生・大学院生を想定して、グレブナー基底の入門的な説明と、その理論が結びついた統計学の最初の問題のひとつを紹介する。前提とする数学の知識は、基本的な線形代数のみである。統計学の知識は、線形モデルの理論は理解していることが望ましいが、仮に知らなくても、本書の内容を理解するために最低限必要な説明は本書に与えた。本書は、統計学 One Point シリーズの一冊であるが、内容は統計学よりもやや純粋数学寄りかもしれない。しかしその分、定義と定理よりも、具体例と考え方の説明や、ソフトウェアによる実際の計算の紹介に紙面を割いた。筆者としては、本書を、理系全般の学生に広く読んでいただき、一人でも多くの方にこの分野の雰囲気を感じ取ってもらえれば嬉しい。

歴史的には、計算代数統計の起源となったのは、90年代の2本の論文、つまり Pistone と Wynn による実験計画法への応用の論文 ([24]) と、 Diaconis と Sturmfels による分割表解析への応用の論文 ([10]) である。これらのうち、前者の論文を理解することが、本書の目標である。本書を読んで計算代数統計に興味をもった読者は、是非、後者の話題についても興味を広げてほしい。グレブナー道場 ([15]) の第4章は、後者の話題についての筆者による入門的な説明である。また、筆者と共同研究者による初期の研究成果の解説記事 [3] や単行本 [4] は、この分野の研究を知るためのきっかけになると思う。

本書では、前述した「実際に計算する楽しさ」を読者に体験してもらうために、代数計算ソフトウェアのひとつ、Macaulay2による計算の例を、なるべくたくさん掲載した。名前から分かる通り、Macaulay2は、

グレブナー基底の理論展開の第一のブレイクスルーのきっかけとなった計算ソフト Macaulay の後継である。そのような歴史的な理由もあり、Macaulay2 は、世界中で最も使われている代表的な代数計算ソフトウェアのひとつである。とはいえ、現在では、Macaulay2 のほかにも数多くの代数計算ソフトウェアの選択肢があり、本来は、目的や計算内容に応じてさまざまなソフトウェアを使い分ける（使いこなす）のが、計算代数統計の勉強・研究の醍醐味である。それでも敢えて、本書で Macaulay2 という特定のソフトウェアを取り上げた理由は、この原稿の執筆時点で、ウェブブラウザ上で利用できる環境 (Macaulay2 online) が整備されている点にある。本書で紹介する計算例では、なるべく基本的な関数のみを使い、具体的な計算コードもすべて記述した。したがって、代数計算ソフトウェアの利用経験が全くない読者でも、ソフトウェアのインストールの手間をかけずに、手元の計算機やタブレット端末からブラウザ上で実際の計算を実行し、この分野の雰囲気を感じとることができるはずである。具体的な方法は 1.7 節を見てほしい。

本書の構成は以下の通りである。第 1 章のグレブナー基底入門では、グレブナー基底とはそもそもどのような性質をもち、どのようにして得られるのかを、多項式の割り算の視点から説明し、グレブナー基底の強力な応用例である消去定理を説明する。多項式環のグレブナー基底を導入するためには、多項式環の定義と Dickson の補題から始めて、イデアルの定義、単項式順序の導入を経るのが王道だと思われるが、本書は、抽象的な議論をなるべく避けて、高校数学などで馴染みのある具体例を通して理解を深めることを意図している。第 2 章では、グレブナー基底の理論が最初に統計学に応用された研究である、Pistone と Wynn の論文 [24] の内容を説明する。端的に言えば、この研究は、実験計画法における母数の推定可能性（識別可能性）の問題と、多項式環のイデアルに対するイデアル所属問題の関係を明らかにしたものである。これを理解するために必要となるのは、多項式環の剰余環の概念と Macaulay の定理である。本書ではこれらを、できるだけ平易に説明する。

本書の第 1 章の内容は、第 2 章を理解するために最低限必要なグレブ

ナー基底の理論であるが、これをもう少し具体的に書けば、「代数方程式系を一般的に解くための理論」となる。統計学に関連した代数方程式系の例をひとつ挙げれば、統計モデルの母数表現、あるいは陽的 (explicit) 表現から、陰的 (implicit) 表現を導くための一般的な方法は、代数方程式系を解き「ある変数を消去する」ことに対応する。例えば、2元分割表のセル確率を表す母数 $\{p_{ij}\}$ について、行と列との独立モデルは、通常は母数表現で

$$p_{ij} = \alpha_i \beta_j, \quad \forall i, j$$

などと表すことが多いが、これを代数方程式系

$$p_{ij} - \alpha_i \beta_j = 0, \quad \forall i, j$$

とみて $\{\alpha_i\}, \{\beta_j\}$ を消去すれば、陰的表現

$$p_{ij} p_{i'j'} - p_{i'j} p_{ij'} = 0, \quad \forall i, i', j, j'$$

が得られる (1.8 節で、実際の計算を確認する)。このようにして導かれた陰的表現の応用的な意味のひとつを明らかにしたのが、計算代数統計のもう一方の起源である、分割表解析への応用の論文 ([10]) である。このように、統計モデルや、本書で扱うような実験計画、あるいは、尤度方程式や構造方程式など、統計学に現れるさまざまな代数方程式は、それらが定義する代数多様体とみて、代数幾何学的な研究対象として扱うことができる。代数幾何学における基本的な概念には、例えば、次元、既約性、特異点、閉包などがあるが、統計学に現れるさまざまな代数多様体について、これらの概念の統計学的な意味については、明らかにされていない部分が多い。統計学においては基本的な対象であっても、代数幾何学的には、高次元で複雑な研究対象となるものが多いからである。計算代数統計は、グレブナー基底の理論を武器に、計算代数ソフトウェアを駆使して、統計学の諸問題の代数幾何を研究する学問である。筆者は、本書で扱う「代数方程式系を一般的に解くための理論」や Macaulay2 による計算が、計算代数統計のスタートラインに立つための「最初の武器」になると信じてい

る。これを磨き、より強力な武器を入手するための道標となるであろう文献をあとがきで紹介する。

グレブナー基底に関する定評のある教科書は多いが、計算代数統計への応用を念頭に書かれた教科書としては、前述したグレブナー道場 [15] が、初学者向けの最適な入門書のひとつであろう。本書で紹介する定理の証明のいくつかは、このグレブナー道場の第 1 章を参考に構成している。その他の参考書に関しては、あとがきに説明を加えた。

本書をまとめるにあたって、多くの方にご助力をいただいた。神戸大学大学院理学研究科の高山信毅先生、立教大学理学部の野呂正行先生、小市民雄先生には、草稿段階の本書を読んでいただき、有益なコメントをいただいた。特に、「この説明は、数学者にとっては分かりにくい」という指摘は貴重であり、筆者自身が学び直すきっかけとなった。編集委員会と読者の方々には、小さなミスから本質的なものまで、多くの丁寧なコメントをいただき、本書の内容をより良いものに引き上げていただいた。読者のお一人には、本書に掲載した Macaulay2 のコマンドを、一つひとつ実行してチェックしていただいた。共立出版編集部には、辛抱強く原稿の完成を待っていただき、編集では大変細やかな対応をいただいた。ここに感謝いたします。また、日頃から筆者の研究活動を支えてくれている妻と娘にも、この場を借りて厚く感謝の意を表します。

2018 年 6 月

青木 敏 (神戸大学)