● はしがき

データ分析のツールは日々進化しており、次々と新しい分析手法が経済学の実証研究に取り入れられています。特に最近の大きな変化は、機械学習の隆盛でしょう。機械学習の手法を用いた実証研究は、この数年で急速に増加しています。ところが、教育の側はそのような流行に追いついているかというと、どうでしょうか。特定の教員によって単発的に授業が行われることはあっても、計量経済学の標準的なトピックとして機械学習の手法が扱われるまでには至っていないのではないかと思います。

もちろん、最近では機械学習関連の書籍は日本語でも数多く出版されていますので、それらを読めば済むのかもしれません。しかし、私の知っている範囲では、それらの本は一部を除いては、経済学者の興味とは異なるところに重点を置いていたり、あるいは、基礎的過ぎて物足りなかったりと、計量経済学を学ぶ人たちにとって丁度良い塩梅のものは、依然として多くはない印象です。

本書はこのような昨今の状況を鑑みて書かれたものです。経済学をバックグラウンドとする人たちを主たるターゲットとして、従来の計量経済学の手法の拡張という視点から機械学習の手法を解説しています。ですが、経済学の知識は特段必要としません。

本書は７つの章から成り、大きく前半（１～４章）と後半（５～７章）に分けられます。前半は伝統的な計量経済学・統計学の手法に関するパートです。ただし、伝統的とは言っても、通常の授業では詳しく取り上げられない内容に焦点を当てており、最近の研究についても触れています。１章で本書の問題意

識を述べた後、2章で線形回帰モデルの変数選択、3章でノンパラメトリック回帰、4章でセミパラメトリック回帰について解説しています。2〜4章で扱う手法は、計量経済学や統計学で以前から用いられてきたデータ駆動型の分析手法です。これらはそれ自体で重要であるとともに、後半の内容の理解の助けにもなると考えて、トピックとして採用しました。

後半は機械学習の手法に関するパートです。機械学習が対象とする範囲は膨大で、私の理解が及ぶ範囲にも限りがありますので、教師あり学習と呼ばれるカテゴリーに属するものの一部のみを扱っています。教師あり学習の主たる目的は予測ですが、計量経済学では区間推定や検定などの統計的推測も重要ですので、機械学習の手法を基にした統計的推測についても解説しています。具体的には、5章で回帰木などのツリーベースの手法、6章で線形回帰モデルの正則化推定量を紹介し、7章ではそれらの手法を応用して、興味のあるパラメータの妥当な信頼区間を求める方法を考察します。

私がはじめて機械学習と接点を持ったのは2010年で、Ph.D.を取得した直後でした。その年に開催された情報論的学習理論ワークショップ（IBIS 2010）で、企画セッション「計量経済と機械学習」に登壇者として招待していただいたのがきっかけです。そのときに機械学習に対して抱いた印象は、自分の関心とかなり近いことをやっているというものでした。機械学習の主眼はデータ駆動的なモデリングにありますが、元々私は変数選択やノンパラメトリック・セミパラメトリックな分析手法の研究をしていたこともあり、データ駆動的という考え方はすでに馴染みのあるものだったからです。最近では機械学習の手法を自分の研究にも取り入れるようになりましたが、私にとっての機械学習はデータ駆動型の分析を行うためのいくつかある手段のうちのひとつであり、従来の計量経済学や統計学を無用にしてしまうようなものであるとは考えていません。とは言え、機械学習の有用性を過小評価しているわけでもありません。

「データ駆動型」と並ぶ本書のもうひとつのキーワードは、「一様性」です。統計分析の手法が満たすべき望ましい性質のひとつとして、分析者が想定するどのような分布からデータが発生したとしても、その手法がうまく機能するというものがあります。そのような性質は一様妥当性（uniformly validity）と言われ、近年の計量経済学の理論研究では重要視されている性質なのですが、大

抵の教科書では触れられていないのではないかと思います。理論の詳細までは踏み込みませんが、雰囲気だけでも伝わればと思い、本書のいくつかの箇所でこの問題を取り上げています。

本書では学部上級レベルの計量経済学の知識を前提としています。『計量経済学』（西山慶彦他、有斐閣、2019 年）くらいを想定していますが、それにプラスして、線形代数の知識も多少必要です。一部ではそれだけでは理解が困難かもしれない発展的な内容も扱っていますが、それらは主として理論研究に関心がある人やプロの研究者向けに書いています。星印（＊）の付いている節は難易度が高めで、星印が 2 つ（＊＊）付いている節はかなり難しいです（星印が付いていない部分が決して易しいわけでもないですが)。星印の有無にかかわらず、難しそうな部分は適宜読み飛ばしてもらってもよいでしょう。

本書のスタイルは、基本的には拙著『計量経済学：ミクロデータ分析へのいざない』（日本評論社、2015 年）のそれを踏襲しています。証明はほぼ省略し、興味のある人には参考文献を挙げるにとどめています。その理由は、話の全体像が見えづらくなることを避けるためであり、厳密性を犠牲にしているというわけではありません。証明をしない代わりに、説明は丁寧にしています。新しく登場する事柄については、可能な限り数式と言葉の両方を用いて解説をするようにしています。また、どのような分析手法であれ、特定の目的には有用であっても、その他の目的にとっては有害にもなりうるので、分析手法の優れた点だけではなく、問題となりうる点についても極力言及するようにしています。

本書の執筆にあたり、多くの方々にお世話になりました。東京大学の奥井亮氏と坂口翔政氏、慶應義塾大学の岡達志氏、滋賀大学の新久章氏、ニューヨーク大学博士課程の池上慧氏、カリフォルニア大学サンディエゴ校博士課程の林田光平氏には、本書の草稿に目を通していただき、多くの有益なコメントをいただきました。また、本書は私のサバティカル中に執筆されたものです。サバティカル中に一緒に仕事をさせていただいた株式会社サイバーエージェント AI Lab の安井翔太氏、竹浪良寛氏、松木一永氏からも有益なコメントをいただきました。紙幅の都合上、あるいは、私の能力の問題で、いただいたコメントのすべてを最終稿に反映させることは叶いませんでしたが、貴重な時間を割

いて原稿を読んでいただいた皆様に心から感謝申し上げます。

また、前著に続いて執筆の機会を与えてくださった日本評論社の吉田素規氏には、今回も大変お世話になりました。原稿の数々の間違いを吉田氏に発見していただかなければ、大変なことになっていたでしょう。もちろん、まだありうる本書の間違いについては、すべて私の責任です。

回帰分析と言えば、佐和隆光先生の『回帰分析』（朝倉書店）が有名です。初版が出版された 1979 年は、私が生まれた年でもあります。それから 45 年が経過し、見識は佐和先生にはるかに及ばないものの、私なりに現代の回帰分析のあり方を模索し苦闘した跡を、本書に読み取っていただければ幸いです。

2024 年 3 月

末石直也

●目　次