

数学とも実データとも違う統計の入り口を探る：

科学哲学・帰納推論・集団的思考

島谷健一郎・統計数理研究所

東京都立川市緑町 10-3

050-5533-8590

shimatan@ism.ac.jp

統計学の入り口

大学レベルの統計学において、その入り口としてよく用いられているものは、以下の 2 つであろう。

1. 数学

古くから数学教員による統計授業として広く行われている。学生からの「数学が難しい」という声以上に問題なのは、そこで教わる統計がいざ実データを手にした時どのように生きてくるか、数学の学習時には想像できないこと、また、実データを手にした時、数学から入った統計を手元の実データに適用した後に結果をどう解釈し何を新たに考察しないといけないか、数学の授業の中では教えられない、といったあたりである。

2. 実データ

4 回生や大学院あるいは実社会の現場で実データを手にし、改めて統計を学び直している人も多い。実際の問題・課題と実データから入る場合、比較的多く見かけるのは、統計手法の計算ソフトの使い方を覚え、数値結果を出してしまう所から入るというものである。それから数学的根拠を踏まえた数値結果の解釈と考察に移行すればいいのだが、しばしば、数値結果が出たところで何かわかった気分、マスターした錯覚、既に克服した達成感に見舞われ、思考停止に陥る。

現在統計を学ぶ学生の大半は、数学や情報以外を専攻とし、数学は苦手科目である。そうした学生に入りやすい別な入り口の候補に、科学哲学がある。

科学哲学から入る自然科学や数学の一例として、科学史から学ぶ物理学が挙げられる。物理学はし

ばしば計算中心になる。それは物理学専攻や工学系の一部には適切かつ必須だが、物理学の計算を必要としない専攻も多い。計算中心の物理学では、学生はついていけないし、物理学周辺の有益な素養を学ぶ機会を奪う結果も招く。そこで例えば、天動説と地動説を巡ってどんな論争が繰り広げられたかなどの歴史から入ることで、科学における仮説と、その検証、立証、反証といった、科学的推論を学ぶという物理教育が提唱されている。

科学哲学の視点からの統計学再入門

同じように、統計学にも、科学哲学からの入り口があるはずである。発表者は、森元良太（北海道医療大）や松王政浩（北海道大）らと、科研費「科学教育の一環としての統計学認識論に関する教育カリキュラム構築」（基盤 B、代表：松王政浩、16H03050）で、これを模索してきた。そして、「科学哲学の視点からの統計学再入門」を提唱し、これを 2018 年度、統計数理研究所における公開講座として開講（森元良太講師）した。

本発表者は、この中の帰納推論と集団的思考を 2 つのキーワードとし、講座の中の該当部分を学部 2 回生及び大学院生（主に修士課程）の授業で実践した。

そもそも、科学においてどうして統計学が必要なのか。統計思考力を有しないとどういう場面で困るのか。そうした概念的な問題を入り口とするものである。これは、「再入門」でなく「入門」である学部生にとっても有効な入り口と思われる。

帰納推論

統計を伴う推論で、何とはなしに歯切れの悪さや違和感を感じた人は多いはずである。その要因のひとつが、帰納推論にある。

データや観察結果から帰納的に仮説を検証する推論は、その根本的問題を広く認識されながらも科学で広く使われてきている。ところが、そこに統計が加わると、格段に歯切れが悪くなる。その理由を科学哲学研究者と議論していくと、単純な原因に到達した。統計の中の数学は演繹推論である。統計を用いると、仮説検証の過程に、演繹と帰納が混在する。かつ、統計解析の中でも、データからの推定など帰納的な逆問題と、演繹的な **forward simulation** が混在する。これが違和感や歯切れの悪さの一因である。

これを逆に見ると、本質的には帰納推論の中に演繹推論が混在する統計を意識することで、違和感や歯切れの悪さは軽減される。帰納と演繹を認識することで、推論の進め方も円滑にできるはずである。

集団的思考と誤差論

統計教育現場において、しばしば用いられる方法に、同じ物を複数の生徒で測り、その計測値のばらつきを見るというものがある。多くの場合、計測値は真値の周辺に正規分布状に散らばる。格好の正規分布入門コースである。

ところで、ばらつきの中には、計測値のように真値がまずあり、ばらつきは計測の悪さの尺度のような世界と別に、人を含む生物の形質などに多く見られる、集団内の変異という世界がある。そこには、真値もなければ、最良値もなく、また、数値がばらついていることはそのまま集団の変異の大きさという集団自体の特性を表す。

そして、統計のひとつの目的は、変異を有する集団の特性を、データから推定することにある。統計においても真値という概念はあるが、それは集団の特性である平均や分散の真値であり、集団自体に真値や最良値があるわけではない。

ガウスが最初に正規分布に到達したとき、そこには真値があり、分散は計測の不完全さの尺度だった。その後、ダーウィンの進化学においては、ばらつくことは生物集団の本質であり、ばらつき

があるから自然選択と浮動により進化が起こる。同じように正規分布を用いていても、背景にある考え方は、誤差論と集団的思考で根本的に異なる。

この2点は、学部生の早い時期でも十分に吸収できるし、おそらく高校や中学の段階でも可能と思われる。本発表では、このような統計教育実践と、その背景にある理念について発表する。