

〈ポスト p 値時代〉の統計学

— t 検定、 F 検定、 χ^2 乗検定相当の分析をどうするか —

企画：秋山 隆（早稲田大学）
 話題提供：豊田秀樹（早稲田大学）
 指定討論：向後千春（早稲田大学）

司会：久保沙織（早稲田大学）
 指定討論：楠見 孝（京都大学）

キーワード：ベイズ統計学, ポスト p 値時代, 統計教育

統計的方法を学ぶことは、これまで、すなわち有意性検定を学ぶことでした。長期に渡りこの大前提はゆるぎなく磐石で、無条件に当たり前で、無意識的でした。しかし、ときは移り、有意性検定や p 値の時代的使命は終わりました。アメリカ統計学会 ASA は、2016年3月7日に、 p 値の誤解や誤用に対処する6つの原則に関する声明をだしました (Wasserstein & Lazar, 2016)。この声明は『『ポスト $p < 0.05$ 時代』へ向けて研究方法の舵を切らせることを意図している』(R. Wasserstein (ASA News Releases, 2016)) ものだと言明されています。2016年現在、統計学における著名な学術雑誌バイオメトリカ (Biometrika) の過半数の論文が、ベイズ統計学を利用しています。多くの著名な学術雑誌も同様の傾向です。スパムメールをゴミ箱に捨て、日々、私たちの勉強・仕事を助けてくれるのは、ベイズ統計学を利用したメールフィルタです。ベイズの画像処理によってデジタルリマスターされ、劇的に美しくよみがえった名作映画を私たちは日常的に楽しんでます。ベイズ理論が様々な分野で爆発的に活用されています。ベイズ的アプローチなしには、もう統計学は語れません。

有意性検定にはどこに問題があったのでしょうか。3点あげます。

I. p 値とは「帰無仮説が正しいと仮定したときに、手元のデータから計算した検定統計量が、今以上に甚だしい値をとる確率」です。この確率が小さい場合に「帰無仮説が正しくかつ確率的に起きにくいことが起きたと考えるのではなく、帰無仮説は間違っていた」と判定します。これが帰無仮説の棄却です。しかし帰無仮説は、偽であることが初めから明白です。それを無理に真と仮定することによって、検定の論理は複雑で抽象的になります。例えば2群の平均値の差の検定におけ

る帰無仮説は「2群の母平均が等しい ($\mu_1 = \mu_2$)」というものです。しかし異なる2つの群の母平均が、小数点以下を正確に評価して、それでもなお等しいということは科学的にありえません。帰無仮説は偽であることが出発点から明らかであり、これから検討しようとするのが既に明らかであるような論理構成は自然な思考にはなじみません。 p 値は土台ありえないことを前提として導いた確率なので、確率なのに抽象的で実感が持てません。このことが p 値の一番の弊害です。以上の諸事情を引きずり、「有意にならないからといって、差がないとは積極的にいえない」とか「有意になっても、 n が大きい場合には意味のある差とは限らない」とか、いろいろな言い訳をしながら有意性検定をこれまで使用してきたのです。しかし、これらの問題点はベイズ的アプローチによって完全に解消されます。ベイズ的アプローチでは研究仮説が正しい確率を直接計算するからです。

II. n を増加させると p 値は平均的にいくらでも0に近づきます。これはたいへん奇妙な性質です。 n の増加にともなって、いずれは「棄却」という結果になることが、データを取る前に分かっているからです。有意性検定とは「帰無仮説が偽であるという結論の下で、棄却だったら n が大きかった、採択だったら n が小さかったということを判定する方法」と言い換えることすらできます。ナンセンスなのです。これでは何のために分析しているのか分かりません。 n を増加させると、 p 値は平均的にいくらでも0に近づくのですから、BIG データに対しては、あらゆる意味で有意性検定は無効です。どのデータを分析しても「高度に有意」という無情報な判定を返すのみです。そこで有意性検定では n の制限をします。これを検定力分析の事前の分析といいます。事前の分析では有意になる確率と学術的な対象の性質から逆算し

て n を決めます。しかし検定力分析によるサンプルサイズ n の制限・設計は纏足¹と同じです。統計手法は、本来、データを分析するための手段ですから、たくさんのデータを歓迎すべきです。有意性検定の制度を守るために、それに合わせて n を制限・設計することは本末転倒です。ベイズ推論では n が大きすぎるなどという事態は決して生じません。

Ⅲ. 伝統的な統計学における平均値の差・分散の比・クロス表の適合などの初等的な統計量の標本分布を導くためには、理系学部で 2 年生程度の解析学の知識が必要になります。すこし複雑な統計量の標本分布を導くためには、統計学のために発達させた分布論という特別な数学が必要になります。それでも、どの統計量の標本分布でも求められるという訳ではなく、導出はとても複雑です。検定統計量の標本分布を導けないと、(教わる側にとっては) 統計学が暗記科目になってしまいます。この検定統計量の確率分布は何々で、あちらの検定統計量の確率分布は何々で、のように、まるで歴史の年号のように、いろいろと覚えておかないと使えません。暗記科目なので、自分で工夫するという姿勢が育つはずもなく、紋切り型の形式的な使用に墮す傾向が生じます。でもベイズ統計学は違います。マルコフ連鎖モンテカルロ (MCMC) 法の本質は、数学 II までの微積分の知識で完全に理解することが可能です。標本分布の理論が必要とする数学と比較すると、それは極めて初等的です。生成量を定義すれば、直ちに事後分布が求まり、統計的推測が可能になります。文科系の心理学者にとっても、統計学は暗記科目ではなくなります。

学問の進歩を木の成長にたとえるならば、平行に成長した幾つかの枝は 1 本を残して冷酷に枯れ落ちる運命にあります。枯れ果て地面に落ちた定理・理論・知識は肥やしとなり、時代的使命を終えます。選ばれた 1 本の枝が幹になり、その学問は再構築されます。教授法が研究され、若い世代は勞せず易々と古い世代を超えていく。そうでなくてははいけません。統計学におけるベイズ的アプ

ローチは、当初、高度なモデリング領域において急成長しました。有意性検定では、まったく太刀打ちできない領域だったからです。議論の余地なくベイズ的アプローチは勢力を拡大し、今やその地位はゆるぎない太い枝となりました。

しかし統計学の初歩の領域では少々事情が異なっています。有意性検定による手続き化が完成しており、いろいろと問題はあるけれども、ツールとして使えないわけではありません。なにより、現在、社会で活躍している人材は、教える側も含めて例外なく有意性検定と頻度論で統計教育を受けています。この世代のスイッチングコストは無視できないほどに大きいのです。このままでは有意性検定と頻度論から入門し、ベイズモデリングを中級から学ぶというねじれた統計教育が標準となりかねません。それでは若い世代が無駄な学習努力を強いられることとなります。教科教育学とか教授学習法と呼ばれるメタ学問の使命は、不必要な枝が自然に枯れ落ちるのを待つのではなく、枝ぶりを整え、適切な枝打ちをすることにあります。ではどうしたらいいのでしょうか。どのみち枝打ちをするのなら、R. A. フィッシャー卿の手による偉大な「研究者のための統計的方法」にまで戻るべきです。「研究者のための統計的方法」の範囲とは、「データの記述」「正規分布の推測」「独立した 2 群の差の推測」「対応ある 2 群の差の推測」「実験計画法」「比率・クロス表の推測」です。これが統計学の入門的教材の初等的定番です。

文 献

- Wasserstein, R. L. & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose, *The American Statistician*, DOI:10.1080/00031305.2016.1154108
- ASA News Releases (2016). American Statistical Association releases statement on statistical significance and p-Values. (<http://www.amstat.org/newsroom/pressreleases/P-ValueStatement.pdf>)
- R. A. フィッシャー(著) 遠藤健児・鍋谷清治(訳) (1970). 研究者のための統計的方法 森北出版 (Fisher, R. A. (1925). *Statistical Methods for Research Workers*, Oliver and Boyd: Edinburgh.)

¹ 纏足(てんそく)とは、幼児期から足に布を巻き、足が大きくならないようにして小さい靴を履けるようにした、かつて女性に対して行われていた非人道的風習です。靴は、本来、足を保護するための手段ですから、大きくなった足のサイズに靴を合わせるべきです。靴に合わせて足のサイズを制限・整形することは本末転倒であり、愚かな行為です。他の靴を履けばよいのです。