

20世紀が終わる頃まで注目されなかった異端の統計学：ベイズ統計学

食料領域 主任研究官 鈴木 均

1. はじめに

本稿では、人工知能から暗号解読、画像解析、ビッグデータ解析などで応用されているベイズ統計学について、専門的な説明になるべくならないようにご紹介してみたいと思います。

まずは、統計学というのは、大まかに言って、記述統計学、そして推測統計学という2つの統計学に分類されます。記述統計学というのは、観測されたデータを複数集めてきて、その平均を計算するか、分散（データの散らばり具合のこと）を計算するとかして、データを集めてきた元となるデータの集団、データの塊の特徴を表すことになります。データの塊の中心はどのあたりなのか、データがどの程度散らばっているか集中しているかといったことを表現するものです。

一方、推測統計学というのは、記述するだけでなく推測もするものだ、と言ってしまえばそのままなんです。ざっくり言いますと、観測されたデータから、そのデータの背後に潜む規則や構造（データの塊の特徴など）を確率という道具を使って推測するものです。

2. 否定されたベイズ統計学

さらに、推測統計学が、大雑把に2つの考え方に分かれており、頻度主義統計学、そしてベイズ統計学という2つに分類されています。頻度主義統計学はデータの頻度に基づいて推測を行う統計学になります。さらに、歴史的な経緯をみると、20世紀の統計学の主流派は、頻度主義統計学でした。

現に、現代統計学の父と言われた有名な統計学者であるロナルド・A・フィッシャーという人は、頻度主義統計学の基礎を築いた人で、頻度主義統計学の考えを強く支持し、その発展に大きく貢献しました。ベイズ統計学のことを逆確率の理論と称して「逆確率の理論はある誤謬の上に立脚するものであって、完全に葬り去らなければならない」と宣言したそうです。こうしてベイズ統計学は20世紀までは、頻度主義統計学者たちに否定され、タブー視されていました。ベイズ統計学の元となるベイズの定理は250年も前に発見されたのですが、長い間使われていませんでした。

この傍流にあり注目されていなかったというベイズ統計学ですが、20世紀も終わりに近くなった頃に注目され始めたのです。1996年にマイクロソ

フトの元代表のビル・ゲイツ氏が「自社が競争優位に立っているのはベイズ統計によるものだ」と言及し、その後の2001年の基調講演において「21世紀のマイクロソフトの戦略はベイズ統計であること」や「世界中からベイズ統計の研究者をヘッドハンティングしたこと」を公言したのだそうです。こうして、ベイズ統計が注目されるようになり、たとえば、迷惑メールかどうかの判定、ネット検索での自動翻訳システムや、ネット購入をする顧客の購買行動などを推定など様々なところで利活用されています。特に最近では、深層学習（ディープラーニング）の中に組み入れることによって、より精度の高いAIによる予測が実現できるようになってきています。

3. ベイズ統計学とは

このベイズ統計学におけるベイズ推定という推定方法を説明していきます。ベイズ推定とは、以下の（1）式でよく表されるベイズの定理というものを使った統計的推定方法だと言われています。

$$P(A|R) = \frac{P(R|A)P(A)}{P(R)} \dots\dots (1)$$

まずは、ベイズの定理は、上記のような条件付き確率で表すことができる式なのですが、実は「この式ならよくわかるよ」、「この確率の等式が成り立つのは理解している」、「ベイズの定理は、このような条件付き確率のことだというのは見たことある」という人はいると思います。

ベイズの定理から、ベイズ推定の考え方にもっていくために、この上記の（1）式を以下の（2）式に捉えなおすことにします。

$$P(A|R) = \frac{P(R|A)P(A)}{P(R)} \dots\dots (2)$$

Aは知りたい情報パラメータのことで、Rは得られたデータ（観測したデータ）のことです。p(R)は定数と考えます。すると、（2）式は、事後確率は、尤度と事前確率の積を定数p(R)で割ったものであるという式になっています。

ここで、「事前確率」、「尤度」、「事後確率」といった難しい用語が出てきましたので、少しわかり

やすくするために別の表現に置き換えて説明してみます。「事前確率」というのは、実際に知りたい情報（パラメータA）ではあるのだが、直接観測できないので、まずは、今までの知識や情報（事前の知識）から考えて、どうなりそうなのか、というものです。「尤度」については、今起こった結果（データR）が、どれくらい起こりやすいかというものです。「事後確率」というのは、新しい情報（データR）を追加することによって、事前に考えていたどうなりそうか（パラメータA）を更新したあとの確率のことです。

そうすると、(2) 式をさらに言い換えますと、事前確率（事前に知られている知識や仮定に基づく確率）と尤度を利用することで、事後確率（未知の新たなパラメータの確率）がわかるということになります。すなわち、この事前にわかる確率と新しいデータ（尤度）を組み合わせることで、より正確な推測（新たなパラメータの確率）を導き出すことができますということなのです。

ここで、ベイズ推定をより理解するために、それぞれ確率として捉えていたものを確率分布として考えるようにします。すると、(2) 式は、以下の(3) 式となります。

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \dots\dots (3)$$

この(3) 式は、事後分布は、事前分布と尤度から求められることを示しています。これは、どういうことかということ、知りたかった新たなパラメータθの確率分布は、事前のパラメータθの確率分布（事前分布）と尤度によって、計算できるということなのです。

シンプルに言うと、ベイズ推定というのは、「事前確率分布と尤度関数から、事後確率分布を求める」ことだとも言えます。

4. ベイズ統計学の復権のきっかけ

ここで、今日のベイズ統計が認識され広まるきっかけの一つとなったMCMC法（マルコフ連鎖モンテカルロ法（Markov chain Monte-Carlo methods））について簡単に紹介します。MCMC法は、難しい積分計算を行って確率分布を求める代わりに、コンピュータ・シミュレーションを使って（ランダムサンプリングを行って）確率分布を推定する方法です。

ベイズ推定においては、上記の事後分布を求めることができたのですが、この事後分布による知りたい推定量を算定する時には積分計算が含まれており、この積分計算が解析的に求まらない場合がほとんどでした。MCMC法では、この時に、難しい積

分計算をマルコフ連鎖に置き換えることで、事後確率分布を求める計算を可能にしました。この計算が可能となった背景には、コンピュータの発達によって計算能力の向上したことで、複雑な計算を短時間で行うことが可能となったことがあげられます。MCMC法が1990年代から急速にベイズ推定に展開されました。このように、MCMC法がベイズ統計学に大きな変化を与えたのではないかと思います。

5. まとめ：ベイズ統計学の特徴とその応用

ベイズ統計学の特徴を述べると、頻度主義統計学が、推定値を点で捉えるのに対して、ベイズ統計は分布で捉えて表現していると言えます。すなわち、ゆらぎをもったものとして表現しています。知りたいパラメータを一つの点で推定するのではなく、知りたいパラメータを確率分布として扱うのです。知りたいパラメータは直接観測できないが、観測されたデータからその背景にある知りたいパラメータを推計することができるのがベイズ統計学です。

ベイズ統計は、21世紀に入り見直されてきており、医療、金融、製造業など様々な分野で応用され注目されています。機械学習のアルゴリズムにおいて、ベイズ推定はモデルのパラメータを推定するために使用され、事前分布と事後分布を利用して新しいデータに基づいてモデルを更新しています。また、ディープラーニング（深層学習）においても、ベイズの理論は不確実性の評価やモデルの正則化に利用されており、ベイズ的なアプローチを用いることで、過学習を防ぎ、より精度の高い予測を行うことができます。今流行の生成AIにおいても、ベイズ推定は生成AIの性能を向上させるためのツールであり、生成AIのより高精度のデータ生成や予測を行うことに役立っています。

ベイズ統計学はAI、機械学習、深層学習の分野で今後も重要な役割を果たし続けるでしょうし、応用される分野もその範囲は今後も広がるものと思われます。

【参考文献】

- 岩波データサイエンス刊行委員会（編）（2015）「岩波データサイエンスVol.1」岩波書店。
- 大塚淳（2020）「統計学を哲学する」名古屋大学出版会。
- 久保拓弥（2012）「データ解析のための統計モデリング入門」岩波書店。
- シャロン・バーチュ・マグレイン、富永星 訳（2013）「異端の統計学ベイズ」草思社。
- 須山敦志、杉山将（監修）（2017）「機械学習スタートアップシリーズ ベイズ推論による機械学習入門」講談社。
- 須山敦志（2019）「ベイズ深層学習」講談社。
- 豊田秀樹（2015）「基礎からのベイズ統計学」朝倉書店。
- 松原望（2024）「改訂版 入門ベイズ統計」東京図書。
- 松原望（2018）「ベイズの誓いーベイズ統計学はAIの夢を見る」聖学院大学出版会。