ベイズ統計学による心理学研究のすゝめ

岡田 謙介

2016年4月18日発行（Ver. 1.0）●発行元：ちとせプレス

いま，心理学の中でベイズ統計学を用いた研究が増えています。長年「傍流」とされてきたベイズ統計学が，最近になってなぜ見直されてきたのでしょうか。また，ベイズ統計学とはいったいどういった特徴をもつのでしょうか。「ベイズ統計学って聞いたことはあるけれど……」，そんな人に向けた，専修大学の岡田謙介准教授による「ベイズ統計学による心理学研究のすゝめ」。

Section 1

あれ，家を出る前にオーブンの電源スイッチ，切ったかな？

お気に入りのレーズンベーグルとチーズベーグル。この原稿を書いている今日も食べました。

私は現在，米国カリフォルニア州で在外研究をしています。近くのスーパーマーケットで安くて美味しいベーグルが買えるので，朝食に焼いて食べるのが好きです。つい先日の朝，職場に着いてから，ベーグルを焼いた後にちゃんとオーブンの電源を切ったかどうか，ふと心配になりました。うちのオーブンはアパートに作りつけの大型のタイプで，ダイヤル型のスイッチを回して切らない限り，いつまでも電源が入りっぱなしになってしまうのです。

こうしたことで気をもんだ経験があるのは，おそらく私だけではない……と思うのですが，いかがでしょう。家を出るときに玄関の鍵を閉めたかどうか，エアコンの電源を切ったかどうか，ふとしたはずみに気になって，思い出そうとしてもよく覚えていなくてモヤモヤすること。そして，実際には習慣化された行動として無意識のうちに鍵を閉めたり電源を切ったりしており，杞憂に終わることが多いこと。もしもこれが慢性化すると，強迫とよばれるような臨床的な症状に数えられてしまうかもしれません。ですが，頻繁にではないにしてもこれに類する経験が思い当たる方は，少なからずいらっしゃるのではないでしょうか……？

今回も，結局のところオーブンの電源スイッチは切ってありました。このヒヤリとした朝を反省しながら，私はふと，こうした思考はもしかすると，心理学でいわれている基準率無視の認知バイアスの観点からとらえることもできるかもしれない，と思いました。この認知バイアスは，比率や確率についての一般的な情報と，より個別的・特定的な情報が与えられたとき，多くの人が合理的な解よりももっと前者の情報を軽視し，後者を重視しがちになってしまうことを指します。2002年にノーベル経済学賞を受賞したダニエル・カーネマンと，エイモス・トヴェルスキーによる研究でよく知られており，心理学を学んだ方は「タクシー問題」や「感染者問題」といった代表的なカバーストーリーを耳にしたことがあると思います。近年でも日本発の研究の貢献も多い，面白い分野だと感じています。

基準率無視は，「確率計算によって求められる合理的な解」と「人間による直観的な認知」とが食い違ってしまう，という意味で，認知バイアスの1つであると考えられています。進化の過程で獲得されてきた人

間の直観的な認知は，私たちにとって有利に機能することが多くありますが，一方で合理的な答えからずれていることも少なくありません。このずれが系統的に観察される場合，それは認知バイアスと呼ばれます。

もっとも，今回私が注目したいのは，認知バイアスについてではなく，むしろ「合理的な」解を与えるための確率計算についてです。基準率無視の文脈における合理的な「正しい」解は，ベイズの定理を用いた確率計算によって求められます。ベイズの定理は，条件付き確率に関する，基本的な数学の定理です。そして，ベイズ統計学という名は，このベイズの定理の原型を発見した，18 世紀イギリスの牧師にしてアマチュア数学者，トーマス・ベイズに由来します。ベイズの定理について，少し歴史を紐解いてみることにしましょう。

ベイズ（Thomas Bayes, 1701-1761）の肖像画としてこの画像が使われることが多いのですが，実際にはおそらくこれは別人の画であろうと考えられています。残念ながら，たしかなベイズ本人の肖像画は現代に伝わっていません。[(1)]

ベイズの定理の発見

ベイズは，18 世紀の主に前半に生きたイギリスの，数学好きの牧師さんでした。もっともこれはイギリス国教会がそれ以外の教派への弾圧を強めていた時代であり，ベイズの属する長老派教会の信者の大半は大学の公職に就くことができませんでした。そのため当時，非国教徒の数学者や科学者は，牧師などで生計を立てつつ趣味として研究を続けることも少なくなかったのだそうです。統計学の中で非主流派の時代が続いてきたベイズ統計学の系譜が，やはり非主流派の宗教家に由来することは面白い偶然ですね。

彼が考えたのは，試行の回数と，ある結果が生じた（生じなかった）回数がわかっているときに，将来の試行においてその結果が生じる確率を推論する問題です。例えば，当たりつきのお菓子をいくつか購入し，そのうち当たりだったものの個数がわかっている状況において，次にお菓子を買って当たりが出る確率を推論するような場合です。

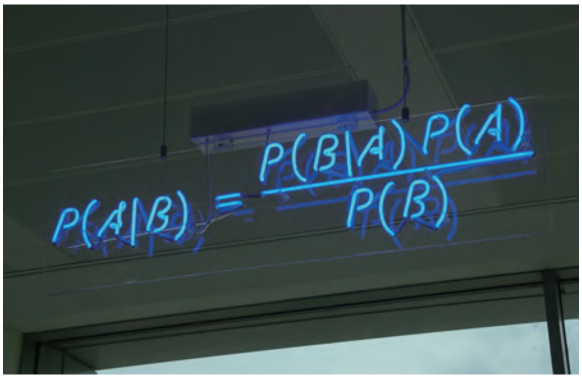
ベイズは大学の職業研究者ではなかったので，発見した定理について書き残しはしたものの，特にそれを対外的に発表はしませんでした。ベイズの死後，その数学関係の遺稿を整理していたリチャード・プライス[(2)] がこの発見の重要性に気づき，彼の手により 20 年近くを経てやっと論文が公刊されました[(3)]。もっとも，このプライスが世に出したベイズの論文には，あまり現実場面での具体的な例は書かれていません。したがってベイズ自身がどのような問題意識でベイズの定理を着想したのか，その詳細はわからないのですが，デイヴィッド・ヒューム[(4)] の提起した懐疑論により科学が因果関係をどのように扱えるのかに関心が集まっていた，という時代背景が背後にあると考えられています。

もっとも，ベイズの論文が出版された当時，残念ながら研究者の間ではこの論文の重要性はほとんど顧みられることはなかったようです。ベイズの発見が広く知られるようになったのは，後に数学をはじめさまざまな分野で業績を挙げた高名な研究者ピエール゠シモン・ラプラス[(5)] が独力でベイズの定理を再発見し，さらに体系化して発表してからです。ラプラスはベイズの定理を再発見しただけでなく，当時流行していた賭けトランプの問題から惑星の運行に関する問題まで，さまざまな不確かさを伴う推論の問題にベイズの定理を用いた推論が利用できることを示し，いわばベイズの定理を実用化しました。これが，ベイズの定理を用いて，不確実性を伴う現実の問題についての推論に取り組む，ベイズ統計学が最初に確立したときだと見ることができると思います。

ベイズの定理とベイズ統計学

ベイズの定理は，確率の更新についての定理です。データ（情報）を得る前から得た後で，不確実性のある事象や変数についての確率がどのように更新されるのかについて，ベイズの定理はその答えを与えてくれます。データを得る前の確率を事前確率，データを得た後でベイズの定理によって更新された確率を事後確率とそれぞれよびます。

冒頭で挙げたオーブンの例は，ベイズの定理を説明するうえで適切だと胸を張って言えるものではありませ



英国ケンブリッジの Autonomy 社にある，ネオンサインになったベイズの定理。[6]

せんが，私の冷や汗経験の反省かたがた，この例を引き続き使わせていただくことにしましょう。このとき，私が新しく入手した（……というか，入手してしまった）情報は，「私は今朝オーブンの電源スイッチを切ったか，切らなかったか」という不確実性のある変数について，「今朝は切らなかったのではないか」と考えたことです。これが今回のデータに対応します。一方，私は日々の生活の中で，使い終わったオーブンに特別な注意を払うことなく電源スイッチを切っています。これはデータを得る前に私がもっている，事前確率に対応します。そして，今回のデータを得たことによって，オーブンの電源を実際に切らなかった確率が合理的にはどのように更新されるのかを，ベイズの定理は教えてくれるわけです。

使い終わったオーブンのスイッチを切らないという事前確率がそもそも小さいため，「今朝は」切らなかったかも，というデータを得た後でも，スイッチを実際に切らなかった合理的な事後確率は依然としてそう大きくはなりません。それにもかかわらず，基準率無視の認知バイアスをもつ人間（私です）は，事前確率を軽視し，今朝はスイッチを切り忘れたかもしれないと必要以上に心配になってしまうのです。

ところで「今朝オーブンの電源スイッチを切らなかった確率」は，「オーブンの電源スイッチを切らなかった確率」に対して，それが「今朝」であるという条件を付加したものです。このように，特定の条件が付け加えられたもとでの確率のことを，条件付き確率といいます。そして，条件付き確率をその意味を考えて自然な形で数学的に定義すると，ベイズの定理はそこから容易に導くこと，つまり証明することができます。

新しいデータが得られたときに，事前確率を事後確率にどう更新したらよいのか。その数学的で合理的な答えは，ベイズの定理を使うことなのです。そこで，このベイズの定理を使って，新しいデータ（情報）が来るごとに，不確実な事柄についての確率を更新していこう。不確実性を伴うどんな問題に対しても，このようにベイズの定理を用いてデータから確率を更新していこう。これが，ベイズ統計学の考え方です。

ベイズ統計学の 2 大原則を紹介しましょう。それは，①現実問題の「不確実さ」を確率で表現すること，②データに基づきベイズの定理によって確率を更新することです。確率は，0 から 1 の間の値をとり，独立事象であれば足し算ができるなど，数学的な取り扱いの仕方がくわしくわかっています。私たちが日々向き合う，さまざまな世の中の不確実さを定量的に扱うにあたって，確率を使うことは最も自然な方法といえるでしょう。また，ベイズの定理は容易に証明できる，新たな情報に基づく確率の更新の仕方についての定理です。しがって，このベイズ統計学の 2 大原則は，おそらく多くの方に納得してもらえる，自然な考え方だということができると思います。

ベイズ統計学はほぼ 20 世紀を通して，主流というよりはむしろ傍流，場合によっては異端の扱いでした。中には，あまりよく知らないけれど，ベイズ統計学は主観的で，あまり科学的とはいえないのではないかと感じている人もいるかもしれません。しかし，これはむしろ逆なのです。ラプラスがさまざまな分野の問題に一貫してベイズの定理を用いて取り組んだように，ベイズ統計学は，2 つの自然な原則を，どんなデータ分析場面にも一貫して適用していき，認知バイアスのような非合理的な推論を避けることのできる，私の師匠の言葉を借りれば「骨太」で一貫した体系なのです。

それでは，そんなベイズ統計学が，どうして長らく傍流に甘んじていたのでしょうか。そして，なぜ現代，再び注目され利用されてきているのでしょうか。次回はこのあたりの話題を，心理学との関連にも焦点をあてつつ，見ていくことにしたいと思います。

文献・注

(1) Wikipedia「Thomas Bayes」より。
https://en.wikipedia.org/wiki/Thomas_Bayes

(2) プライス（Richard Price：1723-1791）。ウェールズの哲学者，数学者。

(3) Bayes, T., & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53**, 370-418.

(4) ヒューム（David Hume：1711-1776）。スコットランド，エディンバラ出身の哲学者。

(5) ラプラス（Pierre-Simon Laplace：1749-1827）。フランスの数学者，物理学者，天文学者。

(6) Wikipedia「Bayes' theorem」より。
https://en.wikipedia.org/wiki/Bayes%27_theorem

判したのかについてです。

ベイズ統計学の冬の時代

冬といえばクリスマス。当地では，こんなふうに車にトナカイの角と赤鼻をつけるのが流行していました。

この原稿を書いているいまは 1 月末。冬真っ盛りのはずの時期ではありますが，西海岸の南カリフォルニアは温暖です。晴れの日が大半なこともあり，半袖で川沿いのサイクリングロードを毎朝気持ちよく通勤しています。もっともカリフォルニアでは水不足が深刻で，川といっても水はほとんど流れていないのですが……。

さて，前回は 18 世紀にベイズ牧師がベイズの定理を発見し，ラプラスがそれを広く普及させる形で発表した，というベイズ統計学の誕生についてお話ししました。ベイズ統計学は，現実世界の不確実性を確率という単一の尺度に落とし込んで表現し，データに基づいて確率を更新していきます。この更新に，ベイズの定理を用います。ベイズ統計学の考え方は，そのシンプルさと汎用性の高さから，19 世紀にはいろいろな分野で応用されるようになりました。しかし，広まるにつれ，この枠組みに対する批判も見られるようになりました。

20 世紀の大部分は，ベイズ統計学にとって冬の時代でした。これを決定づけたのは，現代統計学の祖といって過言ではない R. A. フィッシャー[1] と，ジェジー・ネイマン[2]，エゴン・ピアソン[3] らの，1920 ～ 30 年代を中心とした仕事です。フィッシャーは，ベイズ統計学とは異なる，頻度論的統計学の枠組みを確立しました。ネイマンとピアソンはこれを推し進め，仮説検定という統計的意思決定法を確立しました。今回のテーマは，彼らがなぜベイズ統計学の枠組みを批判したのかについてです。

確率のアップデート

下の図は，ベイズ統計学に基づく推論を模式的に示したものです。

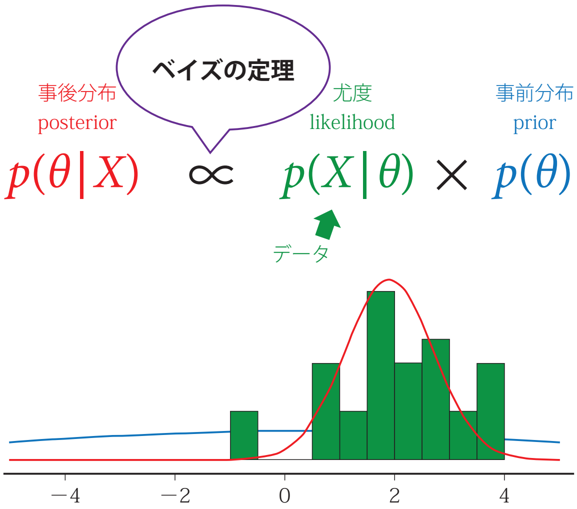


図　ベイズ統計学による推論の模式図

例を挙げましょう。A さんは，海外のはじめて住む土地に引っ越してきたばかりです。家の最寄りのバス停に，定刻では 8 時着の路線バスがあるとします。このバスが実際にはいつ来るのか，についての推論を考えてみます。

図の横軸は定刻の 8 時を基準とした時間（分）を表すとします。青線で描かれているのは，バスの到着時間について，データを得る前に分析者がもっている確率分布，すなわち**事前分布**です。ここでは，この土地の路線バスがどのぐらい遅れるか，もしくは早く来るかについての情報がほとんどなく，0 を中心とした「つぶれた」事前分布になっています[4]。緑のヒストグラムは，実際に 1 カ月の間このバス停で収集した，到着時刻のデータを表します。そして，ベイズの定理は，事前分布にデータのもつ情報を表す**尤度**の項をかけ合わせて，**事後分布**を導きます。なお，「∝」は左辺が右辺に比例する，という意味の記号です。

こうして事前分布とデータから，ベイズの定理を適用して得られた事後分布が，1 カ月間の観測を経た後で，バスの到着時刻についての確率を表す事後分布になります。事後分布は赤線で示されています。

赤線の事後分布は，青線の事前分布よりも分布が釣り鐘型に引き締まっており，到着時刻について，より豊かな情報をもっています。具体的には，事後分布のピークは 2 より少しだけ小さいところにあり，定刻

の 8 時よりも 2 分弱ほど遅れてバスが来る可能性が一番大きいことがわかります。また，事後分布のほとんどの領域は 0 から 4 の間にありますから，8:00 〜 8:04 の間にバスが来ると思っていれば，ほぼ間違いなさそうなこともわかります。この確率を具体的な値として得たい場合には，図の 0 から 4 までの事後分布の面積を求めればよいのです。このように，ベイズ統計学はデータのもつ情報によって，事前から事後へと，ベイズの定理を用いて確率をアップデートしていく枠組みです。

ベイズ統計学への批判点

こうしたベイズ統計学の枠組みにはいろいろ優れた点があるのですが，お楽しみは次回にまわして，今回は批判のお話です。ベイズの定理を用いて確率の更新ができること自体は，証明できる数学の結果ですから，特に批判される点はありません。批判を受けうるのは，ベイズの定理を現実のデータ分析に応用する，その仕方についてです。

ベイズ統計学に対する批判のポイントは，大別して事前等確率の設定についてと，確率の主観的解釈についての 2 点があります。

事前等確率の設定

1 点目から見ていきましょう。ベイズ統計学はデータに基づいて確率をアップデートしていく枠組みですので，データが得られる前の確率を表す，事前分布を設定しなければなりません。素朴に考えると，特に事前情報がない場合には，「すべてのとりうる値について，確率は等しい」という，**事前等確率の設定**が妥当に感じます。事実，ベイズもラプラスも，事前等確率の設定を利用していました。この設定は「理由不十分の原則」とよばれることもあります。特に理由がなければ，事前分布に等確率を設定しよう，ということです。

もし関心があるのが「コインの表（裏）が出る確率」であれば，可能な結果は「表」と「裏」だけですから，事前等確率の設定を利用し，データを見る前には表が出る確率も裏が出る確率もそれぞれ 0.5 である，と考えることに特に問題はなさそうです。しかし，バスの時刻の例ではどうでしょうか？ 机上の空論的ですが，バスが着く時刻は非常に遅く，もしくは早くなってしまう可能性だってあるかもしれません。事前等確率の考え方に基づくなら，－∞（マイナス無限大）から＋∞（プラス無限大）までの範囲に，一様な事前確率を設定することになります。しかし，いくらデータがまだないとはいえ，この設定は妥当なものでしょうか。どんなに情報がなくても，定刻 8 時のバスが 7：55 〜 8：05 の間に来る確率は，8：55 〜 9：05 の間に来る確率よりもよほど大きいと考えるのが自然ではないでしょうか？ こうした考え方は，次回お話しする，より情報のある事前分布の利用につながります。

さらにいえば，じつは「－∞から＋∞までの範囲で一様である」という事前分布は，厳密な意味での「確率」の性質を満たしていません。確率の数学的な定義では，すべての場合について足し合わせると 100 %，つまり 1 になることが要請されています。しかし，「－∞から＋∞までの範囲で一様」の分布は，この要請を満たすことができないのです。こういったおかしな確率分布のことを，**非正則**（improper）**な分布**といいます。

じつは，非正則な事前分布を用いても，多くの応用場面では事後分布は通常の（正則な）分布になります。ですので，非正則な事前分布の利用は，ある 1 つの統計モデルのもとでの推論を考える限りは問題にならないことが多いです。しかし，2 つの統計モデルの間でどちらがより適切かを考えるような場合には，非正則な事前分布を使うとしばしば，答えが 1 つに定まらなくなる，といった大きな問題が生じます。このように，事前等確率の設定は，意味的な問題点だけでなく，モデル選択においては数学的な問題点を抱えています。

この問題を解決する方法は，機械的な事前等確率の設定を避けることです。しかし，そうした理論の体系が作られるようになったのは 20 世紀のフィッシャーらの時代以降のことでした。長い間にわたって，ベイズ統計学で事前等確率の設定を用いるのは半ば当然のことと考えられ，またそれに由来する批判を受けてきたのです。

アメリカではフィッシャーといえば，イタチのようなこの動物のことを指します。その名に反して，魚はあまり食べないのだそうです [5]

確率の主観的解釈

ベイズ統計学の枠組みに対するもう1つの批判点は，確率の解釈についてです。フィッシャーをはじめ，ネイマン，ピアソンらは，データを何度でも繰り返し得られる状況を考え，その極限，つまり繰り返し回数を限りなく大きくした場合の確率を，考慮の対象としました。例えば，さいころを振るという試行は，何度でも同じように行うことができそうです。そこで，「さいころを振って1の目が出る確率」とは，10回，100回，1000回……とさいころを振る回数をどんどん増やしていったときの，1の目が出る頻度の極限であると考えたのです。これを**確率の頻度論的解釈**といいます。頻度，つまり多数回繰り返したうちにそのことが生じる割合の極限として，確率を理解するのです。

しかし，確率の頻度論的解釈の問題点は，多くの現実場面で実際には使えないことです。理想的なさいころの状況と違い，現実世界で起きていることの大半は，一度きりなのです。例えば，「明日の東京での降水確率」を考えてみましょう。降水確率は天気予報などでおなじみですが，明日という日はこれまでのどの日とも違う，ただ1回限りの日です。「明日と同じ日を何度も繰り返す」ことはできないでしょう。したがって，頻度論の立場からは，明日の降水確率という量は，解釈することが困難になってしまいます。降水確率だけでなく，現実に私たちの関心がある多くの問題は，同じことを何度も反復することが難しい，もしくはできないのです。

一方，多くのベイズ統計学者は，確率を単に「確かさ」を数量化したもの，別のいい方をすれば「信念の度合い」だと解釈します。これを，頻度論的確率と対比して，**主観確率**といいます。「信念の度合い」というと非常に個人的で科学とは相容れないものに感じられるかもしれません。しかし，これは応用の仕方に依存し，たしかに「Aさんが固有にもつ信念」として利用することもできますが，より広く「私たちが一般にもつ確かさの度合い」として利用することもできます。そして，後者の使われ方が実際には大半です。主観確率の考え方のポイントは，多数回の繰り返しが想定できない場面においても，確かさを量的に表す共通の単位として確率を使おう，ということなのです。これにより，一回性の事象に対しても，多数回の繰り返しができる状況と同様に，統計学が扱うことができるようになります。統計学の適用範囲が，現実の問題へと大きく広がるのです。

しかし，フィッシャーはこうした利点に目を配らず，主観確率がもちうる主観性を強く批判しました。彼は頻度論の枠組みで，有意性検定という，データによって理論や仮説を定量的に評価する枠組みを作りました。また，ネイマンとピアソンは帰無仮説検定という，2つの仮説から1つを選ぶ枠組みを確立しました。こうして完成した頻度論に基づく仮説検定の枠組みが，20世紀の，心理学をはじめ諸科学のデータ分析を支配したのです。

「不死身の理論」

しかし，ベイズ統計学は息絶えてはおらず，その研究は脈々と続けられていました。そして，特に20世紀の終わり頃から，再び息を吹き返し，大きな盛り上がりを見せました。このような経緯から，ベイズ統計学のことを**不死身の理論**（The theory that would not die）とよぶ人もいます。

なぜこの再興が生じたのか。そして，再興したベイズ統計学は上記2点の批判を，どのように解決したのか。こうした点を，次回は見ていきたいと思います。

文献・注

(1) フィッシャー（Ronald Aylmer Fisher：1890-1962）。イギリスの統計学者。
(2) ネイマン（Jerzy Neyman：1894-1981）。数理統計学者
(3) エゴン・ピアソン（Egon Sharpe Pearson：1895-1980）。イギリスの数理統計学者。父は統計学者のカール・ピアソン。
(4) 特に海外では，定刻より早くバスが来てしまうこともあるようです。要注意ですね。
(5) http://www.ForestWander.com より。

Section 3

2月のアーバインは連日，最高気温が25度を超える夏日。短い冬でした。

世紀末，という言葉は，ともすると終末論的な意味合いで語られがちです。しかし，実際の 20 世紀末には，恐怖の大王がどこからか降臨することはありませんでした。それどころか 1990 年代に入って，心理学を始めとする実証的な学問分野ではベイズ統計学再興の機運がにわかに高まりました。長かったベイズ統計学の冬が終わり，春がやってきたのです。

その原動力は 2 つあります。1 つは頻度論に基づく帰無仮説検定・有意性検定に科学の仕組みが依存しすぎたことに対する反動であり，もう 1 つは計算機の発展と軌を一にしたマルコフ連鎖モンテカルロ法の発展です。

今日は特に前者のお話を，そして次回に後者のお話を書きたいと思います。

従来の仮説検定の問題点

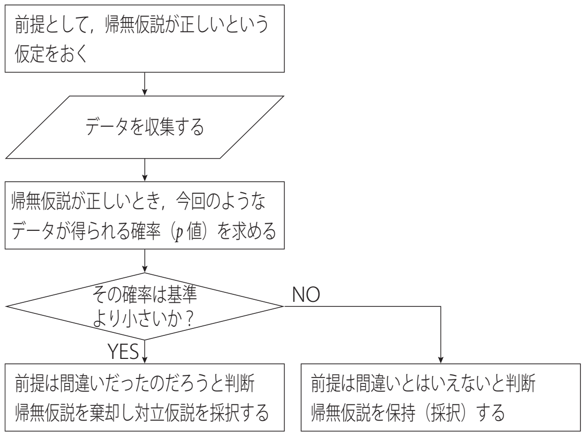
フィッシャーとネイマン，ピアソンらが確立した頻度論に基づく仮説検定は，20 世紀の大半を通じ，統計的データ分析の主役でした。「平均値の差の t 検定」や「クロス表のカイ二乗検定」といった検定は，大学でデータを扱う分野を専攻すれば，一度は触れる機会があるでしょう。科学者コミュニティの中でデータ分析が認められ，論文として世に出るためには，仮説検定を行い「$p < 0.05$」という結果を出す必要がある。それが，「統計学的有意性」を示したことになる。心理学を始め多くの分野で，こうした考え方が主流の時代は長く続きました。

しかし，仮説検定のアイディアはたしかに有用なのですが，科学の仕組みがあまりに従来の検定に依存しすぎたために，その弊害が表面化するようになりました。特にロジックの観点から，今回は 2 つの問題点に注目しましょう。

問題点その 1――仮説の非対称性点

仮説検定では，**帰無仮説**と**対立仮説**という 2 つの仮説が登場します。例えば，血圧を下げる薬の効果を調べるために薬を飲む前後で血圧を測定する状況を考えましょう。このとき，帰無仮説は薬を飲む前後で測定値に「差がない」という仮説，対立仮説は「差がある」という仮説です。

仮説検定には 2 つの仮説が登場するので，公平に 2 つを天秤にかけるのだと思われるかもしれません。しかしじつはそうではなく，仮説検定で考えるのは，帰無仮説が正しい状況だけなのです。仮説検定の流れは，簡略化して書くと次のチャートのようなものです。



ここからわかる通り，仮説検定で検討するのは帰無仮説が正しいという前提のもとでの話だけなのです。仮説検定では，帰無仮説が正しいときに，いま手元にあるようなデータが手に入る確率[1]，すなわち ***p* 値**を求めます。もし p 値が事前に定めた基準（典型的には 0.05）よりも小さい場合には，帰無仮説が正しいという前提のもとではめったに得られないデータが得られたことになります。ですので，そもそもの前提，すなわち帰無仮説が間違いだったのだと結論づけ，代わりに対立仮説を採用します。これが仮説検定のロジックです。はじめにわざと言いたいのとは逆の仮定をし，データがそれが整合的ではないと示すことによって言いたいことを主張する，という点で，このロジックは数学の証明で使う背理法と似ています。

このロジックを 2 つの仮説の間での選択に使ううえで，問題点は何でしょうか。それは，帰無仮説を積極的に支持できないことです。もし p 値が事前に定めた基準より大きい場合，それはデータが帰無仮説のもとでめったに生じないわけではないことを意味します。ですが，このことは，帰無仮説を積極的に支持するわけではありません。実際，ほかの仮説のもとでは，もっと得られやすいデータかもしれないのです。仮説検定は帰無仮説を前提としたもとで話を進めており，対立仮説を含むほかの仮説については直接考えないので，枠組み上，帰無仮説を積極的に支持することができないのです。

問題点その 2――仮説の確率が求められない

ところで，先に「p 値」が登場しました。p 値は，統計用語としては知名度がある用語です。しかし，その意味は誤まって理解されていることも多いです。p 値の値は，いったい何を表すのでしょうか？

答えはすでに先ほど（より正確には注に）書きましたが，よくある誤解に，p 値は「帰無仮説が正しい確

率」であるというものがあります。これは，クジラは魚である，とかオーストラリアの首都はシドニーである，などと同じように，よくある誤解です。

p 値は「帰無仮説が正しい確率」である，という考えが誤りなのは，頻度論における確率の考え方に由来します。統計学では，データ生成の背後に確率的なメカニズムを考えます。前回お話ししたように頻度論では，試行を繰り返したときの頻度の極限として確率を解釈するのでした。ある仮説が正しいときに，データを繰り返し得ることは，少なくとも概念的に可能です。しかし，あるデータが与えられたときに，そこから「仮説」についての繰り返しを想定することはできないでしょう。あるデータを生じたメカニズムについての仮説は，正しいか，もしくは間違っているかのどちらかです。たとえそのどちらなのかが不確実であっても，仮説を繰り返す？ということが考えられない以上，仮説についての確率は考えることができないのです。「帰無仮説が正しい確率」は，頻度論ではそもそも考えることができない量だったのです。

そこで従来の検定では，帰無仮説が正しい確率を考える代わりに，帰無仮説が正しいことを前提としたときに手元にあるようなデータが得られる確率を考えることにしました。これが p 値です。p 値の考え方は，多分に頻度論的なものだったのです。

しかし，しつこくて恐縮ですが，p 値を「帰無仮説が正しい確率」であると思ってしまうことは，私の講義での経験からもとても多い誤解です。データ分析は仮説が支持されるかどうかを検証するために行っているのですから，仮説についての確率を評価したいというのは，むしろ自然に考えることなのだと思います。

じつはこれ，ベイズ統計学ならできるんです。

ベイズ統計学では？

ベイズ統計学の立場からは，従来の検定の 2 つの問題点はどうなるでしょうか？

まず 2 つめの，仮説の確率についてから見ていきましょう。ベイズ統計学では，確率は不確実性の度合いを表す量だと理解します。これにより，多数回の繰り返しが考えられない状況でも，確率的な議論ができるようになるのでした。したがってベイズ統計学の立場からは，手元のデータの生成メカニズムとして帰無仮説が正しい確率を考えることが，問題なくできます。繰り返しが可能かどうかは，ベイズの立場からは問題ではないのです。

ただし，仮説の確率を実際に求めるにあたっては，注意すべき点が 2 つあります。

第 1 に，ある仮説の確率とは，それ単独では考えることはできない相対的なものです。ほかにどんな仮説が候補としてありうるのか，ほかの候補がいくつあるのかなどによって，仮説の確率は変わります。このことは，すべての場合について合計すると 1（100%）になる，という確率の重要な性質に由来します。

第 2 に，単にある仮説の確率よりも，データによって仮説を支持する証拠の大きさがどれだけ変化したのか，の方が実際的にはより重要なことが多いです。例えば，明日の天気は晴れか雨かを考えましょう。ここ南カリフォルニアは，冒頭に示した青空とヤシの木の写真でイメージされる通りの乾燥した気候で，特に夏の間は普通，ほぼまったく雨が降りません。ですので，予報士が夏のある日，気象データに基づいて「明日は晴れでしょう」という予報を出したとしても，もともと晴れの事前確率が高かったですので，それほど価値のある情報が得られたとは言い難そうです。一方，気象データに基づいて「明日は雨でしょう」という予報を出した場合はどうでしょう。今度はデータ分析をする前後で支持される仮説が大きく変わったのですから，データ分析から得られた情報の価値は高いといえるでしょう。このように，単に仮説の確率だけではなく，事前から事後へと仮説の確率がどう変わったのかを評価することは大切なのです。

こうしたことから，S. H. ジェフェリーズ[(2)] はベイズファクター[(3)] という量を使って 2 つの仮説の間での相対的なよさを測ることを提案しました。ベイズファクターはデータをとる前と，とった後とで，一方の仮説が他方に比べて，よりどれだけ支持されるようになったかを確率の比（オッズといいます）の尺度で表す量です。帰無仮説に対して対立仮説を評価する場合には，ベイズファクターが 1 より大きければ，データの情報によって対立仮説の事後オッズが事前オッズよりも大きくなったこと，つまりデータが対立仮説をより支持したことがわかります。ベイズファクターが大きければ大きいほど，支持の度合いは大きくなります。一方，ベイズファクターが 1 より小さい場合には，データが帰無仮説をより支持することになり，その度合いも小さければ小さいほど大きくなります。

ベイズファクターは，2 つの確率についての確率の比であり，どちらの仮説も平等に扱われています。したがって，帰無仮説が正しい状況しか考えていない，という従来の仮説検定の 1 つめの問題も，ベイズ統計学のアプローチでは解消されることになります。

客観事前分布

ただし，ベイズファクターを従来の仮説検定と同じ文脈で使うにあたっては，やはり注意点があります。それは，対立仮説のもとでの事前分布をどう設定するかということです。

平均を 0 に制約する帰無仮説と，制約をおかない（0 ではないという）対立仮説を考えましょう。このとき，対立仮説のもとでは平均にどんな事前分布を設定するのがよいでしょうか？ 平均は理論上どんな大きな・小さな値もとりうるということで，−∞から＋∞までの一様事前分布を設定したくなるかもしれません。しかし，前回少し触れたように，2 つの仮説を比較する文脈で非正則な事前分布を使うと，ベイズファクターの値を 1 つに定められなくなってしまいます。

そこで，発想を転換します。ジェフェリーズは，事前分布がもつべきよい性質をまず挙げ，それを満たすような事前分布を導出して使うことを提案しました。彼はその事前分布に必要とされる性質として，変数変換をしても結果が変わらないことを重視しました。m を cm に直す場合のように，単に単位・尺度を変換しただけで分析結果が変わってしまうのは望ましくないですよね。ですので，そうした事態を避けられる事前分布を使おう，というアイディアです。そうしてジェフェリーズは，分析結果が変数変換の影響を受けない事前分布を導きました。その後，彼のアイディアに端を発し，さまざまな状況のもとで望ましい性質をもつ事前分布を導く研究が進みました。現代では，平均の差，相関，クロス表といったように，よく使われる統計モデルでの典型的な仮説検定に対応するベイズファクターを求める際には，客観事前分布とよばれる，汎用的な分布を利用することができるようになっています。

ま と め

今回見てきたように，本来の守備範囲を超えて使われるようになってしまった頻度論的な仮説検定への反動は，ベイズ統計学への注目を高める大きな要因の 1 つになりました。近年，ヒトの予知的超能力の証拠が実験で示されたと，D. J. ベムという心理学者が報告して[(4)] 大きな議論をよぶ事件がありました。しかしその後，この結果は，今回お話した問題点その 1 のような従来型の仮説検定の特性のために得られた結果，という側面が大きいようであること，ベイズファクターを用いて再分析をすると「超能力」の証拠はむしろ乏しいことが報告され[(5)]，ほとぼりが冷めました。この事件は多くの注目を浴びたぶん，多くの心理学者に，従来の仮説検定の問題点をあらためて気づかせるのに一役かったように思います。

次回は世紀末に始まるもう 1 つ動きである，計算機統計学的なベイズ統計学の再興と，心理学への応用について見ていきたいと思います。

文献・注

(1) *p* 値の意味を自然言語で正確に表現するのはなかなか厄介です。本文とチャートではやや端折った表現をしましたが，より正確性を高めると，*p* 値が意味するのは「帰無仮説が正しいことを前提としたときに，今回得られたのと同じか，またはそれ以上に帰無仮説と整合的でないデータが得られる確率」，となります。

(2) ジェフェリーズ（Sir Harold Jeffreys）。現代ベイズ統計学の基礎を築いた，フィッシャーと同じ時代に生きた統計学者。

(3) ベイズ比，ベイズ因子，ベイズ因数などと訳されることもあります

(4) Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, **100**(3), 407-425.

(5) Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, **100**(3), 426-432.

Section 4

アーバインを走る高速道路の 1 つ，I-405。アメリカでも最も交通量の多い高速道路であり，車社会のロサンゼルス都市圏を象徴しています。

ロサンゼルスはアメリカ第 2 の都市ですが，鉄道網が発達しておらず車移動が中心の，先進国では一風変わった都会です。少し前までは，自家用車を使わずに目的地に行くには，不便なバスを乗り継ぐか，高い料

金を払ってタクシーを呼ぶしかありませんでした。ですが，Uber や Lyft といったスマートフォン上で配車から支払いまで完結するライドシェア・サービスが登場し，学生も大学教員もいまでは日常的に使っています。私も何度も利用しましたが，こうしたサービスの登場以前に当地で暮らすのは，もう少し簡単でなかっただろうなと思います。

Uber や Lyft が当地の人たちの生活を便利にしたように，今回紹介するマルコフ連鎖モンテカルロ法という技術，そして BUGS に代表されるそれを実装したソフトウェアは，ベイズ統計学を用いたデータ分析を，便利で実用的にすることに大きく貢献しました。

積分計算の困難

ベイズ統計学は，ベイズの定理を用いて事前から事後へと確率を更新していく枠組みでした。ここで，ベイズの定理の式の分母は，エビデンスとよばれる，データについての確率です。この項はパラメータのすべての可能な値を考慮して求めるため，連続変数の場合にはそれについて積分する必要があります。特にパラメータが複数ある場合には，多重積分が必要になります。

また，複数のパラメータを含む統計モデルを考えている場合でも，第一義的な関心は，各パラメータについて個別の点推定値（例えば事後平均）やばらつきの大きさ（例えば事後標準偏差）であることが多いです。この計算にもやはり，事後分布をほかのパラメータについて多重積分する必要があります。

こうした確率分布の積分計算は，事前分布をかなり工夫して設定しない限り，一般に容易ではありません。少し複雑な統計モデルになると，事実上この積分計算は解析的に行えなくなってしまうのです。

ベイズ統計学は直感にも合い，一貫して合理的な枠組みでした。ですが，1980 年代まではこの積分計算の困難さによって，単純な統計モデル以外では，求めたい事後確率を実際に計算することができませんでした。このために，ベイズ統計学は理想論であって現実に合っていない，「考え方はわかるけれども実際に使えない」という状況が生じてしまっていました。

マルコフ連鎖モンテカルロ法

この大きな問題に対し，A. E. ゲルファンドと A. F. M. スミスは 1990 年の論文で，困難な積分計算を数値的なサンプリングで置き換える方法を提案しました[1]。彼らのアイディアは，困難な多次元の積分演算を行う代わりに，1 次元ずつの分布から発生させたランダムな数（乱数）をたくさん集めることで，実用上十分な精度で事後分布について知ることができるというものです。この方法を使えば，例えばあるパラメータの事後平均を知りたい場合には，単純に，そのパラメータについて何千個も発生させた乱数の平均を計算すればよいのです。

こうした方法はいまでは，マルコフ連鎖モンテカルロ（Markov chain Monte Carlo: MCMC）法と総称されます。MCMC 法の特長の 1 つは，その汎用性の高さです。以前のように，統計モデルごとに個別に開発されたアルゴリズムを使うのではなく，基本となるアルゴリズムを非常に幅広い統計モデルの推論に適用できるのです。また，MCMC 法は分布の全体から発生させた乱数を利用するので，単に点推定だけでなく，任意の確率を与えての区間推定をすることもでき，さらにモデル選択のために拡張することもできます。MCMC 法の登場によって，ベイズ推論が実用的なものになったのです。

BUGS と Stan

また，統計手法が普及するにあたっては，ソフトウェアの存在が不可欠です。汎用性の高さという MCMC 法の特長を生かせば，さまざまなモデルを柔軟に推定できるソフトウェアをつくることができそうです。

1990 年代初頭から，こうした期待に最初に応えたのが BUGS プロジェクト[2]です。BUGS はスミスのもとで学んだ D. シュピーゲルハルターらのグループが開発した，MCMC 法を用いたベイズ推定を行うためのソフトウェアです。BUGS の大きな特徴は，ユーザーはモデルを記述しさえすれば，ソフトウェアがそのモデルについて MCMC 法を用いた推定を実行してくれ，ユーザーが細かい計算アルゴリズムをプログラミングしなくてよいことです。利用する確率分布や変数の数，変数間の依存関係などを応用場面に合わせて自由に研究者が指定できることになり，統計モデリングがますます便利になりました。特に Windows 版の WinBUGS は長い間，幅広いユーザーに利用されてきました。その後 21 世紀に入って，BUGS プロジェクトはオープンソース版の OpenBUGS に移行しました。また M. プラマーによって独立に開発され，安定さに定評のある JAGS[3]の開発にもつながりました。

一方，コロンビア大学の A. ゲルマンらのグループ

私も参加した，昨夏の世界最大規模の統計学の学会（JSM[5]）で披露された映画のような Stan のトレイラー。講演の最初に流され，会場は大きく盛り上がりました。https://www.youtube.com/watch?v=pWow8Qe1snQ

は近年，新しい MCMC 法を実装したソフトウェア Stan[4] を公開しました。この中で利用されているハミルトニアン・モンテカルロ法という MCMC のアルゴリズムは，階層性や潜在変数を含むような複雑なモデルでも，サンプリングを効率よく行えることが特徴です。Stan の開発チームは精力的に開発を進めており，新しい機能もどんどん実装されていて，いま最も活発な MCMC ソフトウェアのコミュニティとなっています。

なお，ここで述べたソフトウェアはいずれも無償で公開されています。

統計モデリングの重要性

MCMC 法の登場，そして，それを実装した BUGS や Stan のようなソフトウェアの隆盛により，私たちはもっぱらデータに合ったよい統計モデルをつくることに集中できるようになりました。統計モデルとは，データがどのようにして得られたのか，その背後にある規則や生成メカニズムを，確率分布を使って数学的に表現したものです。表現するモデルのベイズ推定は，基本的にソフトウェアに任せることができるからです。このため，現代ではよい統計モデルを構築することの重要性が，これまでに以上に重要になっています。

前回の話とも一部重なりますが，一昔前の典型的な統計分析とは，「○○検定」や「○○分析」といった，いくつかの定型的な方法があらかじめ組み込まれたソフトウェアにデータを読み込み，数回クリックして分析を実行することでした。実験心理学でいえば，分散分析モデルによる F 検定と，その後の下位検定（多重比較）という流れが典型的でした。しかし現代では，柔軟なモデルのベイズ推定ができるのですから，データがどのような機序で観測されたのか，その生成メカニズムを，そして背後の理論や集団差などを組み込んだ統計モデルを構築できるのです。よいモデルは，データからより多くの情報・知見を教えてくれます。そしてモデルのよさは，前回紹介したベイズファクターや，予測の観点から評価をすることができます。有意性検定から，統計モデリングへ――これは現在の統計学の応用における大きな潮流と言ってよいと思います。

色覚の認知モデリングの例

筆者が関わった一例を挙げてみましょう[6]。心理学は古くから色の見え方の研究に取り組んできました。色相を決めるのはおもに光の波長です。可視光の範囲で，物理的に波長の長い光は赤く見え，波長が短い光は紫色に見えます。しかし，人が主観的に感じる色は，赤と紫を両極とした 1 次元的な感覚ではありませんね。むしろ，可視光の範囲で一番波長の長い赤と，一番波長の短い紫とは比較的近い色に感じられると思います。

C. E. ヘルム[7] は実験参加者に 2 つの色を提示し，その色のペアがどのぐらい似ているかを尋ねて回答を記録しました[8]。10 種類の色が用意されたので，可能なペアの組み合わせは 45 通りあり，そのすべてについて参加者たちから主観的な色ペアの類似度データが得られました。このデータを用いて，人間の色の感覚をどのようにモデリングできるでしょうか。

こうした場合に利用できる統計モデルに，多次元尺度法があります。多次元尺度法モデルでは，観測された類似度データの背後に，少数の次元によって張られる心理的空間を想定し，そこで近くに布置される対象はより似ている，遠い対象はより似ていない，という評定が観測されやすくなると考えます。観測される評定データの背後に，直接観測できない心的次元を考える点で，多次元尺度法は認知モデルと考えることができます。

下の図は私たちがベイズ階層モデルに拡張した多次元尺度法によって，この色の類似度データを分析した結果です。黒の点は，それぞれの色の事後平均を表します。また，灰色の点は事後分布から取り出した MCMC サンプルであり，各色の事後分布の範囲を視覚的に表現しています。この中で一番波長が長いのは赤紫（red-purple）で，波長が短いのは紫（purple-1）ですが，物理的な波長が 1 次元的なのに対し，人が感じる色の類似－非類似関係は，2 次元の空間で円環状になっていることがよくわかります。また，事後分布の広がりからそれぞれの色の不確実性はそう大きくないこともわかります。

さらに，このベイズモデルでは 2 つの次元をそれぞ

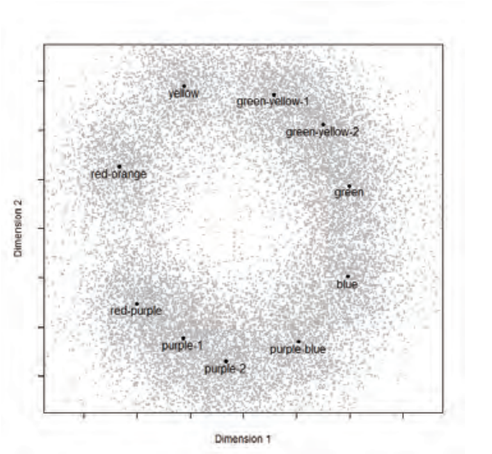


図　色ペアの類似度データを，個人差を考慮したベイズ多次元尺度法によってモデリングし推定した色の心理的空間

れどのぐらい重視するかに個人差を想定しており，2つの次元にも意味があります。ここでは左右方向が赤－緑，上下方向が青－黄色の次元になっています。同じ方法を赤－緑型の色覚異常のある参加者の評定にも適用したところ，想定される通り第1次元への重みがとても小さくなり，赤 - 緑の違いが色の類似度判断に影響を与えなくなっていると解釈できました。

このように，単なる色のペアごとの類似度評定データだけからでも，ベイズモデリングによって，評定の背後にある心理メカニズムや個人差について単純にデータを眺めたり平均を計算したりするだけではわからない，新たな知見を得ることができるのです。

おわりに

4回にわたって，心理学への応用を念頭においた，ベイズ統計学の紹介をしてきました。この連載の終わりとともに，私も日本に戻ります。専修大学の私たちの研究室や，心理学科のメンバーでは，これからますますベイズ統計学を活用した心理学研究を進めていく計画でいます。短い連載でしたが，統計分析・統計モデリングをこれまでよりももっと有効に利用する，何かのきっかけになったら嬉しいな，と思います。ご愛読どうもありがとうございました。

文献・注

(1) Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**(410), 398-409.

(2) BUGS プロジェクトのサイト
http://www.mrc-bsu.cam.ac.uk/software/bugs/

(3) JAGS のサイト
http://mcmc-jags.sourceforge.net/

(4) Stan のサイト
http://mc-stan.org/

(5) アメリカ統計学会の Joint Statistical Meeting のサイト
http://www.amstat.org/meetings/jsm.cfm

(6) くわしくは Okada, K., & Lee, M. D. (2016). A Bayesian approach to modeling group and individual differences in multidimensional scaling. *Journal of Mathematical Psychology*, **70**, 35-44 をご参照ください

(7) Helm, C. E. (1959). A multidimensional ratio scaling analysis of color relations: A Technical Report. Psychology Department, Princeton University and Educational Testing Service.

(8) 実際には色の三つ組を用意した，もう少し複雑なデータ収集法がとられています。くわしくは元論文をご覧ください。

著　者

岡田　謙介（おかだ・けんすけ）：

専修大学人間科学部准教授。主要著作・論文に，A Bayesian approach to modeling group and individual differences in multidimensional scaling.（*Journal of Mathematical Psychology*, 2016，共著），『伝えるための心理統計――効果量・信頼区間・検定力』（勁草書房，2012年，共著）など。web サイト : http://www3.psy.senshu-u.ac.jp/~ken/，Twitter: @kenmetrics。