

# 心理学におけるベイズ統計モデリング

清水 裕 士

関西学院大学

## Bayesian statistical modeling in psychology

Hiroshi SHIMIZU

Kwansei Gakuin University

The use of Bayesian statistics in psychology has been of recent focus. This paper discusses the effectiveness and usefulness of “Bayesian statistical modeling” in psychological studies by exploring its differences with traditional methodologies used in psychological studies. First, we explain differences between two trends in Bayesian statistics: hypothesis testing using Bayesian statistics and Bayesian statistical modeling. Second, Bayesian estimation and probabilistic programming languages may make it easy for psychologists to use statistical modeling. Third, the three advantages of studies using Bayesian statistical modeling in psychology are demonstrated. These advantages include developing mathematical explanations of behavioral mechanisms, valid estimation of psychological characteristics, and improved transparency and replicability of the data analysis. Fourth, it is argued that Bayesian statistical modeling and traditional psychological methodology could coexist by influencing each other.

**Key words:** statistical modeling, Bayesian estimation, MCMC, probabilistic programming language, hierarchical model

キーワード：統計モデリング、ベイズ推定、MCMC、確率的プログラミング言語、階層モデル

### 1. はじめに

本論文では、心理学におけるベイズ統計モデリングの有用性について、心理学の伝統的なデータ分析手法との比較を通して論じることを目的とする。はじめに、最初に本論文の構成についての交通整理をしておきたい。

まず2節では、ベイズ統計モデリングの「ベイズ」部分についての整理を行う。心理学においてベイズ統計学が注目されるようになったのは、著者の考えでは2つの流れがある。それは、頻度主義統計学へのアンチテーゼとしてのベイズ統計学と、MCMCを中心としたベイズ推定によるベイズ統計モデリングである。本論文で強調したいベイズ統計モデリングの「ベイズ」とは後者のことを示す。

次に3節では、ベイズ統計モデリングの「統計モデリング」とは何なのかを論じる。これまでの伝統的な心理学の方法論と何が違うのかについて触れ、数理的な側面では違いはないが、その目的

に違いがあることを示す。

続いて4節では、統計モデリングをベイズ統計学によって解くことのメリットを論じる。この部分で特に強調したいのが確率的プログラミング言語の登場によって、モデル構築とアルゴリズム構築の分離が可能になったことである。それにより、心理学者にとって統計モデリングが非常に強力な武器になりえることを論じる。

これらの議論を踏まえたうえで5節では、本論文のメインの主張である、ベイズ統計モデリングが心理学にどのようなメリットがあるのかを論じる。そのメリットとして、ベイズ統計モデリングによって行動発生メカニズムの数理的な表現および行動の予測ができること、心理的特徴についてより理論的に妥当な推定ができること、そして分析結果の透明性と再生可能性を高めることになること、の3つが挙げられる。

最後に6節では、ベイズ統計モデリングはこれまでの伝統的な心理学の方法とは異なるが、それらは排他的ではなく共存可能であり、また両立す

べきものであることが論じられる。加えて、これからベイズ統計モデリングの研究を実践していくために知っておくべきこと、注意すべきことについてまとめ、今後の展望を述べる。

## 2. 今心理学をにぎわせている「ベイズ」とは何か

近年、日本の心理学の界隈においても、「ベイズ」という言葉を聞く・見る頻度が急増している。たとえば2017年度に開催された日本心理学会第81回大会では、ベイズ統計学に関わる企画が4つ開催された。また、心理学者、心理統計学者が執筆した書籍においてもベイズ統計学についての言及が多くみられる。たとえば大久保・岡田(2012)では帰無仮説検定の限界点を指摘しながらベイズ統計学の可能性に言及しており、豊田(2015, 2016, 2017)の一連の書籍ではさらに強く、伝統的な帰無仮説検定をやめてベイズ統計学を用いることを推奨している。また心理学研究法についての入門的な解説書である三浦(2017)においても、新しい分析手法としてベイズ統計学が紹介されている。海外の心理学に関わる統計の書籍においても、Lee and Wagenmakers(2013)が認知モデルを統計モデリングで表現するという枠組みでベイズ統計学が紹介されている。

一方で、心理学以外の、特にデータサイエンスの分野においてもベイズ統計学についての書籍が日本においても数多く出版されている。その先駆

けとして久保(2012)があり、また岩波データサイエンスのシリーズの第1巻ではベイズ推論とその推定アルゴリズムであるマルコフ連鎖モンテカルロ法(Markov chain Monte-Carlo methods 以下, MCMC)について取り上げている(岩波データサイエンス刊行委員会, 2015)。さらにデータサイエンティストの松浦(2016)はベイズ統計モデリングをStanというソフトウェアを用いて解説している。またそれ以外にも、奥村・牧山・瓜生(2018)や伊庭(2018)など、ベイズ推定やMCMCを用いたモデリングのための書籍は立て続けに出版されており、その関心の高さがうかがえる。

これらの方法論における流行は、同じ「ベイズ」という言葉が使われているが、実は主張しようとしている点はいくつか違いがある。それを大きく分ければ、心理学にまつわる再現可能性の問題から近年改めて生じた「帰無仮説検定へのアンチテーゼとしてのベイズ統計学」の流れと、ベイズ推定の中でMCMCを中心とした汎用的な推定アルゴリズムが容易に利用可能になったことによる「ベイズ統計モデリングの流行」の2つがあると考えられる(図1)。なお、ベイズ統計そのものの有効性についての議論は、ずっと古くからなされているが、本論文で触れているベイズ統計の流行は、「近年再び心理学の研究者の間で注目され始めた」という意味で論じている。また、方法論としてのベイズ統計学以外に、認知メカニズムを表現するさいにベイズの定理を応用するアプローチも存在する。その点については本論文では取り

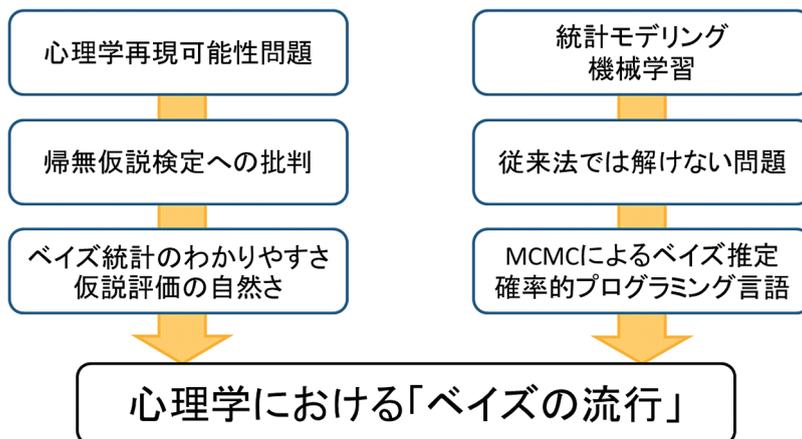


図1 心理学における「ベイズ」の流行

上げないが、本特集号の中村（2018）に詳しい。

そこで、この2つの「ベイズ」の流行について整理しよう。

## 2.1 帰無仮説検定批判とベイズ統計学

心理学の研究では、伝統的に帰無仮説検定が長く使われてきた（大久保，2016）。ほとんどの心理学系雑誌に掲載されている論文では $p$ 値の報告が行われていることからわかるように、帰無仮説検定は心理学、とくに実験心理学とは切っても切れない関係にある。

しかし、Bem（2011）の超心理学についての論文の出版を皮切りに、心理学における再現可能性の低さや「問題のある研究実践（Questionable Research Practices; 以下、QRPs）」の議論が活発になると同時に、帰無仮説検定の方法論的な限界点が指摘されるようになってきた。池田・平石（2016）でまとめられているように、帰無仮説検定では危険率を有意水準 $\alpha$ （通常5%が使われる）に抑え擬陽性を防ぐように設定されている。しかし、現状のデータ分析手法では $p$ -hackingと呼ばれる、 $p$ 値を不当に有意水準未満に導く方法が容易に可能となってしまっている。 $p$ -hackingには、データ改ざんなどの「悪意」があるもの以外にも、研究者が気づかずに犯してしまっているものも多く含まれていることから、帰無仮説検定の方法論的な問題点として指摘されることが多い。

上記のような帰無仮説検定批判の中で、代替案としてベイズ統計学が注目を浴びてきた。Wagenmakers et al.（2011）や Rouder and Morey（2011）では、Bem（2011）の超能力を肯定する論文への批判として、帰無仮説検定が帰無仮説と対立仮説を平等に扱わない点を批判しつつ、ベイズ統計学に基づくベイズファクター<sup>1)</sup>を用いた仮説検定の有用性を主張した。また Simmons, Nelson, and Simonsohn（2011）ではシミュレーションを用いることで、データ取得ごとに検定を行うという手続きが危険率を大幅に上昇させてしまうことを示した。Kruschke（2015）も伝統的な帰無仮説検定において $p$ 値のみに頼ることの問題点を指摘し、ベイズ統計による分析の優位性を主張し

ている。また2016年には Journal of Mathematical Psychology においてベイズファクターを用いた仮説検定についての特集号が組まれている（Mulder & Wagenmakers, 2016）ことから、ベイズ統計学に基づく仮説検定への注目度がわかる。なお、帰無仮説検定の問題点とベイズ統計学の利点についてまとめたものとして大久保（2016）がある。

これらの主張の背景には、帰無仮説検定で用いられる $p$ 値や信頼区間が非常に誤解されやすい概念であること、またそれによってQRPsが無自覚に誘発されやすいこと、などの帰無仮説検定の方法論的な脆弱性などがある。そして、ベイズ統計学を使えばこれらの問題が解決されるだろう、という期待がある。

ベイズ統計学は、帰無仮説検定が立脚する頻度主義統計学と異なる仮定から作られた統計学である。頻度主義統計学との典型的な、かつ、一番大きな違いは、頻度主義統計学では母集団のパラメータは固定値であると考えるが、ベイズ統計学ではパラメータも確率分布によって表現されることである。よって、頻度主義統計学は仮説の正しさについての確率を表現することができず、データの確率的な変動によってのみ仮説を評価できる。一方、ベイズ統計学は仮説の正しさについて確率的に表現することが可能となる。このような理論の違いによって、帰無仮説検定の持つ理解の難しさや、いくつかの限界点がベイズ統計学で乗り越えられると考えられている（たとえば大久保・岡田，2012）。

しかし、本論文では帰無仮説検定、あるいは頻度主義統計学との代案としてのベイズ統計学そのものは議論の対象とはしない。その意味で、上記の議論は本論文の本筋とは実はあまり関係がないが、「この内容の話ではない」と議論を区切ることで、今後の展開を明確にする整理の役割として論じておいた。

## 2.2 統計モデリングとベイズ統計学

統計モデリングにおいてベイズ推定が注目されるようになったのは、Gelfand and Smith（1990）によってベイズ推定のためのアルゴリズムとしてMCMCが利用できることが示されたことが一因である。これまで統計モデルは最尤法による推定が一般的であったが、変量効果を含む一般化線

1) ベイズファクターについては岡田（2014）や岡田（2018）の解説に詳しい。

形モデルである一般化線形混合モデル<sup>2)</sup>では、単純なモデルか、あるいは近似的な方法でしか解くことができなかった（久保，2012）。それに対してMCMCを用いれば、複雑な解法アルゴリズムをそのたびに開発せずとも、汎用的に解けるようになった。それに加え、複雑なモデルを簡単なコードを書くだけでパラメータの推定ができるWinBUGSが90年代半ばに開発されたことも大きい。日本では、久保（2012）において、一般化線形混合モデルをWinBUGSでMCMCによって推定することの利点が述べられ、広くベイズ推定の有用性が理解されるようになった。

心理学に関わるところでいえば、Lee and Wagenmakers（2013）が認知モデルを統計モデリングで表現し、WinBUGSを用いてMCMCで解く、という研究スタイルを示した。また、認知モデリングの発展から、計算論的アプローチの流れを組んだ、計算論的神経科学（Marr, 1982；乾・安藤（訳），1987）や計算論的精神医学（Adams, Huys, & Roiser, 2015）が、心理学に与えた影響もある。これについては国里（2018）に詳しい。さらには、前田・小杉（2017）によってKruschke（2015）が翻訳されたことも、日本の心理学業界でベイズ推定やMCMCの重要性が認知されるきっかけともなった<sup>3)</sup>。その後も、豊田（2017）がMCMCを用いた様々な心理学におけるモデリングを解説している。加えて、日本社会心理学会は2017年に「効果の科学からデータ生成過程の科学へ～心理学者のためのベイジアン・モデリング入門」と題したセミナーを開催した<sup>4)</sup>。これらの立て続けの出版や企画によって、ベイズ統計モデリングが日本の心理学者にも広く知れ渡った。

さて、本論文で論じようとしているベイズ統計モデリングの「ベイズ」はまさにここで論じている、推定法としてベイズ、つまりMCMCによっ

てモデルのパラメータを推定することを指している。なぜ統計モデリングにおいてベイズが重要な役割を果たすのかは、まず統計モデリングそのものについて解説してからのほうがわかりやすいだろう。よって、統計モデリングにおけるベイズ統計の有用性は後の3節に譲り、2節では統計モデリングについての解説を行う。

### 3. 統計モデリングと 伝統的なデータ分析手法との違い

本節では、統計モデリングとは何かを解説したうえで、伝統的に心理学で用いられてきたデータ分析手法、具体的には $t$ 検定や分散分析とどういう違いがあるのかを論じ、統計モデリングが持つメリットについて解説する。

#### 3.1 統計モデリングとは何か

統計モデリングは何も最近になって新たに出てきた方法論ではなく、回帰分析や一般化線形モデルを含む、確率分布を用いたモデリングの総称であり、古くからあるものである。もっと言えば、 $t$ 検定や分散分析も統計モデリングの範疇にあるといえる。日本における統計モデリングの流行は、ベイズ統計学とセットで始まったこともあるため、「統計モデリングはベイズ統計学を使う方法だ」と思われているかもしれない。しかし、あくまでベイズ統計学はパラメータ推定の枠組みのひとつであって、統計モデリングの本質ではない。

松浦（2016）は、統計モデリングとは「確率モデルをデータに当てはめて現象の理解と予測を促す営み」であるとしている。本論文でも松浦の理解を引き継ぎながら、統計モデリングを「確率モデルによってデータ生成メカニズムを記述し、データを使ってそのパラメータを推定し、それに基づいた推論や予測を行うもの」と定義する。そこでまず、「確率モデルによってデータ生成メカニズムを記述」という点について説明する。

確率モデルとは確率分布によって記述されたモデルのことである。たとえば成人男性の身長は、そのデータが平均 $\mu=170$ 、標準偏差 $\sigma=5.5$ の正規分布に従うとする。つまり、

2) 目的変数に正規分布以外の確率分布を仮定する一般化線形モデルに、個人差や集団差を考慮に入れたマルチレベル分析を組み合わせた方法である。詳しくは久保（2012）や日本社会心理学会春の方法論セミナー（[http://www.socialpsychology.jp/sympo/seminar\\_150325.html](http://www.socialpsychology.jp/sympo/seminar_150325.html)）の資料が参考になる。

3) 第81回日本心理学会大会においてKruschke（2015）を用いたチュートリアルワークショップが開催されている。

4) 春の方法論セミナーの資料は以下のURLから閲覧できる。[http://www.socialpsychology.jp/seminar/seminar\\_170314.html](http://www.socialpsychology.jp/seminar/seminar_170314.html)

$$Y_i \sim \text{Normal}(\mu=170, \sigma=5.5)$$

である<sup>5)</sup>。もし母集団の平均値のパラメータである  $\mu$  が未知であり、データから推測する場合は、

$$Y_i \sim \text{Normal}(\mu, \sigma=5.5)$$

となる。この平均パラメータ  $\mu$  は最小二乗法や最尤法などの方法で求められる。

心理学では多くの場合、母集団は人一般や日本人、あるいは、ある地域の有権者など、実在する集団が暗に仮定される。しかし、確率モデルとして現象を記述する場合、データが仮定した確率分布から生成されるという想定さえ成り立てばよい。具体的には次のようなことである。コインを10回投げて表が出た回数は二項分布 (Binomial distribution) に従う。二項分布は試行数  $T$  と成功率  $\pi$  の2つのパラメータを持つ。このとき、表が出る確率  $\pi$  を求めたいとする。つまり、

$$Y_i \sim \text{Binomial}(T=10, \pi)$$

である。先の正規分布の例と同じように、上のモデルでも確率モデルとパラメータが仮定されているが、標本抽出が仮定するような手続きを想定しているわけではない。たとえば、コイントスのモデルを標本抽出の理論の枠組みで理解する場合、表や裏を向いているたくさんのコイン (母集団) がおいてある中から10個だけランダムに選んだとき (標本) に、表が向いていたコインの個数がデータであると考えことになる。しかし、単に10回同じコインを投げたときに表が出た回数も、特定の確率分布に従ってデータが得られているという点では同様である。確率モデルの表現で重要なのはデータを得る具体的な手続きではなく、データがどういう確率分布に従って得られるか、という点だけである。

このことから、データ生成メカニズムを確率モデルで記述するというのは、データがどのように発生するかを、確率分布を使って数理的に表現する、という意味であることがわかる。その点においては2つの平均値の差についての  $t$  検定 (以下、単に  $t$  検定と呼ぶ) も統計モデリングとして表現

することも可能である。 $t$  検定の確率モデルは、

$$Y_{0i} \sim \text{Normal}(\mu_0, \sigma)$$

$$Y_{1i} \sim \text{Normal}(\mu_0 + \Delta, \sigma)$$

となる。ここで、 $\mu_0$  は統制群の平均値、 $\Delta$  は統制群と実験群の平均値の差を表すパラメータ、 $\sigma$  は両群で共通する標準偏差である。このように  $t$  検定は、2つの群でそれぞれ正規分布に従ってデータが生成されると仮定したうえで、2つの群の平均の差を表すパラメータ  $\Delta$  を推定し、さらにその標準誤差を求めることで  $p$  値や信頼区間を算出する、一種の統計モデリングであるといえる。

### 3.2 心理学における伝統的なデータ分析手法との違い

もし統計モデリングが心理学で用いてきた伝統的なデータ分析手法を包括するのであれば、両者には本質的な違いがないようにも思える。しかし、数理的な表現として同じでも、そこで強調される点や目的が異なっている。統計モデリングはデータ生成メカニズムを確率モデルで記述することだけが特徴なのではなく、上で示した定義の後半、すなわち「それに基づいた推論や予測を行う」という点にも重点が置かれているのである (久保, 2012; 松浦, 2016)。

心理統計学で用いられる  $t$  検定や分散分析 (とくに後者) は、もともと実験計画法の文脈で利用されることがほとんどである。実験計画法は、典型的には統制群と実験群を操作によって設定し、無作為に参加者を割り当て、従属変数である行動の出方の違いを比較するものである。この、統制群を設定した操作と無作為割り当てによって、処置や介入が従属変数に対して持つ因果効果を推定できる<sup>6)</sup>。処置や介入が効果を持つことが分かれば、そこから行動の背後にある心理メカニズムについての推論を論理的に行える。よって、実験計画法においては要因の効果の有無 (あるいは大きさ) を知ることが目的とされている。

また、心理学における研究計画は、要因の効果がある場合にどういった心理メカニズムがあるといえるかを導き出すための実験手続きとセットで作

5) 正規分布の尺度パラメータは分散である  $\sigma^2$  を用いることも多いが、ここでは後に述べる Stan の記法との対応から標準偏差  $\sigma$  を尺度パラメータとした。

6) なお、この統計的因果推論については宮川 (2004) や岩波データサイエンス刊行委員会 (2016) が詳しい。

られている。たとえばストループ課題では、認知的葛藤が、色パッチの色を答える課題と色付きの文字の色を答える課題で異なることが理論的に仮定される。そして、2つの課題に反応時間に差があることをもって、そこに認知的葛藤という構成概念からの影響を検証する、という研究手法をとっている。このように、要因の効果・条件間の差から心理学理論を検証できるような研究枠組みは、実験心理学が歴史の中から構築してきた強力な研究パラダイムである。

ただ、実験計画法と $t$ 検定や分散分析とのセットから、行動の予測を行うことは多くない。実験+分散分析は、群間の行動の平均的な違いを見るために用いられるが、次に参加者がどのような行動を行うかについての予測を導く目的で使われることが少なく、またその関心からも外れている。それは、伝統的な心理学の方法論がおもに行動の説明に重点が置かれており、予測が重視されていないからであると考えられる。

一方、統計モデリングが強調しているのは確率モデルからデータ生成メカニズムについて推論・予測することである。ここでわかりやすさのために、心理物理学的な研究例を挙げる。なお、モデルの表記などについては豊田（2017）を参考にしている。

ミューラーリヤー錯視において標準刺激と比較刺激が主観的に等しいと知覚できる主観的等価点（point of subjective equality; 以下 PSE）を求めることを考える。恒常法を用いた場合、参加者はランダムに提示された比較刺激について、どちらが長いかを回答する。このとき行動の生成メカニズムとしては「長い」と答えるか否かの2値であるため、二項分布が仮定できる。もし同じ比較刺激が8回提示されていたとすると確率モデルは試行数（ $T$ ）8回の二項分布となり、

$$Y_i \sim \text{Binomial}(T=8, \pi_i)$$

となる。そして、比較刺激を「長い」と回答する確率 $\pi_i$ は、比較刺激の長さ $X_i$ の関数と考えられる。そこでその関数が標準正規累積曲線によって表現できると仮定して、

$$\pi_i = \Phi(\alpha + \beta X_i)$$

と表現する。 $\Phi$ は標準正規累積関数を表してい

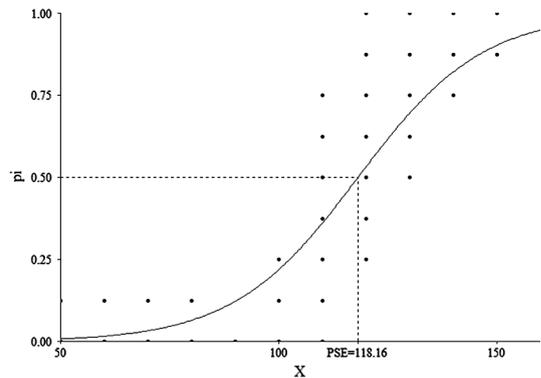


図2 標準正規累積曲線を仮定して推定したPSE

る<sup>7)</sup>。このとき、 $\alpha$ と $\beta$ がデータ $Y_i$ から推定されるパラメータとなる。標準正規累積関数を仮定した場合の「長い」と回答する確率 $\pi_i$ と、比較刺激の長さ $X_i$ の関係を表したのが図2<sup>8)</sup>である。またPSEは $-\alpha/\beta$ で求められることから、図2の例からは、参加者は比較刺激が118cmを超えたあたりから「長い」と回答するだろうことが予測できる。ここからわかるように、確率モデルを仮定したうえでパラメータを推定することで、理論的に知りたい値の推論や予測が可能になる。

また違う観点から心理学の伝統的なデータ分析手法と統計モデリングの違いについて論じ、統計モデリングの特徴を説明する。

統計モデリングでは、データが生成されるメカニズムを、確率モデルを使って「そのまま表現する」ところに重点が置かれる。一方、伝統的なデータ分析手法では、 $t$ 検定や分散分析のような正規分布を仮定したモデルを使うために、データの変換やモデルの補正が行われることがある。それは、正規分布を仮定したモデルが解きやすいこと、それに対応しているソフトウェアが豊富であることもその背景にはある。しかし、実際に得られる行動データが正規分布に従うわけでもないにもかかわらず、 $t$ 検定や分散分析、回帰分析と

7) 正規累積曲線を仮定した二項分布のモデルは、プロビット回帰分析で解くことができる。なお、

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{z^2}{2}\right) dz \text{ である。}$$

8) 図を作成するためのデータについては、大阪大学大学院人間科学研究科の武藤拓之氏が公開されているものを、許可を得たうえで用いた。<http://bayesmax.sblo.jp/article/181795845.html>

いった正規分布を仮定したわかりやすい方法を使うためだけに、データを正規分布に近くなるように変換する(対数変換など)といった処置が行われることがある。また、対応のない  $t$  検定においては群間で等分散の仮定があることが知られているが、その仮定とデータがあていないような場合には、Welch の検定といった、標準誤差や自由度を補正する方法があてがわれることもある。このようなデータの変換、あるいはモデルの補正は、確かに効果の有無を判断するためだけならば十分である。なぜなら、効果の有無を知るだけならば、差についての有意性検定の結果が信頼できればそれでよいからである。

一方で、データ生成メカニズムを確率モデルで記述したいという統計モデリングの立場から見れば、データの変換やモデルの補正といった処置は目的に合わない。なぜなら、反応時間のデータを対数変換して分散分析にかけるといった処置は、そもそもどのようなメカニズムによってその反応時間が生じるかについて、全く考慮がされていないからである。また標準誤差や自由度の補正を行う場合も、データがどのように生成されるかに関心を持つならば、等分散を仮定してからパラメータの標準誤差を推定するのではなく、もともと異分散を仮定したモデルを想定すればよいはずである。たとえば、上にあげた  $t$  検定の確率モデルでいえば、群間の標準偏差が等質でないと仮定する場合には、

$$Y_{0i} \sim \text{Normal}(\mu_0, \sigma_0)$$

$$Y_{1i} \sim \text{Normal}(\mu_0 + \Delta, \sigma_1)$$

というように標準偏差  $\sigma$  を 2 つ推定すればよいだろう。しかし伝統的な手法では、このモデルでは差の標本分布はよく知られた確率分布の形にはならないため、標準誤差の補正という形で対処されることがほとんどである。

ここで、本節で論じたことをまとめておこう。統計モデリングと伝統的な心理学のデータ分析手法との違いを述べた。伝統的な心理学のデータ分析手法は、効果の有無(あるいは大きさ)から心理メカニズムを推論するような研究計画を立て、有意差を検出することで実証していく方法であるといえる。それに対して、統計モデリングは確率モデルを使ってデータ生成メカニズムを記述し、

それを用いて推論や予測を行う方法である。

#### 4. ベイズ推定による統計モデリング

1 節で述べたように、ベイズ統計学は心理学において 2 つの側面から関心を集めている。ひとつは帰無仮説検定批判からきた、それへの代案としての「ベイズ」であり、もうひとつは統計モデリングの推定方法として MCMC が有効であるという意味での「ベイズ」である。本節では、統計モデリングの文脈でなぜ MCMC によるベイズ推定が重要な役割を持つのかについて論じる。

##### 4.1 汎用的な推定アルゴリズムである MCMC

統計モデリングでは一般に、知りたいパラメータについての推定は簡単ではない。統計モデリングはデータ生成メカニズムを確率モデルで表現することに主眼があるため、どうしても  $t$  検定や分散分析のような既存のデータ分析手法に比べて、モデルが複雑になりがちである。すでに上で述べた等分散を仮定しない平均値の差を推定するためのモデルも、正確な標本分布を知ることは難しく、 $t$  分布を用いた近似もうまくいかない。よって、 $t$  検定では Welch の補正などを用いて検定を行っているという側面がある<sup>9)</sup>。一見簡単そうに見える確率モデルでもそうなのだから、実際にそれを解いて一般化線形混合モデルやさらにそれらを応用したモデルでパラメータを推定するとなると、最小二乗法や最尤法でもうまく解けないという問題がしばしば生じる(久保, 2012)。

だが、MCMC の登場によって、これらの問題は解決された。MCMC はパラメータの分布(ベイズ統計では標本分布ではなく事後分布と呼ぶ)を直接解析的に求めるのではなく、パラメータについてのたくさんの乱数値の集積によって分布を推定する方法である。この方法のよいところは、パラメータの事後分布が解析的に表現できなくても推定することが可能なことである。この長所により、数学的に高度な演算から標本分布の導出などをせずとも、パラメータの推定ができるようになった。

9) なお、異分散を仮定した平均値の差の推定については、構造方程式モデルなどを利用すれば最尤法で解くことが可能である。

また、MCMCは非常に汎用性が高い推定アルゴリズムでもある。たとえば最尤法によりパラメータを推定するときは、対数尤度関数の一次（と場合によっては二次）の導関数を求め、ニュートン - ラフソン法などの数値計算によって解く、ということが行われるのが一般的である。よって、確率モデルを作るごとに対数尤度関数について微分を行う、という数学的なスキルが必要とされる。つまり、確率モデルを新しく作るためには、同時に、パラメータを解くためのアルゴリズムの開発も必要になっていたのである。それに対し、MCMCでは確率モデルを用意すれば、あとはすべて同じアルゴリズムによって事後分布からのサンプルを生成できる。したがって、MCMCでは最尤法などに比べて、パラメータを求めるためのアルゴリズムを容易に、汎用的に自動化できるのである。

このようなMCMCの特徴から、ベイズ推定によって統計モデリングを解くことの有用性が知られるようになったのである。

#### 4.2 確率的プログラミング言語の登場

MCMCが汎用的に使える推定アルゴリズムであることを利用して、BUGS (Bayesian Inference Using Gibbs Sampling) と呼ばれるソフトウェアが開発された。またその後、Windowsで動くWinBUGSが開発された<sup>10)</sup>。WinBUGSはギブスサンプラーやメトロポリス・ヘイスティング法と呼ばれるMCMCのアルゴリズム<sup>11)</sup>を応用して確率モデルのパラメータを求めるソフトウェアである。確率モデルを書いてデータを与えるだけで、パラメータの事後分布が求められる、という画期的なものであった。また統計ソフトウェアR上からWinBUGSを操作できるR2WinBUGSの登場によって、多くのユーザーが利用可能となった。

その後、WinBUGSの開発が2007年以降停滞したこともあり、代わりにWinBUGSとほとんど同じ文法で動くJAGS (Just Another Gibbs Sampler)<sup>12)</sup>

や、統計学者Gelmanが開発したStan<sup>13)</sup>など、確率モデルを定義すればパラメータを自動的にMCMCで推定してくれるソフトウェアがいくつか開発された。またSASでもver9.2以降からproc MCMC<sup>14)</sup>が搭載され、PythonでもPyMCというライブラリが配布されている。これらMCMCを実行するソフトウェアについては、岩波データサイエンス刊行委員会(2015)でまとめて紹介されている。

MCMCの登場によって、上記のようなさまざまな汎用解析ソフトウェアが開発された。これらを特に確率的プログラミング言語と呼ぶ(松浦, 2016)。確率的プログラミング言語とは、松浦(2016)によれば「様々な確率分布の関数や尤度の計算に特化した関数が豊富に用意されており、確率モデルをデータにあてはめることを主な目的としたプログラミング言語」のことである<sup>15)</sup>。確率的プログラミング言語を使えば、確率モデルを書いてデータを渡すだけで、パラメータがほぼ自動的に推定できてしまう。よって、尤度関数を微分したり、パラメータを推定するためのコードを書いたり、といった解を求めるためのアルゴリズム構築に関する労力はなくなり、それらをすべてコンピュータが自動で行ってくれる。

WinBUGSやJAGS, Stanといった確率的プログラミング言語の登場は、数学やプログラミングがそれほど得意ではない心理学者にとって、統計モデリングを使って研究することの敷居を大きく下げることになった。それ以前では、統計モデリングに興味があっても、数学が得意な研究者が開発したモデルとそれを解くためのソフトウェアを、「一切設定を変えることなく」自分のデータに当てはめることしかできなかった。よって、自分の関心によってパラメータを減らしてみたり、当てはめる関数を変えてみたり、といった工夫を行うことができなかった。しかし、確率的プログラミング言語があれば、モデルの開発者もStanなどの言語でモデルを書き、それを見たほかの研究者

10) WinBUGSは以下のURLからダウンロードできる。  
<https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>

11) MCMCもさまざまなアルゴリズムが提案されており、その一つにギブスサンプラーなどがある。

12) 以下のURLからダウンロードできる。<http://mcmc-jags.sourceforge.net/>

13) <http://mc-stan.org/> からダウンロードできる。

14) 非商用利用であれば無償で使えるSAS University Editionでも、proc MCMCは利用できる。

15) 確率的プログラミング言語はここで取り上げたWinBUGSやJAGS, Stan以外にもたくさんある。それらの一覧は、[probabilistic-programming.org](http://probabilistic-programming.org) のWebサイトが参考になる。<http://www.probabilistic-programming.org/wiki/Home>

```

parameters{
  real mu;
  real delta;
  real<lower=0> sigma1;
  real<lower=0> sigma2;
}

model{
  Y1 ~ normal(mu,sigma1);
  Y2 ~ normal(mu+delta, sigma2);
}

```

図3 Stanによる確率モデルの表現

もそれを少し改良するだけで、研究者自身の仮説やデータに合ったモデルのパラメータを推定することもできる。

ここで確率的プログラミング言語ではどのように記述されるのかの例を挙げる。たとえば、先ほどから例に挙げている、2群の分散が異なる平均値の差を推定するためモデルを Stan コード（の一部のみを抜粋）で書くと、図3のようになる。

図3で、parametersとあるのは推定するパラメータを宣言する場所で、4つのパラメータが推定対象であることが明示されている。次にmodelとあるのは確率モデルを定義する場所である。データ Y1 と Y2 をそれぞれが標準偏差の異なる正規分布に従うことが記述できる。

ここで、近年よく使われている確率的プログラミング言語について簡単に紹介する。ここでは JAGS と Stan について、両者を比較しながら解説しよう。

JAGS は WinBUGS と同じような文法で書くことができる。JAGS の Stan に対する長所は、単純なモデルでは収束も早く、のちに述べるコードのコンパイルもいらないので、単純なモデルでは JAGS のほうが Stan よりも早く分析できる点が挙げられる。また、どちらかといえばプログラミング初心者にとってコードが書きやすく、R ユーザーなら比較的すぐに習得できる。一方で短所については、パラメータの相関が高くなりがちな階層モデルなどでは、収束するまでのサンプリング回数が多くなり、結果的に時間もかかること、使い方についての資料（特に日本語のもの）があまり多くないこと、開発者が一人なので開発スピードがそれほど速くないこと、などが挙げられる。

Stan はハミルトニアンモンテカルロ法 (Hamiltonian Monte Carlo method; 以下 HMC) とい

う MCMC のアルゴリズムを搭載しており、複雑なモデルでも少ないサンプリング回数で収束しやすいという特徴がある<sup>16)</sup>。そのため、複雑なモデルの場合は、Stan のほうが分析時間は短くなることが多い。Stan は C 言語に近い文法でモデルを記述するところに特徴があり、使う変数の宣言が必須であったり、ブロックが細かく分かれていたり、コードを記述するためのルールが厳格である。また Stan 言語は機械語にコンパイルするために数十秒程度の時間がかかるため、単純なモデルでは JAGS よりも時間がかかる。Stan のほうが、C 言語などのプログラミングに慣れたユーザーなら使いやすく可読性も高い一方、初心者には少しハードルが高いかもしれない。ただ、マニュアルも丁寧<sup>17)</sup>で、日本語の資料が豊富である。

#### 4.3 階層モデルとベイズ推定

次に、少し違った観点から、統計モデリングをベイズ推定で解くことのメリットを述べる。それは、階層モデルと呼ばれるより広範囲なモデリングがベイズ推定によって容易になった点である。

頻度主義統計学に立脚する最尤法などの方法は、パラメータについての階層性を考慮することが難しい。パラメータの階層性について説明するために、上で述べたミューラーリヤー錯視の恒常性データを例に挙げよう。実験参加者が標準刺激に対して比較刺激  $i$  が「長い」と反応する  $Y_i$  は、

$$Y_i \sim \text{Binomial}(T=8, \pi_i)$$

という二項分布で表現できることはすでに述べた。そして、長いと回答する確率  $\pi$  は

$$\pi_i = \phi(\alpha + \beta X_i)$$

という標準正規累積曲線で表現できるとした。この確率モデルは、比較刺激  $X_i$  によって比較刺激が長いと反応するかどうか確率的に変わることが仮定されているが、実は錯視量についての個人差は一切考慮されていない。これは二項分布が正規分布のように分散パラメータが平均パラメータ

16) HMC については Kruschke (2015) や 豊田 (2015) に詳しい。

17) 2016 年から Stan のマニュアルを日本語訳するプロジェクトも立ち上がり、順次訳が進められている。http://statmodeling.hatenablog.com/entry/introduction-stan-ja

と独立にあるような分布ではなく、平均パラメータ（ここでは $\pi$ ）から一意に決まる分布である<sup>18)</sup>ことが要因である。そこで、PSEに個人差があるという想定の下で、パラメータ $\alpha$ が個人 $j$ によって変化することを考える。すると、

$$Y_{ij} \sim \text{Binomial}(T=8, \pi_{ij})$$

$$\pi_{ij} = \phi(\alpha_j + \beta X_{ij})$$

という確率モデルになる。このモデルは、 $\alpha_j$ が一人ひとり違うと考えるため、個人個人のデータごとに分析をすることになる。つまり、PSEは $-\alpha_j/\beta$ というように、参加者の人数分推定される。しかし、個人ごとに別々のモデルを想定すると、推定するパラメータ数が多くなり、推定が不安定になる（あるいは収束しない）ことがある。また、個人差を表す心理変数の多くが正規分布でよく表現ができることから、 $\alpha_j$ が平均 $\mu_\alpha$ 、標準偏差 $\sigma_\alpha$ の正規分布に従うという仮定をモデルに追加する。すると確率モデルは、

$$Y_{ij} \sim \text{Binomial}(T=8, \pi_{ij})$$

$$\pi_{ij} = \phi(\alpha_j + \beta X_{ij})$$

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \sigma_\alpha)$$

となる。このとき、このモデルは、データについての確率モデルと、そのパラメータについての確率モデル、というように「階層性がある」という。また、そのようなパラメータに階層性があるモデルを階層モデルと呼ぶ。個人差を表すパラメータ $\alpha_j$ に確率分布を仮定することは、 $\alpha_j$ に緩やかな制約を与えることを意味し、ひとつに固定するよりは柔軟に、しかし個人ごとに別々に推定するよりは安定した解析が可能になる。さらに、 $\alpha_j$ の分布の特性（標準偏差 $\sigma_\alpha$ など）から、PSEの個人差がどの程度大きいかを推論することもできる。

しかし、すでに述べたように、階層モデルを最尤法で解くのは簡単ではない。それは、頻度主義統計学がパラメータを固定値と想定することから、パラメータに確率分布を仮定することが頻度主義の理論的枠組みからは困難だからである。よって、最尤法で解くためには個人差を表すパラ

メータを積分で消去した尤度関数について最適化する必要があるが、積分を解析的に評価することは一般には難しい。そこで、最尤法で解く場合はラプラス近似と呼ばれる近似法で推定するか、数値積分によって推定する方法の2つが用いられる。前者は近似が必ずうまくいくとは限らないこと、後者は正確な推定は可能だが単純な階層モデルでしか適用できないこと、などの限界がある。

このような困難も、ベイズ統計学とMCMCを用いれば解決する。ベイズ統計学は、すでに述べたように、パラメータが確率的に変動することを仮定している。よって、階層モデルは特殊なモデルというよりは、むしろベイズ統計学では自然なモデルの拡張としてとらえることができる。ベイズ統計学では、階層モデルでなくとも各パラメータに事前分布という確率分布が仮定される。多くの場合は事前分布には無情報分布が用いられるが、階層モデルではパラメータの事前分布が無情報分布ではなく、それ自体が確率モデルの一種として記述されることになる。そういう意味で、ベイズ統計学では確率モデルと事前分布の区別は明確に分けられるものではない。そして、だからこそ階層モデルが自然に構築可能なのである。そして、階層モデルであっても確率モデルの一種であるので、MCMCによるパラメータの推定が容易になる。

階層モデルの使いどころは、心理学者にとっての関心があるものに絞れば、大きく分けて2つある。ひとつは二項分布やポアソン分布のように、個人差が仮定されない確率分布を使うときに、個人差をモデリングするために用いる場合である。もうひとつは、階層線形モデルなどのマルチレベル分析と同様、データに階層性があり、切片や回帰係数の集団間変動を考慮したい場合である。たとえば、切片の集団間変動を仮定した階層線形モデルは、

$$Y_{ij} \sim \text{Normal}(\mu_{ij}, \sigma)$$

$$\mu_{ij} = \alpha_j + \beta X_{ij}$$

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \sigma_\alpha)$$

となる。個人が $i$ 、集団が $j$ で表されており、切片である $\alpha_j$ が集団ごとに異なることが示されており、またそれが正規分布に従うことがわかる。このように、データが個人と集団というネストさ

18) 二項分布の分散は $T\pi(1-\pi)$ で求められる。このときの $T$ は試行回数であり、この例では $T=8$ である。

れた構造を持っている場合、確率モデルも同様に階層モデルを使うほうが自然である。また、因子分析や項目反応理論など、個人が複数の項目に回答した場合に潜在得点を推定したい場合なども、一種の階層モデルとしてみなすことができる。あるいは、反復測定などの参加者効果を含んだ線形モデルも同様である。

心理学者の関心に直結しない部分でも、階層モデルの使いどころはたくさんある。ここでは簡単に紹介するにとどめるが、状態空間モデルなどの時系列解析、機械学習で用いられるパラメータの平滑化、欠測を含むデータ分析などにも階層モデルが利用されている。これらについては伊庭(2018)にまとめられている。

#### 4.4 最尤推定とベイズ推定の違い

なぜ統計モデリングをベイズで推定することがよいかを明確にするために、階層モデルを解く以外の、最尤推定に対するベイズ推定(ただしMCMCを使った場合)の利点を述べる。最尤推定とベイズ推定には大きく分けて2つの違いがある。ひとつめは、最尤推定がパラメータの事前分布を想定しない(あえていえば、範囲が無限の一樣分布を仮定している)が、ベイズ推定では事前分布を想定する点にある。もうひとつの違いは、最尤推定のアルゴリズムが基本的に点推定値(尤度を最大にする点)を探索する方法であるのに対し、MCMCは分布全体を推定する方法である、という点である。

事前分布を仮定できることのメリットは、過学習(overfitting)を防ぎやすい点にある(松浦, 2016)。過学習とは、統計的推測を手元のデータに合わせすぎてしまうことを指す。最尤推定ではパラメータの推定を「データのみから」行うことしかできない一方、ベイズ推定では、先行研究の成果を事前分布として利用することで、汎化性能を維持した推定ができる。汎化性能とは、次に新しく取得したデータに対しても予測力が高いことを意味している。データは、ある種の客観性を持つが、同時にサンプリングバイアスを内在している。よって手元の限られたデータにモデルを合わせすぎるのは、本来知りたい現象の理解から遠ざかる可能性があることを認識する必要がある。

パラメータの事前分布を想定することは、ベ

イズ推定がもつ弱点として考えられることが多かった。それは、パラメータについて研究者が恣意的に事前分布を設定できてしまい、分析結果の客観性が維持できないという考えがあったからである。しかし、渡辺(2012)も主張しているように、確率モデルではそもそも尤度関数の選択も研究者が行うものであり、恣意性がある。よって、事前分布の設定のみが恣意的であり、尤度関数の選択は客観的である、ということはいえない。確率モデルは「尤度関数と事前分布のセット」を研究者が選択していると考えるのが妥当であろう。

点推定値の探索ではなく、分布全体を推定することのメリットは、パラメータの区間推定などが容易な点、事後分布が多峰分布になる場合にも妥当な推論ができる点、にある。前者の点については、パラメータの区間推定に関わる。点推定値自体は推定が簡単でも、パラメータの分布が正規分布にならないような場合、最尤法では区間推定や標準誤差の推定が困難な場合がある。たとえば媒介分析で用いる間接効果の推定では、点推定値は2つのパス係数の積だが、その標本分布は一般には非対称になり、その形状を正確に知るのは、最尤推定では簡単ではない<sup>19)</sup>。しかし、MCMCによるベイズ推定では、非対称な分布の場合でも、特にモデルの補正や推定アルゴリズムの拡張を必要とせずに推定できる。後者の点については、混合分布モデルのように、確率モデルが複数の確率分布の和で表現される場合に、パラメータの分布が多峰になる場合に関わる。そのようなときには、最尤推定のように点推定を探索するアルゴリズムを用いる方法(たとえばEMアルゴリズム)では、初期値によって推定値が変わってしまうという、初期値依存性が生じ、一意に解を得るのが難しい。こうした複雑な確率モデルでも、MCMCによるベイズ推定を用いることで適切な事後分布が得られることがある。

## 5. 心理学における ベイズ統計モデリングの可能性

本節では、ベイズ統計モデリングが心理学の研

19) 大標本の仮定などを置いて、漸近正規性の性能に頼らざるを得ないことも多い。あるいはブートストラップ法などが利用される。

究でどのように活用できるかについて、いくつか具体例を挙げながら述べる。

心理学でベイズ統計モデリングが活用できる点として、以下の3つの可能性を挙げる。ひとつめは、行動データの生起メカニズムを確率モデルで記述することで、理論の精緻化と行動の予測に役立つ可能性である。もうひとつは、心理的な特徴を測定・推定するときに、より理論的に妥当になる可能性である。そして最後に、確率的プログラミング言語によって確率モデルを記述することによって、より透明性や再生可能性の高いモデル構築が可能になる点を挙げる。また、上記3つ以外にも、本特集論文の竹澤（2018）が心理学におけるモデリングの必要性を論じている。

### 5.1 理論に基づいた行動生起メカニズムを確率的に表現できる

先述のように、統計モデリングの一番の特徴は、データ生成メカニズムを確率モデルで記述し、そこから推論や予測をする点にある。このことから心理学においては、行動の生起メカニズムを確率的に表現し、行動の予測を行えるようになるという点が、そのまま利点となりえる。

心理学で扱う行動データは、そのほとんどが測定誤差および個人差が大きく、行動生起メカニズムについて確定的な関数によって説明することはほとんど不可能である。測定誤差や個人差といった変動は確率的な法則にしたがうと仮定して、確率モデルによって行動生起メカニズムを記述することができれば、行動生起による偶然的な変動と、心理学的に意味のある構成概念同士の確定的な関係を切り分けて表現できる。このことは、心理学理論を検証可能で、かつ、比較可能なものとして構築できるという意味で有効である。

この点について、Lee and Wagenmakers (2013) で紹介されている、Ebbinghaus (1885/1974) の忘却曲線を応用した、記憶保持のモデルを例に説明する。いくつかの刺激（たとえば英単語など）を覚えさせたあと、再認課題を行うとする。このとき、再認課題を行うまでの時間によって正答率が指数的に減衰する変化を、確率モデルで表現することを考える。ある実験参加者が  $j$  時間後の再認課題（課題は  $n$  回）で、正解する場合に  $k=1$ 、不正解の場合に  $k=0$  とすると、

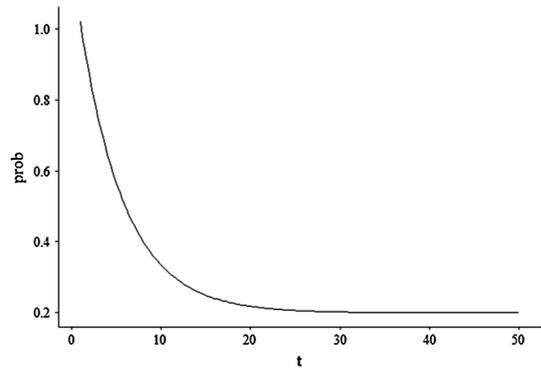


図4 忘却曲線のモデリング

$$k_j \sim \text{Binomial}(n, \theta_j)$$

$$\theta_j = \exp(-\alpha t_j) + \beta$$

というモデルを想定できる<sup>20</sup>。ここで、 $\theta_j$  は時間  $j$  のときの正答率で、これが減衰率  $\alpha$  と記憶のベースライン（時間がどれだけたっても最低限はここまでは記憶できる程度）を  $\beta$  とする。なお、 $\alpha$  と  $\beta$  はそれぞれ  $0 \sim 1$  の範囲を取るものとする。このとき、正答率  $\theta$  と時間  $t$  の関係は図4のような指数的な関係になる。

上のような忘却曲線のモデルでは、まず正解するか否かという行動指標が確率的に変動しつつ、記憶の定着の程度を表すパラメータ（心理的な特徴）である  $\theta$  と関係するとしている。すなわち、行動  $Y$  と心理的概念  $\theta$  が確率的関係性 ( $\sim$ ) で結ばれている。次に、その心理的な特徴  $\theta$  が、時間と記憶についての能力（減衰率  $\alpha$  とベースライン  $\beta$ ）と確定的な関係性 (=) で結ばれている。このように、行動の偶然性を確率分布で表現しながら、心理的構成概念についての法則性を数理的に表現できる。

しかし、実際このモデルは個人ごとの記憶成績のデータにはフィットしない。それは、記憶力についての個人差が考慮に入られていないからである。よって、このモデルは次のように  $\alpha$  と  $\beta$  に個人差が加えられて修正される。つまり、それぞれの個人差が正規分布に従い、ある一定の散らばりになるという仮定を置いて、

20) 実際は  $\theta$  が 1 を超えないよう、 $\min(1, \exp(-\alpha t) + \beta)$  が想定されているが、ここでは見やすさを重視して省略した。

$$\alpha_i \sim \text{Normal}(\mu_\alpha, \sigma_\alpha)$$

$$\beta_i \sim \text{Normal}(\mu_\beta, \sigma_\beta)$$

となる。最初のモデルから比較して、記憶力についての個人差の大きさを表す  $\sigma_\alpha$  と  $\sigma_\beta$  のパラメータが増えることで、よりデータにフィットしたモデルを作ることができる。モデルの優劣をどのように決めるかについては難しい問題であるが、たとえば情報量規準である AIC や WAIC (Watanabe, 2010), あるいは交差妥当化といった予測の観点から評価することが可能である。またそれ以外にも、ベイズファクターを用いたモデル評価なども近年は利用可能になりつつある。ベイズファクターについては岡田 (2018) に詳しい。

また、個人差についての散らばりも考慮に入れた忘却曲線のモデリングによって、新しく記憶課題に取り組んだある人が、 $t$  日後の課題でいくつ正解するかについての確率的な予測ができるようになる。データについての予測を確率的に行うには、確率モデルと推定されたパラメータを用いて計算される予測分布というものを使う。予測分布を使えば、95% ベイズ予測区間から、たとえば 10 日後の再認課題で 20 個中 4 個～6 個の範囲で正答する、というように予測できるのである。

さらに、今回は単純な指数関数を用いたが、記憶のメカニズムから考えてさらに妥当な関数を考えることもできるだろう。たとえば Averell and Heathcote (2011) では指数関数モデル以外に、パレート関数モデルとべき関数モデルを比較している。

統計モデリングは、心理学理論が想定するものをトップダウン的に記述するためだけではなく、逆に、その理論よりももっと単純な規則によって人が行動している可能性を検証できる。たとえば心理学的には複雑な行動生起メカニズム (ヒト特有の心理的なメカニズム) が想定されて検証されてきたものが、実は非常に単純な規則 (動物一般にも適用できるメカニズム) によっても行動が同様に予測できてしまうかもしれない。複雑なモデルは表現力が多様になる一方、パラメータも多くなることで、予測能力が逆に悪くなることもある。少ないパラメータのモデルでも同程度の予測力があるならば、理論の前提となる想定的大幅な修正が要請されるかもしれない。

## 5.2 心理的特徴について理論に妥当な推定ができる

次の利点として、統計モデリングを使うことで、より理論的に妥当な心理的な特徴を測定できるようになることが挙げられる。行動の生起メカニズムにある関数と心理的な特徴によって表現することで、理論的仮定を踏まえながら心理測定を行える。

たとえば、記憶の課題などで使われる信号検出理論を挙げる。信号検出理論 (Tanner & Swets, 1954) では、信号の検出方法について 4 つの反応パターン (ヒットとミス、誤警報と正棄却) を分け、それらを確率モデルで表現する。Lee and Wagenmakers (2013) では、信号検出理論について次のような確率モデルが紹介されている：

$$h \sim \text{Binomia}(\theta^h, S)$$

$$f \sim \text{Binomia}(\theta^f, N)$$

$$\theta^h = \Phi\left(\frac{1}{2}d - c\right)$$

$$\theta^f = \Phi\left(-\frac{1}{2}d - c\right)$$

$h$  はヒットの数、 $f$  は誤警報の数である。これらの行動指標が、ヒット率  $\theta^h$  と誤警報率  $\theta^f$ 、そして信号ありの試行数  $S$ 、信号なしの試行数  $N$  をパラメータに持つ二項分布に従うと仮定している。ヒット率  $\theta^h$  や誤警報率  $\theta^f$  は、信号検出理論の理論的仮定から、信号検出力  $d$  と反応バイアス  $c$  という 2 つのパラメータの確定的な関係性 ( $\Phi$  は標準正規累積関数) によって表現できると予想される。

さらに、信号検出力  $d$  や反応バイアス  $c$  について、確率分布を仮定することで、個人差を階層モデルで表現できる。そうすれば、人間一般の行動傾向を表すパラメータ (たとえば信号検出力の平均や標準偏差) も知ることができる。つまり、 $d$  と  $c$  に事前分布を仮定し<sup>21)</sup>、

$$d_i \sim \text{Normal}(\mu_d, \sigma_d)$$

$$c_i \sim \text{Normal}(\mu_c, \sigma_c)$$

とモデルを拡張できる。

21) この場合、 $d$  と  $c$  に相関を仮定して、二変量正規分布に従うという想定も可能である。そうすれば、 $d$  と  $c$  の相関もデータから推定できる。

このように統計モデリングを使うことで、信号検出や記憶課題における成績を、信号検出理論が想定するモデルに基づいてストレートに推定できるのである。

次に、潜在連合テスト (Implicit Association Test, 以下 IAT) を例に挙げる。IAT は、Greenwald, McGhee, and Schwartz (1998) で提案された、人がもつ認知的表象のネットワークにおける特定の2つの表象と属性同士の連合の強さを推定するものである。この方法により、潜在的な態度やステレオタイプなどを測定できるとされている。IAT は二肢選択分類課題において、たとえばステレオタイプに一致する条件の反応時間と、一致しないときの反応時間の差から、潜在的なステレオタイプが測定される。IAT 効果の算出は、Greenwald, Nosek, and Banaji (2003) が提案した D スコアが用いられることが多いが、その推定方法はアドホックであり、理論的な仮定に基づくものではない。そこで、Klauer et al. (2007) は二肢選択分類課題の反応時間についての確率モデルである diffusion model (Ratcliff, 1978) を用いて、IAT 効果を3つのパラメータ (注意深さ、情報収集率、非決定時間) によって推定することを試みた<sup>22)</sup>。その結果、顕在的指標との相関などから、情報収集率についてのパラメータが IAT 効果の指標として妥当性があり、また D スコアに対しても理論的な意味でアドバンテージがあるとしている。ここでも認知的なメカニズムとの対応を考慮したうえで、反応時間から精緻に心理的特徴を推定することが有効であることが示されている。なお、IAT と diffusion model については、土居・川西 (2016) にまとめられている。

心理尺度を用いたパーソナリティ測定の場合でも、統計モデリングによって、よりバイアスを含まない心理的特徴の推定ができる可能性がある。Bolt, Lu, and Kim (2014) は King et al. (2004) で提案された係留寸描法 (anchoring vignettes method; 以下 AVM) を用いた項目反応理論を提案した。AVM は複数人の仮想人物について描写した文章

を提示し、その仮想人物についての特徴をリッカート尺度で回答させる。続いてそれと同じ項目について回答者自身の特徴についても回答させることで、リッカート尺度における反応バイアスを推定することが可能である。ここでいう反応バイアスとは、実際に測定したい回答者の心理的特徴とは無関係に、リッカート尺度の midpoint に反応しやすい傾向 (midpoint バイアス) や、「まったく当てはまらない」や「非常に当てはまる」といった極端な反応ラベルに反応しやすい傾向 (極値バイアス)、などのことである。Bolt らのモデルでは、AVM で得たデータから、反応バイアスを確率モデルのパラメータとしてベイズ推定できる。北條・岡田 (2017) は、AVM によって得たデータに対して、反応バイアスを推定した場合とそうでない場合でのモデル比較を行った。その結果、反応バイアスを推定した方法のほうが予測誤差の観点でよいモデルであることを確かめた。

また、上で挙げたような理論的妥当性だけではなく、推定の区間推定ができる点も、ベイズ統計モデリングの利点である。MCMC によるベイズ推定は、得点の事後分布がそのまま得られるため、個人の心理的特徴を点推定ではなく、ベイズ信頼区間などを用いた区間推定によって評価することも容易である。

### 5.3 分析の透明性や再生可能性が高くなる

ベイズ統計モデリングを心理学に生かすことの利点として、Stan などの確率的プログラミング言語によってモデルを記述することによって得られるメリットについて述べる。

確率的プログラミング言語は、確率モデルを記述するだけで、あとは MCMC で自動的にパラメータを推定できることはすでに述べた。この利点は、分析するものにとっての負担が大きく減るだけでなく、モデルを書いたプログラムのコードの共有が簡単であることも意味する。Stan などの確率的プログラミング言語は、その文法さえ理解すれば、数学やプログラミングについての高度な知識を必要とせず、その意味を理解することが可能である。それは推定したいモデルと推定のためのアルゴリズムが、完全に分離されていることによって実現されている。

このことから、確率的プログラミング言語は、

22) 実際には diffusion モデルでは7つのパラメータが推定できるが、そのうち心理学的な解釈がしやすい3つについて取り上げている。また、diffusion model は多くの場合最尤法で解かれることが多いが、ベイズ推定を行うための R パッケージである hBayesDM (Ahn, Haines, & Zhang, 2017) も開発されている。

データ分析についての透明性を高めてくれるといえる。具体的には、Stanなどで記述されたコードを論文に添付することによって、論文で示された確率モデルが正確にコードに表現されているかについて、査読者や読者が確認しやすくなる。もちろん、統計ソフトウェアRもオープンソースであることから、パッケージに含まれている統計分析のコードが正しく機能しているかをすべてのユーザーが確認できる。しかし、多くの場合、どの部分がモデルに関わるコードで、どの部分が推定アルゴリズムに関係するコードであるかを正確に判断するためには、統計学やプログラミングの知識が不可欠である。同じオープンソースであっても（数学があまり得意とはいえない心理学者にとっては）、確率的プログラミング言語が持つ透明性は相対的に高いといえる。

また、確率的プログラミング言語で記述されたモデルとデータが共有されていれば、査読者や読者は容易に同じ分析結果を得られる。ここで、同じデータに対して分析を行ったときに、同じ結果が得られることを、国里（2017）に倣い「再生可能性（replicability）」と呼ぶことにする。分析結果の再生可能性を最大限に高めるためには、データの読み込み、データのハンドリング、解析、解析結果の出力、文書化などすべてがコード化され、できるだけヒューマンエラーが入り込まないようにすることが重要であるとされている。StanやJAGS、そしてそれらのインターフェイスのソフトウェアであるRやPythonなどはフリーソフトウェアであり、どの研究者も自由に無償で使うことができる。また、これらのソフトウェアはCUI（character user interface）と呼ばれる、コードだけで分析の指定すべてを行えるソフトウェアである。よって、確率モデルのコード（とそれを実行するためのRコード）があれば、同じ結果が再生されるかを確認できる。一方、たとえばGUI（graphical user interface）タイプのソフトウェアでは、分析の細かなオプションなどを共有するのは煩雑である。またヒューマンエラーを防ぐことが難しいことなどによって、分析結果の再生可能性は損なわれやすくなってしまふ。

近年、心理学における再現性問題を踏まえて、事前審査制度を導入する雑誌が増えつつある（池田・平石，2016；三浦，2018）。事前審査では、

データを取得する前の仮説と分析戦略について記述することが求められるが、その段階でデータ分析のためのコード、理想的にはパラメタリカバリシミュレーション<sup>23)</sup>などを行った結果なども報告しておけば、よりスムーズに他の研究者がモデルの検証を行える。

このように、確率的プログラミング言語を用いたベイズ統計モデリングは、分析結果についての透明性や再生可能性を高めるのに機能すると考えられる。

## 6. ベイズ統計モデリングの今後の展望

本論文をここまで読んだ読者には、「もう効果の有無だけを知る目的の方法論は時代遅れなのか」と思われた読者もいるかもしれない。あるいは逆に、ベイズ統計モデリングはまだまだ敷居が高く、「自分の研究には無関係である」という感想を抱いたかもしれない。そこで本節では、上記のような反応を想定したうえで、著者の考えを述べる。具体的には、伝統的な心理学の手法との共存はどうなるのか、そしてベイズ統計モデリングを実践・評価するために何が必要か、である。また最後に、ベイズ統計モデリングについての注意点などについて触れ、今後の展望を論じる。

### 6.1 伝統的なデータ分析法との共存

これまで、本論文ではあたかも「これからの心理学はベイズ統計モデリングの時代だ！」といわんばかりに、心理学におけるベイズ統計モデリングの利点を述べてきた。しかし、著者自身も、心理学におけるデータ分析法がすべてベイズ統計モデリングにとってかわられるとは、全く思っていない。それは、伝統的なデータ分析手法が、心理学の研究パラダイムとセットになって発展してきており、容易にそれらを変容させることが難しいこと、またなにより従来のデータ分析手法そのものにも、他に代えがたい利点があるからである。

23) パラメタリカバリシミュレーションとは、ある確率モデルとあらかじめ決めておいたパラメータから乱数によってデータを生成し、そのデータを想定した確率モデルによって推定することで、設定したパラメータ通りの値を得られるかを確認する手続きのことである。

心理学のデータ分析手法では、すでに述べたように、要因の効果の有無（や大きさ）を知るのが第一の目的となっている。これは、心理学がこれまで培ってきた研究パラダイムが、実験+分散分析を前提に組み立てられてきたからである。実験によって要因の効果を知ることは、統計的因果推論の枠組みから理解すればわかるように、因果関係を知るうえで非常に有益な方法である。心理学の測定方法がどれほど正確になって、理論がどれほど精緻に検証できるようになったとしても、実験による因果推論が完全になくなることはないであろう。

その中で、ベイズ統計モデリングがすぐに心理学で利用されると考えられるのは、主に心理的特徴の妥当な推定という枠組みである。すなわち、理論的基盤を持った心理測定のツールとしての利用である。すでに例で挙げたIATを統計モデリングで推定する方法なども、認知メカニズムを考慮に入れた、より理論的に妥当なIAT効果の推定を可能にしていた。そして、これらの方法で推定された心理的特徴について、群間に差があるか否かは伝統的な心理学の枠組みで検討することになる。つまり、洗練された心理測定ツールとして統計モデリングを活用しながら、因果推論のために実験+分散分析を利用できるのである。その意味で、伝統的な心理学の方法論である「要因の効果の有無（や大きさ）を知る」ことは、ベイズ統計モデリングと共存可能であり、かつ、さらに洗練されたものになりえる。

またベイズ統計モデリングで仮定される確率モデルは、当然のことながら、伝統的な心理学の方法論によって実施された実験から見出された知見によって構築されることも多い。ある特定の操作によって行動に違いが見られたことから、これまでの確率モデルには足りなかった別のパラメータの存在が示唆される、などである。その後、新しいパラメータを含めた確率モデルをデータから検証するといったように、実験による効果の検証と確率モデルによる推定は、互いに貢献しあえるだろう。

ただし、一部に見られる、「やみくもに正規分布を仮定した線形モデルにデータを当てはめて、有意性だけを確認するような分析手法」は、ベイズ統計モデリングの登場で影を潜めていかもし

れない。一般化線形混合モデルを使えば、より行動指標に合った確率分布の選択、より理論的に妥当であると想定できるリンク関数の選択、そして個人差や集団差の推定などを行える。正規分布を仮定した推論しかできないという時代はすでに終わり、より行動データがどのような性質を持ち、どのようなメカニズムでそれが表れてきたのかを考えることができる時代になったのである。既存のデータ分析手法が正規性、線形性、帰無仮説検定などを要請することによって制限されてきたリサーチクエスチョンや実験計画・調査計画の立て方は、格段に自由になりつつあるといえるだろう。まさにCarlin and Chib (1995) が述べたような「どんなモデルを利用するかどうかは、ユーザーの想像力によってのみ制限される」が現実的になろうとしている。逆に言えば、研究者側にその自由に答えるための知識や洞察力、数学的な素養などが求められていくのかもしれない。また、久保 (2012) が批判する「ブラックボックス統計学」、つまりデータや確率分布の性質を理解しないままにソフトウェアにデータを入れて有意差のみを検出する方法、に対する批判はこれから強くなっていくかもしれない。

ここから、さらに考察を進めて、今後起きうるかもしれない、心理学と統計モデリングの「折り合いの悪さ」についても触れておこう。もし、伝統的な心理学のデータ分析手法とベイズ統計モデリングが、ある種の対立を生むとすれば、それは説明のための科学なのか、あるいは予測のための科学なのか、という立場の違いによるものかもしれない。伝統的な心理学のデータ分析手法は、行動の「予測」よりは、行動の「説明」を行うためのものが多い。一方、統計モデリングが力を発揮するのは、解釈可能な行動についてのモデルを作りながらも、そこからさらに行動の予測をも視野に入れているときである。たとえば実験室で反応時間を測定するとき、それは群間の平均的な反応時間の差が見たいのであって、個人の反応時間そのものを予測したいわけではない。一方で、統計モデリングは予測分布を作り行動を予測できるため、マーケティング分野における購買行動や、政治科学における選挙の得票数といった、他の社会科学の分野でよりその強みを発揮できることは容易に想像できる。しかし、心理学ではそのような

予測の観点からの利用が未だ想定しにくい、というのが現状かもしれない。このことは、心理学にとって統計モデリングが不要に映りうるひとつの要因となる。しかし、本論文では、統計モデリングの流行が、心理学においても行動の予測に関心を持つよい機会でありうる、という点を強調しておきたい。実際、心理学はどちらかといえば「状況に依存しない、一般的な行動傾向を知るための学問」というスタンスをとってきた（と著者には見えている）。よって、特定の状況の特定の行動だけを予測することには、あまり関心を示さない心理学者が多いように思う。そのなかで、統計モデリングが、むしろ状況を絞り込み、現実で起こりうる行動を心理学理論で予測していくという研究の方向性を開いていく可能性があるのではないかと思われる。言うまでもなく、現実社会における特定の行動を確率的にでも予測できることは、有用な側面があるはずである。今回は詳しくは触れなかったが、統計モデリングのなかには状態空間モデルやガウス過程回帰といった、より予測を重視したモデルも含まれている。これらのモデルを心理学者が利用することで、より実践的な心理学の知見を提供できる可能性がある。

## 6.2 バイズ統計モデリングを実践するために

すでに述べたように、今すぐすべての心理学者が統計モデリングを使えなければならない、ということはない。伝統的な心理学的方法論はこれからも活用されていくことだろう。しかし、統計モデリングを用いた論文がますます増えている事実を前に、われわれ心理学者はどのように対応していけばいいだろうか。

これからバイズ統計モデリングを利用して研究に役立てたい、という研究者がどのような知識を持っておく必要があるかについて、筆者の考えを簡単に述べる。バイズ統計モデリングを実践するのに必要なことは、1. バイズ推定についての知識、2. 確率分布や関数形についての理解、3. 確率的プログラミング言語の知識、4. データ生成メカニズムを想定するための発想力、などが最低限求められるだろう。

バイズ統計モデリングを使うにあたり、バイズ推定の知識は必須である。それは結果を理解するとき、論文に書くときに、最低限の知識が求めら

れるからである。ただし、単にバイズ統計モデリングを使う、ということであれば数学的な理解（MCMCの原理や共役事前分布などの理解）はそれほど深くなくてよい。MCMCが事後分布から乱数を発生させることで推定を行うアルゴリズムであることを理解したうえで、MCMCによる推定の結果をどのように処理すればよいかなど、実践的な知識でまずは十分である。これらについてのバイズ推定についての知識は、近年では日本語でも多くの書籍から得ることができる。Kruschke (2015) の翻訳である前田・小杉 (2017) の第二部、あるいは豊田 (2015) や奥村・牧山・瓜生 (2018) など、バイズ統計学の入門からMCMCによるバイズ推定について、わかりやすくまとめられている。また、統計モデリングの観点からみたMCMCの解説は、久保 (2012) にもまとまっている。

バイズ統計モデリングが確率モデルを利用する以上、確率分布についての知識もある程度は必要である。まずは離散分布と連続分布の区別、分布から生成されるデータの範囲などは最低限知っておく必要がある。加えて、それらの確率分布がどのような物理現象を表現するために作られたのか、心理学の行動との関係は何かなどについて理解していれば、確率モデルを自ら構築するのに役立つ。書籍では、豊田 (2015) に加えて、豊田 (2017) が事例も含めて解説されていてわかりやすい。また、松浦 (2016) でも多くの確率分布が利用されており、またその実践方法がStanのコードとともに解説されている。

また、確率分布そのものの理解とは異なるが、心理学現象を説明するうえでどのような関数形が利用されているかについての知識があるとよい。回帰分析では単純な線形結合のみが用いられるが、それ以外にも一般化線形モデルでは、リンク関数と呼ばれる、線形モデルを多様な非線形モデルに変換するための関数が利用される。ロジットリンク関数、対数リンク関数、正規累積リンク関数などがそれである。これらは多くの場合、確率分布とセットで用いられるが、心理現象によりフィットした関数形を想定できるほうが理想的である。これらについてはLee and Wagenmakers (2013) にさまざまな事例がある。応用的には、ソフトマックス関数や、差分方程式あるいは微分

方程式を用いた関数形も利用することができるだろう。

確率的プログラミング言語については、WinBUGSやそれをもとに開発されたJAGS、あるいはそれらとは少し違う言語をもつStanなどが使いやすい。WinBUGSやJAGSは久保(2012)をはじめ、Lee and Wagenmakers(2013)やKruschke(2015)に多くのモデルについての例が掲載されている。Stanについては松浦(2016)がもっともわかりやすくまとまっている。Stanは日本のデータサイエンス業界や心理学でもユーザーが多く、Webにも多くの情報がある<sup>24)</sup>。

最後に、データ生成メカニズムを想定するための発想力についてである。心理学は伝統的に、群間の平均的な差を検出することによって心理メカニズムを推論してきたため、心理学者の多くはこの枠組みで研究計画を考えるようにトレーニングされている。もしかすれば、そのトレーニングを受けた年月が長いほど、この枠組みから抜け出るのが難しいかもしれない。そのような場合には、統計モデリングを柔軟に吸収した若手の研究者や、数理モデルの構築が得意な研究者とのコラボレーションが有効であると考えられる。

またすでに述べたように、統計モデリングの中でも比較的馴染みやすいのは、心理的な特徴を統計モデリングで推定し、推定されたものを従来の方法で分析するというアプローチである。それらには、すでに紹介した信号検出理論を用いた記憶成績の推定、認知モデリングを利用したIAT効果の推定、それ以外にも項目反応理論を応用した態度推定などがある。これまでは実装が難しかったが故に簡便的な指標化しかできなかった心理的特徴を、本来の理論の想定に基づいた確率モデルによって推定することは、統計モデリングに慣れていない研究者にもその強みが理解しやすく、またすぐに利用可能である。このアプローチから、統計モデリングのエッセンスを理解していきながら、利用範囲を広めていくことが有効だろう。

### 6.3 ベイズ統計モデリングについての注意点と展望

最後に、ベイズ統計モデリングの注意点に触れながら、今後の展望を述べる。

ベイズ統計モデリングは、モデル構築部分とアルゴリズム構築部分を分離することで、自由なモデリングが可能になった。しかし、このようなモデリングの自由度の高さは、1990年代に構造方程式モデリング(structure equations modeling; 以下SEM)が登場してきたときにも、同じような謳い文句で「利点」として売り出されていた。現在でもSEMは利用されているが、当時の盛り上がりからすれば、それを利用した論文の数は思ったほど増えていないのかもしれない。さらにいえば、実験心理学者からはやや敬遠されている側面もあるかもしれない。ベイズ統計モデリングも同じような結末に至るだろうか。もちろんその可能性は否定できないが、SEMを反面教師にしながら、ベイズ統計モデリングを実施する上での注意点を挙げる。

ひとつは、SEMを用いるときに、適合度を上げるため「だけ」のモデル選択がしばしば行われていたことが、ともすればSEMが不毛なデータ分析手法であると研究者に映った原因のひとつかもしれない。もちろん、統計モデリングでもモデル選択はとても重要な問題である。実際に、ベイズ推定でも簡単に計算できる情報量規準である、WAIC(Watanabe, 2010)が提案され、活発に利用されている。しかし、ベイズ統計モデリングでも、SEMで見られたようなWAICをただ小さくするためだけのモデル選択がもし行われるとするならば、やはりそれは不毛な研究実践であるといえるだろう。本来モデル選択は、小さいサンプルを一度とっただけでできるほど簡単なものではない。情報量規準は、平均的にはよいモデルを選択できるとはいえ、サンプリングバイアスの影響は一般の心理学者が想定するよりも大きい。モデル選択を統計指標によって安易にするよりも、まずモデルがデータ生成メカニズムを理論的に表現できているのか、また事後予測チェックなどを用いて、データと確率モデルがフィットしているのかを視覚的にも確認していきながら、複数の研究で慎重にモデルの選択を行っていくのが望ましいと考えられる。

24) たとえば、Stan advent calendar ではStan コードで書かれたさまざまな統計モデルの事例を閲覧できる。<https://qiita.com/advent-calendar/2017/stan>

SEMに比べて、ベイズ統計モデリングはさまざまな長所を持っている。多変量正規性を仮定しなくてもよいこと、線形モデル以外のモデルを構築可能なこと、階層モデルを容易に構築できること、などがある。これらの長所は、確かにSEMに対して持たれていた一部の批判（たとえば杉万, 2005）を乗り越えてはいる。しかし、結局のところ方法論のよさを引き出すのは研究者自身であって、研究実践をよりよいものにしていくためには、研究者間での「方法論の使い方についての合意形成」が重要になっていくと思われる。

方法論が自由になればなるほど、研究計画を立てる研究者側に求められるものは大きくなっていくのかもしれない。しかし、それぞれの研究者が持つ長所を持ちよりながら、よりよい研究実践に対する合意を形成することが、心理学をよりよいものに発展させていくはずである。

## 文 献

- Adams, R. A., Huys, Q. J. M., & Roiser, J. P. (2015). Computational psychiatry: Towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery, and Psychiatry*, *87*, 53–63.
- Ahn, W. Y., Haines, N., & Zhang, L. (2017). Revealing neuro-computational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Computational Psychiatry*, *1*, 24–57.
- Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, *55*, 25–35.
- Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425.
- Bolt, D. M., Lu, Y., & Kim, J. S. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods*, *19*, 528–541.
- Carlin, B. P., & Chib, S. (1995). Bayesian model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 473–484.
- 土居淳子・川西千弘 (2016) 拡散モデルに基づく潜在的連合テストデータの分析 京都光華女子大学京都光華女子短期大学部研究紀要, *54*, 31–42.
- Ebbinghaus, H. (1885/1974). *Memory: A Contribution to Experimental Psychology*. New York: Dover.
- Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approach to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398–409.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.
- 北條大樹・岡田謙介 (2017) 評定尺度における反応傾向を考慮した係留寸描データのベイズ的項目反応モデル データ分析の理論と応用, *6*, 1–13.
- 伊庭幸人 (2018) ベイズモデリングの世界 岩波書店.
- 池田功毅・平石 界 (2016) 心理学における再現可能性危機：問題の構造、現状と解決策 心理学評論, *59*, 1–34.
- 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.
- 岩波データサイエンス刊行委員会 (編) (2016) 岩波データサイエンス Vol.3 岩波書店.
- King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*, 191–207.
- Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology*, *93*, 353–368.
- Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A tutorial with R, JAGS, and Stan*. (2nd ed). London, UK: Academic Press. (前田和寛・小杉考司 (監訳) (2017) ベイズ統計モデリング: R, JAGS, Stanによるチュートリアル 共立出版.)
- 久保拓弥 (2012) データ解析のための統計モデリング入門：一般化線形モデル・階層ベイズモデル・MCMC 岩波書店.
- 国里愛彦 (2017) 統計解析の再現可能性を高める取り組み 臨床疫学研究における報告の質向上のための統計学の研究会, 発表資料 [Web page] <https://www.slideshare.net/YoshihikoKunisato/ss-77835559>
- 国里愛彦 (2018) 臨床心理学と認知モデリング 心理学評論, *61*, 55–66.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press. (井関龍太 (訳) (2017) ベイズ統計で実践モデリング：認知モデルのトレーニング 北大路書房.)
- Marr, D. (1982). Vision: a computational investigation into the human representation and processing of visual information. New York: W. H. Freeman and Company. (乾 敏郎・安藤広志 (訳) (1987) ビジョン：視覚

- の計算理論と脳内表現 産業図書.)
- 松浦健太郎 (2016) Stan と R でベイズ統計モデリング 共立出版.
- 三浦麻子 (2017) なるほど! 心理学研究法 北大路書房.
- 三浦麻子 (2018) 心理学におけるオープンサイエンス「統計革命」のインフラストラクチャー 心理学評論, 61, 3–12.
- 宮川雅巳 (2004) 統計的因果推論—回帰分析の新しい枠組み 朝倉書店.
- Mulder, J., & Wagenmakers, E.-J. (2016). Editors' introduction to the special issue "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments." *Journal of Mathematical Psychology*, 72, 1–5.
- 中村國則 (2018) 高次認知研究におけるベイズのアプローチ 心理学評論, 61, 67–85.
- 岡田謙介 (2014) ベイズ統計による情報仮説の評価は分散分析にとって代わるのか? 基礎心理学研究, 32, 223–231.
- 岡田謙介 (2018) ベイズファクターによる心理学的仮説・モデルの評価 心理学評論, 61, 101–115.
- 大久保街亜 (2016) 帰無仮説検定と再現可能性 心理学評論, 59, 57–67.
- 大久保街亜・岡田謙介 (2012) 伝えるための心理統計: 効果量・信頼区間・検定力 勁草書房.
- 奥村晴彦・牧山幸史・瓜生真也 (2018) R で楽しむベイズ統計入門 技術評論社.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59.
- Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, 18, 682–689.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- 杉万俊夫 (2005) 社会構成主義と心理学 —「内なる心」の観念を超えて 下山晴彦 (編著) 心理学論の新しいかたち (pp. 66–84) 誠信書房.
- 竹澤正哲 (2018) 心理学におけるモデリングの必要性 心理学評論, 61, 42–54.
- Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401–409.
- 豊田秀樹 (2015) 基礎からのベイズ統計学: ハミルトニアンモンテカルロ法による実践的入門 朝倉書店.
- 豊田秀樹 (2016) はじめての統計データ分析 ベイズ的〈ポスト p 値時代〉の統計学 朝倉書店.
- 豊田秀樹 (2017) 実践ベイズモデリング—解析技法と認知モデル— 朝倉書店.
- 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, 426–432.
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.
- 渡辺澄夫 (2012) ベイズ統計の理論と方法 コロナ社.

— 2018. 2. 5 受稿, 2018. 3. 7 受理 —