

特集

ベイズ統計学が拓く 調査データ解析の 最前線



マイクロソフトの共同創業者であるビル・ゲイツは、1996年にマイクロソフトが競争優位に立っている理由としてベイズ統計学の技術を挙げた。そして、2001年には21世紀のマイクロソフト社の戦略はベイズ統計学であると宣言した。現在、ベイズ統計学は自動運転や迷惑メールフィルターなどで使われ、人工知能開発においても重要な役割を果たしている。データサイエンスが隆盛を極める現代において、ベイズ統計学はその中心的存在とも言えるだろう。

一方で、ベイズ統計学は社会科学分野の調査データ解析研究でも活用されているが、そのことは研究者の中でもあまり知られていない。「ベイズ統計学」という言葉は知っているが、それが一体なんであるかは知らないという読者も多いだろう。本特集は、心理学・教育学・マーケティング分野でベイズ統計学による研究を行っている6名の執筆者をお招きし、各分野における調査データを題材としてベイズ統計学の活用方法や有効性について論じてもらった。また、ベイズ統計学による分析を具体的にやりたい読者のために、ソフトウェア案内についても企画した。読者にベイズ統計学への注目を持ってもらい、自身のデータ分析に役立ててもらうことができれば幸いである。

(尾崎幸謙)



特集論文

1

バイズ統計学の考え方

岡田謙介

東京大学大学院教育学研究科 准教授

バイズ統計学は、不確実性を確率によって定量化し、観測データに基づいてその確率を更新するという原則を、さまざまな個別のデータ分析場面に適用していく体系である。本稿では、このことをパラメータの推定、将来のデータの予測、そしてモデル選択（仮説検定）という3つの場面において論じる。具体的に話を進めるため、世論調査データの分析を例として用いる。その上で、点推定と区間推定、事前分布の設定、そして推定アルゴリズムについての補足的な議論を行う。

1 はじめに

本稿の主要な目的は、バイズ統計学の考え方を紹介することである。その基本的な考え方は、次の2つの一般原則からなる（e.g., Lee & Wagenmakers, 2013, 井関訳, 2017）。(1) 不確実性は確率によって表現されること。(2) 観測データで条件付けて、その確率を更新すること。

すなわち、バイズ統計学では、関心の対象についての不確実性を、観測データで条件付けた条件付き確率によって定量的に評価する。この原則を、あらゆるデータ分析場面に適用することによって、首尾一貫した統計的データ分析の体系が構築される。

次節では簡単な具体例を用いながら、この一

般原則が「パラメータの推定」、「将来のデータの予測」、そして「モデルの選択・仮説検定」という、3つの代表的なデータ分析場面においてどのように実現されるのを見えていく。その後3節では、限られた紙幅の中ではあるが、実践上重要となる、「点推定と区間推定」、「事前分布の設定」、そして「推定アルゴリズム」について論じる。

2 観測データで条件付けた不確実性の評価

2.1 推定

世論調査を行う場合を考えてみよう。目的は、対象集団における内閣支持率を推定することである。集団のサイズは十分に大きく、構成員全体からデータを観測することは現実的ではないとする。このとき、推定対象である内閣支持率は、分析者には正確な値がわからず、データに基づいて推定したい変数である。言い換えれば、不確実な変数である。バイズ統計学では、こうした不確実性を、データで条件付けた確率によって定量化して評価する。

集団全体での内閣支持率を θ で表すことにしよう。この θ は比率であるため、0 (0%) から1 (100%) までの値をとる連続的な変数である。この集団から選ばれた n 人のデータを観測し、そのうち支持すると回答した人の数が y 人

だったというデータが得られたとする。このとき、 y の確率分布は二項分布

$$y \sim \text{Binomial}(n, \theta) \quad (1)$$

となる。(1)式は、確率関数を陽に書けば

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y} \quad (2)$$

である。等価な表現である(1)式および(2)式は、観測データ y が生成される確率的メカニズムを表している。そうした確率分布をデータ分布(data distribution)、あるいは θ の関数と見て尤度関数(likelihood function)、もしくは単に尤度(likelihood)という。本稿では以下、尤度と呼ぶことにする。

次に、データを観測する前に我々が θ について持っている不確実性を、確率によって表すことを考える。一つの合理的な方法として、その台、すなわち、とりうる値の全範囲について、確率密度が一様であるという設定

$$p(\theta) = \begin{cases} 1 & (\theta \in [0,1] \text{ のとき}) \\ 0 & (\text{それ以外のとき}) \end{cases} \quad (3)$$

を考えることができる。これはすなわち、図1の最も薄い灰色線で示すとおり、 θ が

$$\theta \sim \text{Uniform}(0,1) \quad (4)$$

と、0と1をそれぞれ最小値・最大値とする一様分布にしたがうということである。このような、データを観測する前のパラメータの不確実性を表現する確率分布を事前分布(prior distribution)という。事前(prior)という語は、「データを観測する前の」という意味である。特段利用できる情報がない場合には、台上の一様分布を事前分布として利用する、という考え方は、無差別性原理(principle of indifference)と呼ばれる(Janes, 2003; Lee & Wagenmakers,

2005)。とりうる各値について差をつける理由がないことから、等しい確率もしくは確率密度を割り当てるという考え方である。

我々が推定したいのは θ についての確率であり、用いることができる情報は観測データ y である。すなわち、観測データ y で条件付けた、パラメータ θ についての確率 $p(\theta|y)$ を求めたい。この $p(\theta|y)$ を事後確率(posterior probability)という。事前確率に対して、データ y を観測し、その情報を用いて更新した後の確率という意味である。

この事後確率について、ベイズの定理を適用すれば、

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (5)$$

が成立する。今回の例では、我々はこの(5)式右辺の分子 $p(y|\theta)$ を(2)式(もしくは、その別表現である(1)式)によって、また $p(\theta)$ を(3)式(もしくは、その別表現である(4)式)によって定めた。一方、分母は

$$p(y) = \int_{\Theta} p(y|\theta)p(\theta)d\theta \quad (6)$$

と表すことができる。つまり、分子が設定され観測データ y が与えられれば、 θ を積分消去することで分母も計算できる。あるパラメータについて積分消去することを、周辺化する(marginalize)ともいうことから、(6)式の量は周辺尤度(marginal likelihood)と呼ばれる。

このように(5)式は、分析者が尤度 $p(y|\theta)$ と事前分布 $p(\theta)$ を定め、実際に観測されたデータ y の値を代入すれば、関心の対象である確率、すなわち事後分布 $p(\theta|y)$ を求められることを意味する。このことから、尤度と事前分布の組 $p(y|\theta)p(\theta) = p(y, \theta)$ をモデル(model)という(e.g., Gelman et al., 2013; O'Hagan & Forster, 2004)。すなわち、ベイズ統計学的な推論のために分析者が行うべきことは、データとパラメータの同時分布であるモデルを設定し、データ y を観



測することである。そうすれば、観測データで条件付けて、(5) 式のベイズの定理によって事後分布を求めることができる。

世論調査の例において、表1に示す全部で $n=15$ のデータを観測し、 θ の事後分布を求めるとしよう。左から3人分の $n=3$ のデータとして分析したときの事後分布、左から6人分の $n=6$ のデータとして分析したときの事後分布というように、全15人分の $n=15$ のデータとして分析したときの事後分布を、重ね描きして図1に示した。データのサンプルサイズが増えるにつれて、不確実性が更新され、 θ の事後分布における確率密度の大きい値と小さい値とのコントラストが明確になっていく様子を見ることができる。なお、このモデルのもとでは θ の事後分布はベータ分布になることを示すことができ、 $n=15$ のデータで

条件付けた事後分布は $Beta(5,12)$ である。

2.2 予測

分析者の関心対象は、パラメータよりもむしろ再度データを収集したときにどのような値が得られるか、にある場合もある。たとえば、前節で述べた $n=15$ のデータを観測したもとの、同じ集団からさらに10人のデータを観測しようとするとき、このうち何人が内閣支持と回答するのかが研究関心であるとしよう。このとき、将来の観測における内閣支持者数を変数 y^{new} で表すことにする。

分析者の関心対象はパラメータから将来のデータに変わったが、その値が不確実であり、したがって、観測データで条件付けて更新したその確率を求めるというベイズ統計学の原則は変わらない。すなわち、求めたい量は観測データ y で条件付けた将来のデータ y^{new} の確率 $p(y^{new}|y)$ である。これは、

$$p(y^{new}|y) = \int_0^1 p(y^{new}|\theta)p(\theta|y)d\theta \quad (7)$$

と表現できる。ここで $p(\theta|y)$ は前節で求めた事後分布である。他方、 $p(y^{new}|\theta)$ は将来のデータの尤度であるが、将来のデータも今回観測されたデータと同じデータ生成過程によって生成されたとすれば、今回観測されたデータに用いた尤度と同じ確率分布が設定できる。したがって $p(y^{new}|\theta)$ に二項分布を設定すると、(7) 式の積分計算によって $p(y^{new}|y)$ を求めることができる。パラメータについての、データで条件付けた不確実性の更新を推定 (estimation) というのに対し、将来のデータについてのデータで条件付けた不確実性の更新は予測 (prediction) という。このことから、(7) 式で求める $p(y^{new}|y)$ は事後予測分布 (posterior predictive distribution) と呼ばれる。

世論調査の例における、将来10人を観測した場合での事後予測分布 $p(y^{new}|y)$ を図2に示す。

表1 世論調査の回答データ。1は支持を、0は不支持を表す。本文における y は「回答」の行に示したダミー変数値の合計となる。

	◀ n=15のとき ▶			◀ n=12のとき ▶						◀ n=9のとき ▶			◀ n=6のとき ▶			◀ n=3のとき ▶			◀ n=0のとき ▶			
回答者番号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15							
回答	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0							

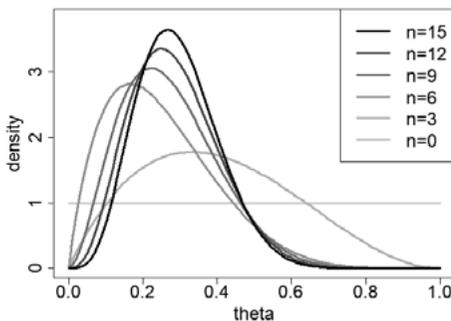


図1 θ の事前分布 ($n=0$) に加え、表1における $n=3$ のデータでの θ の事後分布、 $n=6$ のデータでの θ の事後分布、 \dots 、 $n=15$ のときの事後分布を重ねてプロットした図。観測データで条件付けた不確実性の更新の様子を見てとることができる。

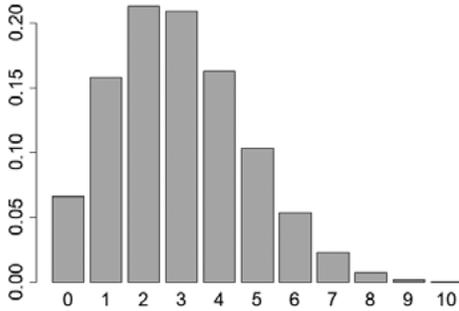


図2 2.1節でのn=15のときの θ の事後分布に基づいて求めた、将来の10人の観測における y^{new} の事後予測分布

θ の事後分布(図1)と y^{new} の事後予測分布は、異なる確率変数についての、異なる意味を持つ分布であるが、類似した形状を示していることがわかる。両確率変数の意味を考慮すれば、これは今回の文脈において納得のできることであろう。なお、このモデルと観測データの下では、事後予測分布も解析的に求めることができ、ベータ二項分布 $BetaBinomial(10,5,12)$ である。

2.3 モデル選択・仮説検定

ここまで、パラメータについての事後分布を求める推定と、将来のデータについての事後予測分布を用いる予測を扱ってきた。データ分析におけるもうひとつの分析者の重要な関心事に、モデル選択(model selection)がある。

前節までで考えていたモデルは、

$$H_1 \begin{cases} y \sim Binomial(n, \theta) \\ \theta \sim Uniform(0,1) \end{cases} \quad (8)$$

であった。これに対して、 $\theta = 0.50$ であるという、もうひとつのモデル

$$H_0 \begin{cases} y \sim Binomial(n, \theta) \\ \theta = 0.50 \end{cases} \quad (9)$$

を考えよう。モデル H_1 は θ を未知のパラメータとして一様事前分布のもとで推定するモデルであるが、モデル H_0 は θ を0.50として固定するモデル

である。仮に、回答者がでたらめに支持か不支持かを選ぶ場合、 $\theta = 0.50$ となる。このように、仮にでたらめな回答がされた場合のモデル H_0 よりも、モデル H_1 がデータから大きく支持された場合には、回答行動は単にでたらめに行われたのではなく、回答者集団の特徴を反映したものだとする証拠のひとつとなるだろう¹⁾。

2.1節において、(5)式によって行ったパラメータの事後分布を求める推定は、モデル H_1 を所与としたもとで行われたものである。この推定がモデルを前提としていることを明示的に表せば、(5)式は

$$p(\theta|y, H_1) = \frac{p(y|\theta, H_1)p(\theta|H_1)}{p(y|H_1)} \quad (10)$$

と表現しなおすことができる。同様に、2.2節で扱った予測も、全ての確率をモデル H_1 で条件付けた形で書くことができる。複数のモデル間での比較・選択を考えない場合には、こうしたモデルによる条件付けは、明示的には表記されないことが多い。しかし、実はあらゆる推定や予測は、設定されたモデルのもとで、すなわちモデルで条件付けて行われる。この認識を持つことは重要であり、ベイズモデル平均化(Bayesian model averaging; e.g., Hoeting, Madigan, Raftery & Volinsky, 1999)のようなモデルの不確実性への対応を可能にする拡張へも繋がる。

本節ではモデル選択の問題を考える。我々が知りたいのは、2つのモデル H_0 と H_1 の間での不確実性である。繰り返して述べているように、不確実性を観測データで条件付けた確率によって評価することが、ベイズ統計学の一般原則である。これにしたがって、モデル選択も、観測データで条件付けた事後モデル確率(posterior model probability) $p(H_1|y)$ を求めて行うことができる。なお、いま2つのモデルだけを考えているので、もう一方のモデルの事後モデル確率は $p(H_0|y) = 1 - p(H_1|y)$ となる。つまり一方のモデルの確率が決まれば、もう一方のモデルの



確率も自動的に定まるので、一般性を失わずにモデル H_1 について考える。

事後モデル確率は、再びベイズの定理を用いることにより、

$$p(H_1|y) = \frac{p(y|H_1)p(H_1)}{p(y)} \quad (11)$$

と書くことができる。ここで (11) 式の分子の $p(y|H_1)$ とは (5) 式、もしくは等価な (10) 式の分母である周辺尤度である。また (11) 式の分母は、 $p(y) = p(y|H_1)p(H_1) + p(y|H_0)p(H_0)$ と求められる。したがって、観測データを得る前のモデルの確率である $p(H_1)$ が決まれば、求めたい確率 $p(H_1|y)$ を求めることができる。これまでの用語の使い方と同様にして、 $p(H_1)$ をモデル H_1 の事前モデル確率 (prior model probability)、 $p(H_1|y)$ を事後モデル確率という。

事前モデル確率を具体的に設定するにあたり、ひとつの合理的な方法は、無差別性原理にしたがって $p(H_1) = p(H_0) = 0.50$ と、比較する2つのモデルに等しい事前モデル確率をおくことであろう。このとき、表1のデータに基づいて更新された事後モデル確率は $p(H_1|y) = 0.60$ となる。

事後モデル確率は、不確実性を確率によって評価し、データで条件付けてその確率を更新するというベイズ統計学の原則から直ちに導かれるモデル評価のための量である。ただし、モデルの事後確率は、可能なモデルがいくつあるかに依存する。たとえば比較するモデルがもうひとつある場合には、無差別性原理に基づけば各モデルに割り当てられる事前確率は $1/3$ ずつとなり、対応してモデル H_1 の事後モデル確率も、比較するモデル数が2つであった場合よりも小さくなる。

ここで、(11) 式に対応する式を H_0 についても書き、両者の比をとると、分母が打ち消し合って、2つのモデルの事後確率の比が

$$\frac{p(H_1|y)}{p(H_0|y)} = \frac{p(y|H_1) p(H_1)}{p(y|H_0) p(H_0)} \quad (12)$$

と書ける。左辺の $\frac{p(H_1|y)}{p(H_0|y)}$ は、2つのモデルの事後確率の比である。確率の比のことをオッズ (odds) というため、これを事後モデルオッズ (posterior model odds) という。事後モデルオッズは、モデル H_1 の事後確率が、モデル H_0 の事後確率の何倍であるかを表す比である。2つのモデルの事後確率が等しいとき、すなわちデータの情報で更新したときに2つのモデルの確率が同じになるとき、事後モデルオッズは1になる。同様に、右辺の $\frac{p(H_1)}{p(H_0)}$ は事前モデルオッズ (prior model odds) という。

(12) 式を変形すると

$$\frac{p(y|H_1)}{p(y|H_0)} = \frac{p(H_1|y) p(H_0|y)}{p(H_0|y) p(H_1|y)} \quad (13)$$

となる。すなわち、2つのモデルの周辺尤度の比は、2つのモデルの事後モデルオッズと事前モデルオッズの比である。このモデルオッズの比は、観測データの情報を用いることによって、モデルオッズが事前から事後へ何倍変化したのかを表す。すなわち、その値が2であるということは、モデル H_0 に対して H_1 を支持するオッズが、観測データ y によって2倍になったと解釈できる。このように解釈される、(13) 式の周辺尤度の比のことをベイズファクター (Bayes factor) という。したがって、ベイズファクターは1より大きいときに H_1 が、1より小さいときに H_0 が、データから支持される比の形の量であり、その支持の大きさが大きいほど、1からの乖離の程度が大きくなる。なお、ベイズ情報量規準 (Bayesian information criterion, BIC) は、ベイズファクターの粗い近似である (Kass & Raftery, 1995)。

連続的なパラメータ θ を点帰無仮説に対応する形で制約する (9) 式のようなモデル H_0 と、制約しない対立仮説に対応する、(8) 式のようなモデル H_1 との間で、ベイズファクターによってモデル選択を行うことは、ベイズ統計学的に仮

説検定 (hypothesis testing) を行う最も代表的な方法である。このような、点帰無仮説に対応するモデル H_0 と、対立仮説に対応するモデル H_1 との間のベイズ的仮説検定では、図3に示すように、ベイズファクターは H_0 で制約する点における、 H_1 のもとでの事前分布と事後分布の確率密度の比という単純な量に帰着し (e.g., Wagenmakers, Lodewyckx, Kuriyal & Grasman, 2010), $B_{10} = 1.50$ となる。この値は1に近いものであり、どちらかと言えばデータからモデル H_1 の方が H_0 よりも支持されてはいるものの、依然としてモデルの不確実性が大きく、一方のモデルを他方よりも積極的に選択する理由には乏しいといえる (岡田, 2018も参照してほしい)。

3 実践上のいくつかの論点

本節では、前節で述べたベイズ統計学の推定・予測・モデル選択に関連する、いくつかの重要な点について補足的な議論を行う。

3.1 数値要約 (点推定値・区間推定値)

本稿で繰り返し述べているように、ベイズ統計学は、不確実性を確率によって表現し、データの情報を用いて (データで条件付けて) 確率を更新していく体系である。したがって、推定については図1に、予測については図2に示したような、データで条件付けて得られる事後の確率分布それぞれ自体が、最も豊かな情報を持つ。しかし、実際の応用で用いるモデルでは、関心のあるパラメータが多数あることも多い。また、独立でない2パラメータ以上の同時確率分布は、視覚的に表現したり、理解することがしばしば困難である。こうしたことから、確率分布の情報を簡潔に数値で要約したいという要求が生まれる。

このような目的で、確率分布を各種指標によって数値要約することができる。たとえば図1の事後分布を要約する方法として、事前分布と $n = 15$ のデータを用いた事後分布の各種要約指標を表2に示した。とくに、事後分布の平均はEAP (expected a posteriori), 事後分布の最頻値はMAP (maximum a posteriori) と呼ばれ、頻繁

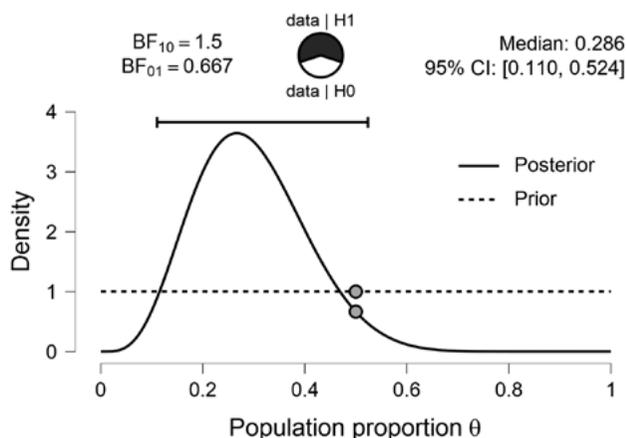


図3 JASP (JASP Team, 2020; Wagenmakers et al., 2017; 清水・山本, 2020) によって、表1のデータを用いて (8) 式の H_1 と (9) 式の H_0 との間でベイズ的仮説検定を行った場合に出力される図。確率密度は H_1 のもとでの事前分布 (点線) と事後分布 (実線) である。ベイズファクター B_{10} は、 H_0 で点帰無仮説として設定する点 $\theta = 0.50$ における、 H_1 のもとでの事前分布と事後分布の確率密度の比として求められ、 $B_{10} = \frac{1}{0.667} = 1.50$ となる。



に利用される事後分布の要約値である。2変数(以上)間の関係も、共分散や相関係数などによって同様に数値要約して報告できる。また、パラメータが所与の確率で含まれる区間を信用区間(credible interval)という。(100 - α)%信用区間の代表的な構成の仕方には、最も短い区間となるように構成される事後最高密度区間(highest posterior density, HPD) と、両裾に $\alpha/2$ ずつの確率密度をとった間の区間である中心事後区間(central posterior interval, CPI) がある。このように、事後分布自体の数値要約値を報告することが、ベイズ統計学における点推定や区間推定の考え方となる。

なお、伝統的な統計学においては、点帰無仮説の仮説検定において有意水準 α の検定で帰無仮説が棄却されるか否かと、(100 - α)% 信頼区間(confidence interval) が当該の点を含むか否かとが一致する。したがって、推定の方法である信頼区間が特定のパラメータ値を含むか否かによって、2つのモデル間の選択を行うことがある。しかし、ベイズ統計学の考え方にしたがって推定と仮説検定を行う場合、両者の結果は一般に一致しない。したがって、モデルおよび仮説の評価・選択のためには、パラメータの推定ではなく、モデル選択のための方法を用いて行うのが適切である。両者が一致しない一例は Wagenmakers, Lee, Rouder and Morey (2020) に見ることができる。

3.2 事前分布

事後分布 $p(\theta|y)$ をベイズの定理を用いて求めるためには、モデルの設定が必要であり、したがっ

て事前分布 $p(\theta)$ の設定が必要である。これをどのように設定するかは、しばしば議論の対象となる。しかしここで、分析者が設定しなければならぬのは、事前分布だけではなくモデルも必要であること、すなわち尤度と事前分布の両方であることに注意する。統計モデルを構成する2つの要素のうち、事前分布だけが議論の対象とされがちであるのは、尤度は伝統的な統計学の応用においても設定されてきたのに対し、事前分布の設定はベイズ統計学に特徴的であるためと思われる。しかし、分析者がデータ分析における自らの仮説や理解に基づいて設定しなければならない点、また、そこでなされる仮定の根拠を説明することが求められる点において、尤度と事前分布の間が変わるところはない。(θ を所与とした) データ y の確率的生成過程を表現するのがデータ分布(尤度)であり、パラメータ θ の確率的生成過程を表現するのが事前分布である。

また、事後分布は、データで条件付けて事前分布を更新することで導かれる。この際、データのサンプルサイズが大きくなればなるほど、相対的に事前分布の影響は小さくなる。ある程度のサンプルサイズがあれば、推定における事前分布の設定は、しばしば、実用上あまり問題にならなくなる。事前分布の設定の違いによって推定結果がどの程度影響されるのかを調べるためには、感度分析(sensitivity analysis)を行うことが有用である(e.g., Roos, Martins, Held & Rue, 2015)。

さらに、事前分布を設定することは、より積極的な実用的利点もある。第一に、現実のデータは、たとえば社会調査において、個人は家庭に、家庭は自治会等に、自治会は地方自治体に、といったネストした階層構造があると考えられる。各階層においては、群内での類似性が認められ、これをモデル上で表現する必要が生じる。こうした構造は、階層事前分布を用いることで自然に表現できる。久保(2012)はそうした流れで階層ベイズモデルを導入している。第二に、大規

表2 事前分布と事後分布の数値要約値。信用区間の算出にはCPIを利用した。

	事前分布(n=0)	事後分布(n=15)
平均値	0.500	0.294
最頻値	—	0.267
標準偏差	0.289	0.107
95%信用区間	[0.025, 0.975]	[0.110, 0.524]

模データを扱う上で、多数の説明変数候補がある中で変数選択と推定を同時に行うことができる統計解析法として正則化 (regularization) が注目されている (川野・松井・廣瀬, 2018)。正則化では罰則項を導入するが、これはベイズ統計学の観点からは事前分布と理解することができる。すなわち、適切な事前分布を導入することによって、変数選択と推定を同時に行うことが可能になる。こうした方法はベイズ正則化 (Bayesian regularization) と呼ばれる (e.g., Erp, Oberski & Mulder, 2019)。

3.3 推定法

2節で示した数値例は、いずれも事後分布や事後予測分布、事後モデル確率などが解析的に求められる場合であった。しかし現代において、実際のデータに適用されるモデルは、その生成過程を反映した、よりパラメータ数が多く、また関数形の複雑なモデルであることが多い。そうしたモデルでは、多くの場合、事後分布を解析的に求めることができない。これは、1980年代以前において、ベイズ統計学の普及を妨げる大きな要因であった。

しかし、Gelfand and Smith (1990) によってマルコフ連鎖モンテカルロ (Markov chain Monte Carlo, MCMC) 法を用いたベイズ推定の実用性が示され、またそれを汎用的に実装した WinBUGS などの確率的プログラミング言語が普及したことによって、ベイズ統計学は実用的な学問へと変貌を遂げた。JAGS や Stan といった、MCMC 法による推定を実装した汎用確率的プログラミング言語が、ベイズ推定の普及に果たしている役割は大きい (岩波データサイエンス刊行委員会, 2015, 清水, 2018)。和書においても、たとえば伊庭 (2018), Lee and Wagenmakers (2013, 井関訳, 2017), 豊田 (2018) などにさまざまな分野でのそうした例が報告されている。

一方で、MCMC 法がどんなときでも実用的であるわけではなく、ベイズ推定のよりよい方

法に関する研究は活発に行われている。我々の研究から例を挙げると、Okada and Mayekawa (2018) では多変量解析における MCMC 推定時の不定性の問題を解決するため、因子回転のアイデアを拡張し適用する方法を提案した。また、Yamaguchi and Okada (2020) では認知診断モデルを対象に、変分ベイズ推定という機械学習で発展した近似計算法のアルゴリズムを開発して、MCMC 法に匹敵する精度の計算を、大幅に高速に行えることを示した。

4 むすびに

ベイズ統計学は、不確実性を確率によって評価し、データで条件付けてその確率を更新するという原則を、さまざまな応用場面に對して適用していく、シンプルで首尾一貫した体系である。本稿ではこのことを論じ、パラメータの推定、将来のデータの予測、モデルや仮説の評価という、基本的な3つのデータ分析上の目的について、実例を用いて示した。その上で、点推定と区間推定、事前分布の設定、そして推定法について、実践上の補足的な議論を行った。少数の一般原則を、データ生成過程を反映した豊かなモデルとともに具体的な応用場面に適用していくベイズ統計学は、現代の統計的データ分析において大きな訴求力を持つ。今後さらに諸分野での応用例が蓄積され、我々がさまざまな現象を理解し予測するのに役立つことが期待される。

謝辞

本稿の初稿に対して有益なコメントをいただいた野村圭史、野々田聖一の両氏 (共に東京大学大学院教育学研究科) に感謝いたします。

注

- 1) ただし、今回の例では、実際に集団中での支持率が50%である場合にも H_0 が支持されることに注意が必要である。



文献

- Erp, S. Van, Oberski, D. L. & Mulder, J., 2019, "Shrinkage priors for Bayesian penalized regression", *Journal of Mathematical Psychology*, 89: 31-50. (<https://doi.org/10.1016/j.jmp.2018.12.004>).
- 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. (<https://doi.org/10.2307/2289776>).
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B., 2013, *Bayesian data analysis* (3rd ed.), Boca Raton, FL: Chapman and Hall/CRC.
- Hoeting, J. A., Madigan, D., Raftery, A. E. & Volinsky, C. T., 1999, "Bayesian model averaging: A tutorial", *Statistical Science*, 14(4): 382-417. (<https://doi.org/10.1214/ss/1009212519>).
- 伊庭幸人編, 2018, 『ベイズモデリングの世界』岩波書店。
- 岩波データサイエンス刊行委員会, 2015, 『岩波データサイエンス Vol.1 特集: ベイズ推論とMCMCのフリーソフト』岩波書店。
- Jaynes, E. T., 2003, *Probability theory: The logic of science*, Cambridge, UK: Cambridge University Press.
- JASP Team., 2020, JASP (Version 0.12.2) [Computer software], (Retrieved from <https://jasp-stats.org/>).
- Kass, R. E. & Raftery, A. E., 1995, "Bayes factors", *Journal of the American Statistical Association*, 90(430): 773-795. (<https://doi.org/10.1080/01621459.1995.10476572>).
- 川野秀一・松井秀俊・廣瀬慧, 2018, 『スパース推定法による統計モデリング』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。
- 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.)
- Lee, M. D. & Wagenmakers, E.-J., 2005, "Bayesian statistical inference in psychology: Comment on Trafimow" 2003, *Psychological Review*, 112(3), 662-668. (<https://doi.org/10.1037/0033-295X.112.3.662>).
- 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(1): 101-115. (https://doi.org/10.24602/sjpr.61.1_101).
- Okada, K. & Mayekawa, S., 2018, "Post-processing of Markov chain Monte Carlo output in Bayesian latent variable models with application to multidimensional scaling", *Computational Statistics*, 33(3): 1457-1473. (<https://doi.org/10.1007/s00180-017-0759-6>).
- O'Hagan, A. & Forster, J., 2004, *Kendall's advanced theory of statistics. Volume 2B: Bayesian inference* (2 ed), New York, NY: Wiley.
- Roos, M., Martins, T. G., Held, L. & Rue, H., 2015, "Sensitivity analysis for Bayesian hierarchical models", *Bayesian Analysis*, 10(2): 321-349. (<https://doi.org/10.1214/14-BA909>).
- 清水裕士, 2018, 「心理学におけるベイズ統計モデリング」『心理学評論』61(1): 22-41. (https://doi.org/10.24602/sjpr.61.1_22).
- 清水優菜・山本光, 2020, 『研究に役立つJASPによるデータ分析: 頻度論的統計とベイズ統計を用いて』コロナ社。
- 豊田秀樹編著, 2018, 『たのしいベイズモデリング——事例で拓く研究のフロンティア』北大路書房。
- Wagenmakers, E.-J., Lee, M. D., Rouder, J. N. & Morey, R. D., 2020, "The principle of predictive irrelevance, or why intervals should not be used for model comparison featuring a point null hypothesis. In C. Gruber (Ed.), The theory of statistics in psychology: Applications, use and misunderstandings, New York: Springer.
- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H. & Grasman, R., 2010, "Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method", *Cognitive Psychology*, 60(3): 158-189. (<https://doi.org/10.1016/j.cogpsych.2009.12.001>).
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J. et al., 2017, "Bayesian inference for psychology. Part II: Example applications with JASP", *Psychonomic Bulletin and Review*: 1-19. (<https://doi.org/10.3758/s13423-017-1323-7>).
- Yamaguchi, K. & Okada, K., 2020, "Variational Bayes inference for the DINA model", *Journal of Educational and Behavioral Statistics*, Online Ahead of Print. (<https://doi.org/10.3102/1076998620911934>).

2

アフターコロナ時代のテスト ベイズ統計学を利用した 項目反応理論に基づく コンピュータ適応型テスト

登藤直弥

筑波大学人間系 助教

1 はじめに

2020年の上半期、新型コロナウイルスの流行により、日本をはじめとする世界各地において、これまでの生活様式が一変した。効果的なワクチン・治療薬の見つかっていないこの病気の感染拡大を防ぐため、人々はできるだけ自宅に籠っての自粛生活を余儀無くされ、社会・経済活動の一部がオンラインで行われるようになった。このオンライン化の波は教育活動にも押し寄せてきており、私の勤務する大学を含めた多くの教育機関では、オンラインでの授業が行われるようになってきている。加えて、受験者の人生を左右するような入試や就職活動、資格試験等に関しても、オンラインで行われるようになってきている。

この原稿を執筆している2020年6月現在、日本における新型コロナウイルスの流行は沈静化しつつあり、徐々にではあるが、社会・経済活動も、新型コロナウイルス流行前の姿に戻つつある。ただし、その戻つつある先が流行前の姿と完全に一致するかというと、どうやらそうはならないようである。たとえば、新型コロナウイルスの流行以降、街を歩き交う人々のほとんどはマスクを着用するようになったし、依然としてリモートワークや在宅勤務を行う方々も多くいる

ようである。このように、このコロナ禍を契機として、我々の社会・経済・教育活動は、それがどの程度のものとなるかはわからないのだが、変化していく事になるのだろう。

ところで、私が専門としている教育測定学では、テストや質問紙などの尺度を利用した人間の能力や態度といった目に見えない心理学的構成概念の測定に関する研究が行われている。こういった尺度の実施に関しても、アフターコロナの時代には、何らかの変化がみられるのであろうか。たとえば、先述の通り、一部の資格試験や入試に関しては、既にコロナ禍の影響でオンラインによる実施が進められているし、質問紙を使った研究でも、私の聞く限りでは、オンラインでの実施が多く出ているようである。このような新たな実施の形態は、アフターコロナの時代においても、続くのであろうか。今後の推移を見守るしかその答えを知る方法は無い。一教育測定学者としては、この新型ウイルスの流行によって尺度、特にテストの実施に制約が生まれ、従来のように多くの受験者を一箇所に集めてテストを実施することが難しくなったことは、今まで教育測定学の中で研究されてきた様々な研究成果を社会に還元する、つまり、実際のテスト運用に実装する大きな契機になるのではないかと考えている。



たとえば、そのような研究成果の1つに、「項目反応理論 (item response theory, 以下, IRT)」（e.g., Lord & Novick, 1968）に基づく適応型テストがあろう。現在でも、多くのテストを分析する際に、各項目の正答率や平均点といった指標が算出されているが、これらの指標に基づく分析には限界があることがかねてから指摘されていた。この課題を克服するためのテストデータの分析枠組みとして、1900年代半ばから教育測定学の分野で盛んに研究が行われてきたのがIRTである。IRTに基づくテストの作成・評価・運用には、たとえば、異なる集団が異なるテストを受験した場合にもその結果が比較できるようになる、テストを実施する前にその平均点を予測できるようになる等、様々な利点があるのだが、そのような利点の1つとして、適応型テストと呼ばれるテストを実施できるようになることが挙げられる。

適応型テストとは、受験前に出題する項目を全て決めておく従来のテストとは異なり、受験者のそれまでの解答に基づいて次に出題する項目を適応的に決めていくテストを指している。たとえば、コンピュータ上で適応型テストを実施した場合、このようなテストを特にコンピュータ適応型テスト (computerized adaptive testing, 以下, CAT) と呼ぶ。CATでは、その時点での受験者の学力などの受験者特性値をそれまでの解答結果に基づきリアルタイムで推定し、その推定結果に基づいて、次に出題する項目を当該受験者の特性値を推定するうえで最適になるように選択して出題することが可能になる。

実際に利用されている例は、本邦に関しては、CASEC (Computerized Assessment System for English Communication) やRST (Reading Skill Test) (e.g., Arai et al., 2017) など、まだまだ少ないというのが実情だと思われる。しかし、上記の説明からも明らかなように、一般に、CATを用いると、従来のテストに比べてより少数の項目で精度の高い特性値の推定を行うことができ

ようになるし、各受験者が異なる項目から成るテストを受験する事になるため、カンニング等の不正への対応もより強固になる。つまり、従来のテストをCATの形式で運用することにより、不正に対応しつつ、多くの受験者を一箇所に集めることなくテストを実施することが可能となる。しかも、テストの受験者と実施者、双方の負担も、出題数を減らせるという意味で、軽減できるようになるのである。このような特徴を有するCATは、アフターコロナの時代における「新しいテストの形」になり、仮に従来のテストが実施できるようになったとしても、その利用はますます広がっていくのではないだろうか。

そこで、本稿では、そのような「新しいテストの形」が普及する一助になればとの思いから、このIRTに基づくCATの概要について解説を行う。また、実際にIRTに基づくCATを運用していく際には、本特集のテーマであるベイズ統計学の考え方が非常に重要となってくる。そこで、IRTに基づくCATにおけるベイズ統計学の利用についても、その初歩的内容について解説を行うことにする。

以下、本稿の構成は以下の通りとなっている。まずは第2節にて、CATと関係する部分を中心に、IRTの概要について解説を行う。続いて、第3節にて、IRTに基づくCATの基本的な事項についての解説を行う。その後、第4節にて、IRTに基づくCATでのベイズ統計学の利用に関して基礎的な事項を解説し、最後の第5節にて、本稿のまとめを行う。

2 項目反応理論 (IRT)

項目反応モデル

IRTとは、先述の通り、テストを作成・評価・運用していくうえでの様々な利点を有する、テストデータを分析するためのテスト理論の1つである。たとえば、偏差値を算出する際など、我々がテストデータを分析する際には、通常、そ

の合計点を対象にすることが多いのだが、IRTでは、正答、誤答など、各項目への反応に焦点を当てることになる。もう少し具体的に述べると、IRTでは、そのテストで測定しようとしている受験者特性値と各項目への反応との間にある関係が、受験者特性値等を未知の母数として含むある数学的な関数により表現できると仮定する。この関数のことを、項目反応理論で用いられるモデルということで「項目反応モデル」という。そして、実際にテストを実施して収集された項目反応データに対してこの項目反応モデルをあてはめて母数の値を推定し、得られた推定値に基づいて、テストの評価・作成・運用等を行っていくことになる。

では、この項目反応モデルとして、実際には、どのようなものが利用されるのであろうか。項目反応モデルとしては、先行研究において、実に多くのものが提案されているのだが、ここでは、その中でも実際に利用されることの多い代表的なものに関して、登藤 (2019) をベースに、更に平易な解説を行なってみたい。

1パラメタ・ロジスティックモデル

IRTに基づき項目反応データを分析する際によく利用される項目反応モデルの1つに「1パラメタ・ロジスティックモデル (以下、1PLM)」と呼ばれるものがある。たとえば今、受験者の語学力を測定するための5つの項目からなるテストがあったとすると、1PLMにおいては、受験者特性である語学力と各項目への正答率との関係が図1のようになっていると仮定する。なお、IRTにおいては、受験者特性値と個々の項目への正答率との関係を表す図1中の曲線のことを「項目特性曲線 (item characteristic curve, 以下、ICC)」と呼ぶ。

1PLMにおいては、その項目に対する正答率が0.5 (50%) となる受験者特性値のこと、図1の例でいえば語学力のことを「困難度」と呼ぶ。たとえば、図1において語学力が平均である0と

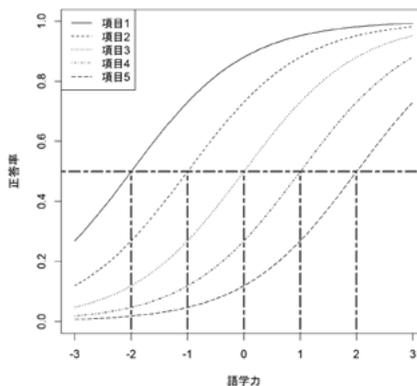


図1 語学力テストに含まれる5項目の1PLMに基づくICCと各項目の困難度

なる場合、つまり横軸が0の時の各項目のICCの値に着目してみると、困難度の高い項目ほどその正答率が低くなっていることがわかる。このように、困難度は、文字通り、各項目の難しさを表す値となっており、1PLMにおいては、この困難度が項目毎に異なることが想定されていることがわかる。また、図1からは、この困難度の値が高い項目ほどICCが右のほうにずれていくこともわかる。

加えて、1PLMにおいては、困難度における項目特性曲線の傾きに比例する値のことを「識別力」と呼ぶ。図1からは、各項目の困難度付近におけるICCの傾きが5つの項目間で等しくなっていることが伺える。つまり、1PLMにおいては、各項目の識別力が等しいことが想定されており、受験者特性値の増加に伴う困難度付近での正答率の上昇具合が全ての項目において等しいことが想定されている。

2パラメタ・ロジスティックモデル

2パラメタ・ロジスティックモデル (以下、2PLM) とは、1PLMと同じく、実際に項目反応データを解析する際によく利用される項目反応モデルである。ただし、1PLMとは異なり、項目間で困難度と識別力が異なることを想定しているモデルとなる。たとえば今、受験者の語学力を測定す



るための5つの項目からなるテストがあったとすると、2PLMにおいては、受験者特性である数学力と各項目への正答率との関係が図2のようになっていると仮定する。

1PLMと同様、2PLMにおいても、その項目に対する正答率が0.5 (50%) となる受験者特性値、図2の例でいえば数学力のことを「困難度」と呼ぶ。図2からは、2PLMにおいても、各項目の困難度が異なることを想定していることがわかる。また、2PLMにおいても、各項目の困難度付近におけるICCの傾きに比例する値のことを「識別力」と呼ぶ。図2からは、先述の通り、項目間でこの傾きが異なっていることが伺える。このように、2PLMにおいては、受験者特性値の増加に伴う困難度付近での正答率の上昇具合が項目によって異なることが想定されている。

3パラメタ・ロジスティックモデル

3パラメタ・ロジスティックモデル(以下、3PLM)も1PLMや2PLMと同じく、項目間で困難度と識別力が異なることを想定する項目反応モデルである。ただし、このモデルでは、それらに加えて、受験者特性値の低い受験者が偶然正答する、つまり、当て推量により正答する状況をも想定している。たとえば今、受験者の読解力を測定

するための5つの項目からなるテストがあったとする。ただし、これらの項目は4つの選択肢から1つの正答選択肢を選ぶタイプの項目となっており、読解力が低く本来正答する能力の無い受験者でも当て推量で正答する可能性があったとしよう。このような時、3PLMにおいては、受験者特性である読解力と各項目への正答率との関係が図3のようになっていると仮定する。

図3からは、各項目のICCの位置がずれ、その傾きも異なっており、3PLMにおいても、各項目の困難度と識別力は異なっていることが想定されていることが伺える。また、図3においては、読解力が低い受験者の正答率が、1PLMや2PLMを仮定した場合に比べて高くなる傾向にあることが伺える。このことから、3PLMを利用することにより、受験者特性値が低い場合の当て推量による正答が考慮できるようになっていることが伺える。

項目情報量

項目反応データと受験者特性値との関係を表すものとして上述のような項目反応モデルを仮定し、困難度や識別力などの項目母数を推定したら、次は、得られた項目母数の推定値を利用してテストの作成や評価、運用を行っていくこ

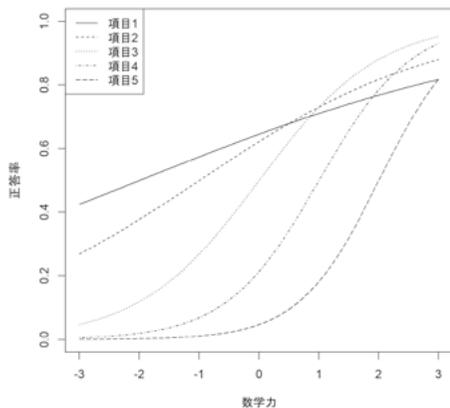


図2 数学力テストに含まれる5項目の2PLMに基づくICC

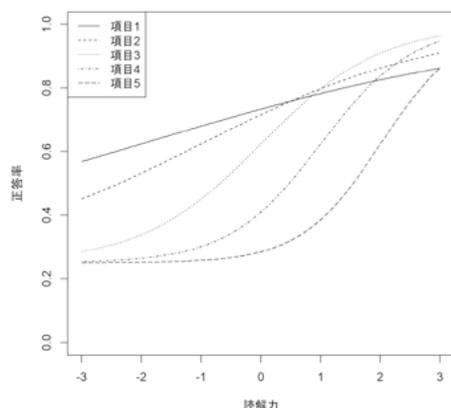


図3 読解力テストに含まれる5項目の3PLMに基づくICC

となるのだが、この時利用される統計量の1つに「項目情報量」と呼ばれるものがある。

項目情報量とは、その項目への解答結果から解答者の受験者特性値を推定した際の推定精度の指標となるものである。もう少し具体的にいうと、ある受験者特性値の最尤推定量の標準誤差の二乗の逆数を与えるものである。ここで、標準誤差とは、もう一度全く同じ条件下で同じ項目へ解答した場合に、その解答結果から受験者特性値を最尤推定すると推定値がどの程度変化するかを表すものである。つまり、この項目情報量を算出することにより、たとえば、受験者特性値が0の受験者に項目Aを出題した場合にはその解答データから得られる最尤推定量の標準誤差が0.1になり、1の受験者に出題した場合には0.2になる等、受験者特性値の水準毎にその推定精度を検討できるようになる。そして、項目情報量は受験者特性値の関数となっており、したがって、ICCと同様に、項目情報量と受験者特性値との関係はグラフとして表現することが可能である。このようなグラフのことを「項目情報量曲線 (item information curve, 以下、IIC)」という。たとえば今、2PLMの解説時に用いた、受験者の数学力を測定するための5つの項目からなるテストがあったとする。これらの項目に

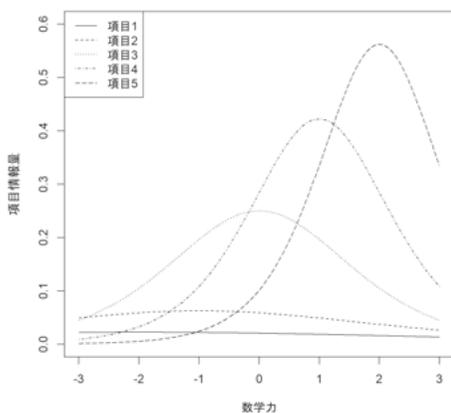


図4 数学力テストに含まれる5項目の2PLMに基づくIIC

対して、個々の項目の正答率と受験者特性値との間の関係が2PLMで表現できると仮定し、項目情報量を計算、IICを描いたとすると、以下の図4にあるような5つの曲線が得られることになる。

ここで、図4中の5つのIICのうち、特に項目3のIICに着目してみると、項目3の項目情報量は受験者特性値が0の時に最も大きくなっており、受験者特性値が0から離れていくにしたがって、項目情報量の値も小さくなっていくことが伺える。このことから、受験者特性値が0近辺の受験者に対して項目3を出題した場合、その解答データから受験者特性値を最尤推定した際の標準誤差は小さくなるが、0から離れた受験者に対して項目3を出題した場合には、最尤推定値の標準誤差は大きくなるであろうことが伺える。このことが示唆するように、一般に、項目にはその項目を出題するのに適した受験者特性値の水準が存在し、どのような受験者層に対しても有効な項目というのは存在しない。

3 項目反応理論に基づくコンピュータ適応型テスト(CAT)

図4に示したIICの形状等からは、「その受験者にとって適当な項目だけを出题することができれば、従来のテストより少数の項目であっても、同程度の精度で受験者特性値の推定ができるのではないか」という発想が生まれてくる。実は、このようなアイデアを実現するための方法の1つがIRTに基づくCATなのである。先に述べたように、本節では、このIRTに基づくCATの基本的な事項についての解説を行うが、紙幅の関係上、本稿で解説できる事項はごく限られたものにならざるを得ない。したがって、読者が実際にIRTに基づくCATを作成・運用しようとする場合には、更なる学修が必要になると考えられる。そういった場合には、たとえば豊田 (2012) やvan der Linden & Glas (2010) にて、IRTに基づくCATのまとまった解説がなされ



ているので、ぜひそちらも参照していただきたい。

まず、IRTに基づくCATを作成し運用する場合には、項目プールと呼ばれる項目母数が推定済みであるたくさんの項目の集合を用意する必要がある。たとえば、表1には、2PLMを仮定した場合の項目母数の推定値が得られている10項目 (item1から10) の項目名、識別力、困難度に関する情報がまとめられている (例示のために作成した仮想的なものである)。このような情報をデータベースの形でまとめたものを、一般に、「項目プール」と呼ぶ。もちろん、項目プールに収録される情報はこれらの情報に限られるわけではない。作成し運用しようとしているCATの仕様に合わせて、たとえば、各項目の問題内容が収録されることもあれば、その項目が出題された受験者の総数が収録されることもある。

このような項目プールが用意できたならば、次に、受験者がコンピュータ上でテストを受験し項目に解答できるシステムを構築することになる。たとえば、このようなシステムを構築する手段として、個人で利用できるものとしては、質問紙調査を行う際に利用されることの多くなったGoogleフォームなどが挙げられるかと思う。しかしながら、以下で解説する項目選択アルゴリズムの実装や解答データの安全な保管という観点からは、多くの受験者が見込まれるハイステークスなCATの作成・運用を行う場合などには、一からこのようなテスト受験システムを構築するのが望ましいと思われる。

表1. 項目プールの例

項目名	識別力	困難度
item 1	0.56	-0.98
item 2	1.32	-0.18
item 3	0.92	0.16
item 4	0.98	-0.73
item 5	0.52	1.30
item 6	0.70	-0.34
item 7	1.44	1.42
item 8	1.13	-0.04
item 9	1.15	-1.04
item 10	0.63	-0.95

そして、このようなテスト受験システムが構築できたならば、いよいよIRTに基づくCATが実施できるように、このシステムに対して、以下のようなフローが実行可能となるよう、改修を加えることになる。これらの手続きを経て構築されるシステムにより運用されるテストのことを、一般にCATと呼ぶ。

1. 受験を開始してから数間はシステム側であらかじめ決定しておいた項目に解答する
2. それまでの解答結果から、受験者の受験者特性値を推定する
3. 得られた受験者特性値に基づいて、次に出題する項目を選び、出題する
4. 以降、2と3の手続きを、テストの終了基準が満たされるまで繰り返す

なお、CATシステムを構築する際に上記1から4のステップの詳細をどのように定めるのかによって、最終的に利用可能となるCATとしては実に様々なパターンのものを考えることができる。たとえば、上記のステップ1に関していえば、初めの数問をa) 全ての受験者に対して同じ項目にするのか、b) 学年などの受験者の属性によって変更するのか、などといった様々な仕様の選択枝が考えられるし、ステップ4に関していえば、テストの終了基準をa) 10分経過したら終了といった制限時間に基づくものにするのか、b) 10問出題したら終了といった出題された項目数に基づくものにするのか、などといった様々な仕様の選択枝が考えられる。このように各ステップにおいて取り得る様々な選択枝の組み合わせを考えてみると、先に述べたように、最終的に利用可能となるCATとしては実に様々なものを考えることができるわけである。したがって、実際にCATを構築し運用する際には、無数にある選択枝の中から、実施者の運用目的に最も沿う仕様を選び出し、それを実現するためのCATシステムを構築していくことになる。

4

ベイズ統計学を利用した項目反応理論に基づく適応型テスト

IRTに基づくCATの仕様としては実に様々なものが考えられるわけだが、その中でも最も基本的な仕様の1つが、先のフローの3つ目のステップにおいて、「推定された解答者の受験者特性値においてその項目情報量が最大になる項目を次の問題として出題する」というものである。項目情報量の項で解説したように、このようにして次に出题する項目を選ぶことにより、その解答者にとって有効な項目をより多く出題することが可能となり、より精度の高い受験者特性値の推定が可能となる確率が高くなる。なお、この項目選択アルゴリズムを実行するためには、項目プールに含まれる各項目の受験者特性推定値における項目情報量を計算する必要があるわけだが、項目プールに含まれる各項目の項目母数は既に推定済みであるため、初めの数問に回答して受験者特性の暫定的な推定値が出た後であれば、この項目選択アルゴリズムは実行可能となる。

ところで、IRTに基づくCATを実施するにあたっては、先のフローの2つ目のステップにおいて、それまでの解答結果から受験者特性値の推定を行うことになるわけだが、この際、気を付けなければならないことがある。たとえば、最初の数問に解答した直後などにおいては、出題数の少なさゆえに解答結果が全問正答や全問誤答となることがあるわけだが、実は、このような場合に受験者特性の値を多くのデータ解析で使用される最尤推定すると、推定値が $+\infty$ あるいは $-\infty$ となってしまう。その結果、項目情報量の値も計算できなくなってしまい、ひいては、項目情報量に基づく項目選択、もといえはIRTに基づくCATも行えなくなってしまう。そこで、このような状況に対応するための方法が今までに色々と提案されてきた。その中の1つが「受験者特性値をベイズ推定する」

というものになる。

受験者特性値をベイズ推定する際には、最尤推定する時とは異なり、各項目への解答結果、つまり、尤度を計算するために必要な情報だけではなく、項目反応データを取得する前にデータの解析者が抱いていた受験者特性値に関する信念、たとえば受験者特性値がある値である確率は標準正規分布で表されるといったものに基づいて、受験者特性値の推定を行うことになる。もう少し具体的に説明すると、受験者特性値をベイズ推定する際には、ベイズの定理と呼ばれる条件付き確率の定理を利用して、各項目への解答結果等から計算された尤度と事前分布と呼ばれる事前に抱いていた受験者特性値に関する信念から、事後分布と呼ばれる項目反応データを取得した後の受験者特性値に関する信念を求め、この事後分布の平均(期待値)や最頻値をもって、受験者特性値の推定値とすることになる。このように、受験者特性値のベイズ推定においては、各項目への解答結果だけではなくデータ解析者が事前に抱いていた信念も受験者特性値の推定時に利用することになり、解答結果以外の情報も利用できるようになるため、解答結果が全問正答あるいは誤答の場合であっても有限の受験者特性値を推定することが可能になる。その結果として、項目情報量に基づく項目選択やIRTに基づくCATを実施することが可能となる。

このように、ベイズ統計学はIRTに基づくCATを行ううえでは欠かせないものとなり、したがって、アフターコロナ時代のテストを考えるにあたっては必要不可欠なものになると考えられる。

5 終わりに

本稿では、IRTの概要について説明し、これに基づくCATについてその概要を利点と共に解説して、ベイズ統計学がIRTに基づくCATにおいて果たす役割について言及した。



この原稿を執筆している現在、日本をはじめとする世界において新型コロナウイルスの脅威はくすぶり続けており、この感染症の流行を契機に、我々の日常は一変しようとしている。その行き着く先には、政府の述べる「新しい生活様式」が待っているであろう。

このような流れの中においては、最初に述べたように、当然、テストの在り方も変わってくるはずである。今までは、テストを実施する際、多くの受験者を同じ場所に集めて同じ時刻に一斉にテストを実施する形式を採ることが多かったわけだが、今後は、このような、感染リスクの高

い形式でのテスト実施は敬遠されることになるだろう。そして、その代わりに、個々の受験者が個別にテストを受験することが可能なオンラインテストが普及してくるのではないだろうか。このとき、より少ない項目数で受験者特性値を測定でき、なおかつ、受験者ごとに異なる項目群を出題することの可能なIRTに基づくCATは、受験者や実施者の負担といった観点からも必ずや注目を浴びることになるだろう。その際に、本稿が、IRTに基づくCATの導入を目指す方々にとって何らかのお役に立つのであれば、私としては望外の喜びである。

文献

- Arai, H. N., N. Todo, T. Arai, K. Bunji, S. Sugawara, M. Inuzuka, T. Matsuzaki and K. Ozaki, 2017, "Reading skill test to diagnose basic language skills in comparison to machines", *Proceedings of CogSci 2017*: 1556-1561.
- Lord, F. M. and M. R. Novick, 1968, *Statistical theories of mental test scores*, Reading: Addison-Wesley.

- 登藤直弥, 2019, 「テスト得点の分析——古典的テスト理論と項目反応理論」 繁杵算男他編著『公認心理師の基礎⑤——心理学統計法』 遠見書房: 233-259.
- 豊田秀樹, 2012, 『項目反応理論 [入門編] (第2版)』 朝倉書店。
- van der Linden, W. J. and C. A. W. Glas, 2010, *Elements of Adaptive testing*, New York: Springer.

3

現代的な パーソナリティ測定のための ベイズ統計モデリング

分寺杏介

東京大学大学院
教育学研究科 博士課程¹⁾

岡田謙介

東京大学大学院
教育学研究科 准教授

コンピュータ上で行う心理特性の測定には、回答に要した時間も収集できることや、回答バイアスの影響を受けにくい比較判断型項目を自然に実装ができることといった利点がある。従来の紙の調査票をそのままコンピュータに載せるのではなく、こうした特徴を活用して妥当性の高い測定を行えると望ましい。本稿ではこれを実現するための、ベイズ統計的アプローチに基づく項目反応理論モデルの拡張による私たちの研究を紹介する。

1 はじめに

コンピュータによる測定

計算機科学の発達により、コンピュータを用いた調査のニーズは年々増加している。2018年には、マーケティングの分野におけるインターネット調査の売上高はアドホック調査全体の50%に達している（日本マーケティング・リサーチ協会, 2019）。インターネット調査をはじめとしたコンピュータによる測定では、回答者の回答データに加えて、様々な付加情報を容易に入手することができる。その例としては、アクセス時間や画面遷移のログ、マウスカーソルの軌跡などが挙げられるが、中でも最も幅広く用いられている付加情報は回答時間であろう。

教育測定の分野では、回答時間のデータを回

答データと同時に分析するための方法論の研究が古くから行われてきた。特に「あてはまる」、「あてはまらない」の二肢選択の質問によるパーソナリティ測定では、回答者の特性と回答時間との関係性が経験的に知られている。例えば外向性を測定する項目として「盛り上げ役である」にあてはまるかどうかを尋ねると、極端に外向的あるいは極端に内向的な人は迷うことなくいずれかの選択肢を選ぶと考えられるが、中程度の外向性の人ほど回答を迷うと想定される。横軸に回答者の特性値を、縦軸に回答時間の期待値をとると左右対称な山形になることから、この関係は「逆U字 (inverted-U)」と呼ばれている。この逆U字の関係を踏まえると、回答時間が長い人ほど特性値が平均的な方向に近いと考えることができるため、例えば Ferrando & Lorenzo-Seva (2007) などはこの関係性を取り入れて、回答時間データを特性値推定に利用する項目反応理論 (item response theory, IRT) モデルを提案している。

回答バイアス

実際のパーソナリティ測定では、回答者が選んだ選択肢は回答者の心理特性以外の要因を反映していると考えられる。その代表例が回答バイアスである。例えば「全くあてはまらない」から「とてもあてはまる」までの5件法の質問項



目で判断に迷った際には、とりあえず中間の「どちらでもない」を選択する（中心化）傾向を持つ人がいる一方で、「あてはまる」側の選択をする（黙従）傾向の人もあるだろう。他にも、回答内容によって回答者が受ける処遇が異なる状況下（例えば入社試験など）では、意図的に自分をよく見せるように回答する可能性がある（Faking; Dunette et al., 1962）。

こうした傾向は、測定結果を歪める系統的なバイアスとなる危険性があるため、様々な対処法が研究されている（e.g., King et al. 2004; Paulhus et al. 2003）。回答バイアスの影響を抑制できるとされている方法の一つに、「強制選択型項目」の利用がある。通常の5件法などの項目（リッカート型項目）と異なり、強制選択型では複数の文を同時に提示する。例えば「盛り上げ役である」と「すぐに雑用を済ませる」の二文が同時に提示され、回答者は選択肢の中から自身によりあてはまる文を選ぶ。あるいは三肢以上の場合、すべての選択肢をあてはまる順にランキングするといった回答形式も見られる。本研究では多肢強制選択型項目の中でも、特に各文がそれぞれ異なる心理特性を反映した形式を扱う。例えば上記の例では「友達が多い」は外向性を、「すぐに雑用を済ませる」は勤勉性を反映した文である。先行研究では、強制選択型項目がリッカート型項目と比べて回答バイアスを抑制し、その結果測定の妥当性を高める可能性も示されている（e.g., Cao & Drasgow, 2019; Christiansen, Burns & Montgomery, 2005）。

強制選択型項目は認知心理学の分野では広く用いられている実験スキームである。例えば異なる明るさの2つの四角が同時に提示され「どちらがより明るい」かを瞬間的に判断する課題は、二肢強制選択項目となる。そのため、認知心理学ではこのような項目への反応データを分析するための様々なモデルが提案されてきた。特にDiffusionモデル（Ratcliff, 1978）に代表される「連続サンプリングモデル（sequential sampling

model）」は、項目が提示されてから実際に反応が発生するまでの認知過程を数理モデルによって表現する。連続サンプリングモデルでは時間の経過とともに進行する認知過程を数理的に表現しているため、回答時間データが必然的に利用される。しかし、連続サンプリングモデルは、元来心理特性の測定を目的として開発されたものではない。したがって、これを心理特性の測定に使用する場合には、モデルの拡張が必要となる。

本稿では、複数の分野の先行研究で提案されたモデルを融合して私たちが提案したモデルを紹介する。これは、強制選択型項目に対する回答データと回答時間データを同時に利用して回答者の心理特性を推定するためのモデルである。連続サンプリングモデルは一般に、正規分布などのよく知られた確率分布を用いる場合と比べて、データの生成メカニズムをより具体的に想定するため、パラメータの数が多くなる。特に本稿で紹介するモデルは各項目・回答者に複数のパラメータが設定され、合計のパラメータ数は膨大になる。この結果、尤度関数の微分は閉じた形にならない。しかし、マルコフ連鎖モンテカルロ（以下、MCMC）法を用いれば事後分布からの乱数発生を行うことができるため、計算時間はかかるがパラメータの推定値を得ることができる。以下の節で紹介する実データ解析例はいずれも確率的プログラミング言語Stanを用いたMCMC法によるベイズ推定を行い、得られた事後分布の平均値（EAP）を点推定値として用いている。

2 Thurstonian Diffusion IRT モデル

本節では、多肢強制選択型のパーソナリティ測定項目に対する回答データを分析するためのThurstonian IRT（以下、TIRT; Brown & Maydeu-Olivares 2011）モデルと、Diffusionモデルを組み合わせたThurstonian Diffusion IRTモデル（Bunji & Okada, 2020a）を紹介する。この2つのモデルの特徴を組み合わせることにより、回答時間デー

タを用いて、回答が生成されるまでの認知過程を考慮しつつ二肢強制選択型のパーソナリティ測定項目に対する回答データを分析することが可能となっている。はじめに本モデルのもととなった2つのモデルを導入し、その後本モデルを紹介する。最後に私たちが行った実データ解析の一例を示す。

Thurstonian IRT (TIRT) モデル

TIRTモデルは多肢強制選択型のパーソナリティ測定項目に対する回答データを分析するモデルであり、そのコンセプトはサーストンの比較判断の法則 (Thurstone, 1927) に基づいている。比較判断の法則では、複数の選択肢が提示されたときに実際にどの選択肢を選ぶかは、各選択肢の主観的選好の対比較によって決まると考える。確率的効用理論によれば、各選択肢に対する主観的効用は正規分布に従うランダムな値として定義される。具体的には、選択肢 k に対して回答者 i が持つ主観的効用 $u_i^{(k)}$ は一般的な因子分析モデルによる以下の形で表される：

$$u_i^{(k)} = \mu^{(k)} + \lambda^{(k)}\eta_i^{(k)} + \varepsilon_i^{(k)}, \quad \varepsilon_i^{(k)} \sim N(0, \psi^{(k)}).$$

ここで、 $\lambda^{(k)}$ は因子負荷を、 $\mu^{(k)}$ は切片項=選択肢 k の平均的な効用を表している。 $\eta_i^{(k)}$ は選択肢 k が反映している因子について回答者 i の因子得点を表している。この式は、因子負荷が正であれば、因子得点の値が大きい回答者ほど $u_i^{(k)}$ の値も大きくなり、その選択肢への選好が強いことを意味する。また、 $\varepsilon_i^{(k)}$ は平均0、分散 $\psi^{(k)}$ の正規分布に従う誤差項を表している。以後、主観的選好 $u_i^{(k)}$ のうち誤差項を除いた部分を $\delta_i^{(k)} = \lambda^{(k)}\eta_i^{(k)} + \mu^{(k)}$ と表記する。

以上の式から、TIRTモデルでは、選択肢対 (1), (2) が提示されると $u_i^{(1)}$ と $u_i^{(2)}$ の値を比較して大きい方が選択されると考える。 $u_i^{(1)}$ と $u_i^{(2)}$ にはそれぞれランダム要因 $\varepsilon_i^{(1)}$, $\varepsilon_i^{(2)}$ が含まれているため、モデル上では $u_i^{(1)} - u_i^{(2)} > 0$ と

る確率=選択肢1が選択される確率を

$$P(u_i^{(1)} - u_i^{(2)} > 0) = \Phi\left[\frac{\delta_i^{(1)} - \delta_i^{(2)}}{\sqrt{\psi^{(1)} + \psi^{(2)}}}\right]$$

と、標準正規分布の累積確率関数で表現することによって尤度関数を構成する。

Diffusionモデル

図1に、Diffusionモデルで仮定される回答決定プロセスの概念図を示す。Diffusionモデルでは単位時間ごとに各選択肢に対する情報が蓄積され、これが一定の量に到達した段階で回答が決定されると考える。回答決定に資する情報の蓄積量を表すもの (図中のジグザグに動いている線) は「アキュムレータ」と呼ばれる。はじめに項目が提示されると、アキュムレータは初期値 (図中の z) からスタートし、情報の蓄積に応じて上下に移動する。アキュムレータが一方の境界 (図中の a と 0) に到達すると、到達した側の選択肢が選ばれる。実際の回答時間は、アキュムレータが一方の境界に到達する時間に無反応時間 (符号化などの時間) を加えたものとして表れる。

Diffusionモデルには数種類のパラメータがある。ここでは、特に (Thurstonian) Diffusion IRTモデルを特徴づけるために重要なパラメータを2つ紹介する。1つはドリフト率と呼ばれる「単位時間あたりの平均的な情報の蓄積量」(図中の v) である。実際に単位時間ごとにアキュムレー

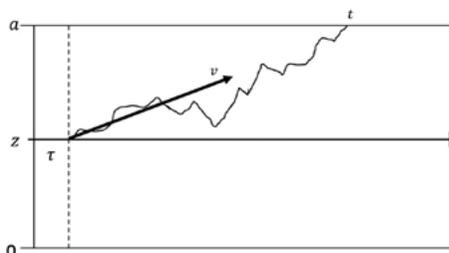


図1 Diffusionモデルで仮定される回答決定プロセスの概念図 (Adapted from Bunji & Okada, 2020a)



タが移動する量は、ドリフト率を平均値に持つ正規分布からランダムに決定されると考えるため、ドリフト率が正の値であればアキュムレータは平均的には上昇していく。また、ドリフト率の絶対値が大きくなるほどアキュムレータは素早く移動するため、境界に到達する時間が短くなり、結果的に回答時間が短くなると期待される。もう1つのパラメータは上側の境界（図中の a ）である。2つの境界間の距離が長くなるほど、回答にかかる時間が長くなる一方で、回答の正確さが高くなる。本来、ドリフト率が正であればアキュムレータは上側の境界に到達することが期待される。しかし、アキュムレータの移動量は正規分布からランダムに決定されるため、瞬間的にはアキュムレータが下側に移動する可能性が存在する。これを繰り返すと、ドリフト率が正であるにもかかわらずアキュムレータが下側の境界に到達する可能性もある。移動量はあくまでも確率的に決定されるため、2つの境界間の距離が長くなるほど、ドリフト率の符号とは反対方向の境界に到達する確率は低くなる。この結果、回答がドリフト率と整合的な形で行われる確率が高くなる。

Diffusionモデルにおけるアキュムレータの移動のプロセスは「ウィーナー過程」と呼ばれる。これはランダムウォークの連続時間への拡張である。アキュムレータが特定の境界に時間 t で初めて到達する確率の同時分布は、上述の2つのパラメータに加えて、初期値（本稿では $z = a/2$ に固定）や無反応時間などのパラメータによって規定される「ウィーナー分布 (Wiener first passage time distribution)」に従う。

Thurstonian Diffusion IRTモデル

Thurstonian Diffusion IRTモデルは、Diffusionモデルをもとに、ドリフト率と上側の閾値が回答者と項目の両者の要素によって決定すると考えている。ここでは、上側の境界は選択肢1に、下側の境界は選択肢2にそれぞれ対応している

とする。TIRTモデルにおける主観的効用の差 $u_i^{(1)} - u_i^{(2)}$ と、Diffusionモデルにおけるアキュムレータの移動量は、いずれも正規分布に従うランダム要因を含んでいる。ここから、Diffusionモデルのドリフト率は自然な形でTIRTモデルの $\delta_i^{(1)} - \delta_i^{(2)}$ に置き換えることができる。上側の境界の位置は、回答者と項目の要素との商の形で表す。ここでの回答者と項目の要素は、それぞれ回答者の慎重さや項目の複雑さなどを表しているともみなすことができる。

以上より、Thurstonian Diffusion IRTモデルの尤度関数は、ドリフト率のパラメータを $\delta_i^{(1)} - \delta_i^{(2)}$ 、上側の境界の位置のパラメータ a を回答者と項目の要素との商の形で与えたウィーナー分布で表現することができる。

実データ分析例

以下では、Bunji & Okada (2020a)による実データ分析例を紹介する。分析に使用したデータは、クラウドソーシングで集めた20代から70代の日本人500名に対してオンライン調査を行い収集された。調査に使用した項目は、Big-Five尺度の日本語版 (Apple & Neff, 2012) の50項目をもとに、異なる因子を測定する文を組み合わせで作成した二肢強制選択項目 (25項目) を用いた。調査では、画面上に2つの文が同時に提示され、回答者は「自分に最もよくあてはまる」文に対応するキーを押して回答した。調査のフレームワークはjsPsych (de Leeuw, 2015) によって作成したもので、一つひとつの項目に対する回答データと回答時間が記録された。

収集したデータを、Thurstonian Diffusion IRTとTIRTの両モデルに適用し、MCMC法によるベイズ推定によってパラメータの推定値を得た。図2に、Big-Fiveのうち外向性の因子得点の推定値の散布図を示した。横軸はTIRTモデル、縦軸はThurstonian Diffusion IRTモデルによる推定値を表している。外向性だけでなくBig-Fiveの全5因子について両モデルでの推定値は高い相

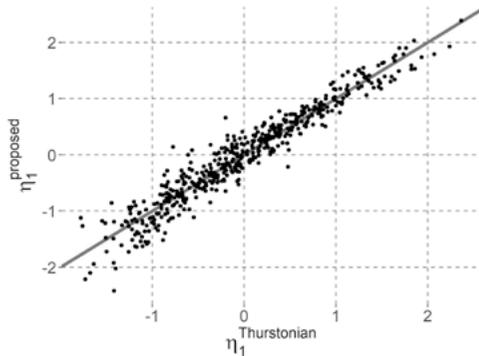


図2 外向性の因子得点の推定値の散布図 (横軸:TIRTモデル, 縦軸:Thurstonian Diffusion IRTモデル)
(Adapted from Bunji & Okada, 2020a)

関を示した ($r=0.954-0.970$)。また、因子負荷についても両モデル間での相関は0.902と十分に高い値を示した。これらの結果から、Thurstonian Diffusion IRTモデルは、回答時間データを用いていないTIRTモデルと比べても基本的には同じ特性を推定していると考えられる。

続いて、回答時間の利用による具体的な推定値の変化を検証した。回答時間の逆U字の関係を本モデルの枠組みにあてはめると、回答時間が長いほど2つの選択肢に対する主観的選好が近いと予測される。ここから、回答時間を使用しないTIRTモデルによる推定値と比べると、回

答時間を使うモデルによる推定値では、回答時間が短いほど選択された文に対応する因子の得点が高くなると考えられる。以上を踏まえて、本分析では「MSSRT」という指標を用いた。具体的な手順はBunji & Okada (2020a) に示されているが、簡潔に示すとMSSRTは各因子の文を含んでいる項目の(標準化された)回答時間の平均値である。例えば外向性因子の文を素早く選択している場合、外向性得点は高いと考えられ、この場合MSSRTは小さな値をとる。外向性因子の文を選択していない場合には、外向性得点が高いほど他の選択肢と判断に迷うために回答時間が長くなると考えられる。そのため、MSSRTでは当該因子の項目を選択しなかった場合の(標準化)回答時間は符号を反転させた上で平均値の計算を行う。こうして得られたMSSRTは、小さな値をとるほどその特性を強く持っていると想定される指標となっている。

図3の散布図は、縦軸に両モデルによる特性値の推定値の差を、横軸にMSSRTを示している。図からはこれらに負の相関が明らかに見られる。この結果から、回答時間を利用するThurstonian Diffusion IRTモデルでは、推定値が回答時間と整合的な方向に変動していることが示された。

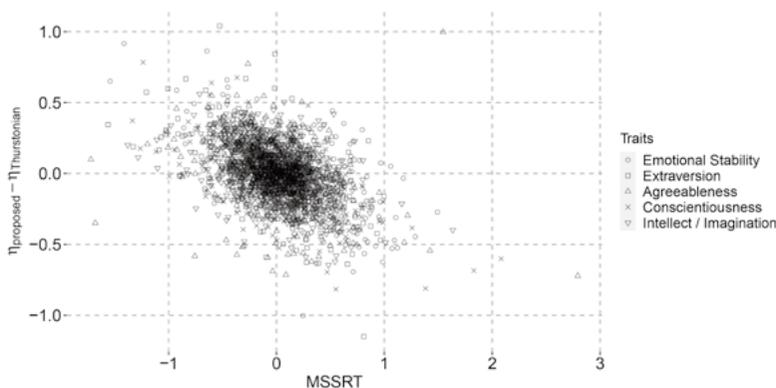


図3 特性値の推定値の差とMSSRTの散布図
(Adapted from Bunji & Okada, 2020a)



3 LBA IRTモデル

本節では、Diffusionモデルと同様、連続サンプリングモデルの一つであるLinear Ballistic Accumulator (以下、LBA) モデル (Brown & Heathcote, 2008) をもとにしたLBA IRTモデル (Bunji & Okada, 2019) を紹介する。Diffusionモデルと比較した際のLBAモデルの最大の特徴は、三肢以上の多肢強制選択項目にも対応している点である。はじめにLBAモデルを紹介し、その後LBA IRTモデルを紹介する。続けて私たちによる実データ分析例を示す。

LBAモデル

図4に、LBAモデルで仮定される回答決定プロセスの概念図を示す。項目が提示されると各選択肢に対する情報が蓄積され、これに応じてアキュムレータが移動するという点はDiffusionモデルと変わらない。Diffusionモデルとの大きな違いとして、LBAモデルでは選択肢ごとに異なるドリフト率の独立したアキュムレータが設定され、各アキュムレータの移動量は時間不変という仮定が置かれる。実際の各アキュムレータの単位時間あたりの移動量(傾き)は、ドリフト率を平均値に持つ正規分布などからのランダムな値となる。また、各アキュムレータの初期値はDiffusionモデルと異なり、0からAの範囲の一様乱数によって決定される。

LBAでは、境界は上側に一つのみ(図中の β)設定され、最初に境界に到達したアキュムレータに対応する選択肢が選択される。Diffusionモデルと同様に、本来はドリフト率の最も高い選択肢が最初に境界に到達することが期待される。しかし、初期値は一様分布から発生するため、境界が低いほど、ドリフト率の低い選択肢のアキュムレータが最初に境界に到達する確率が高くなる。したがって、LBAモデルでも境界パラメータの値が高いほどドリフト率と整合的な回

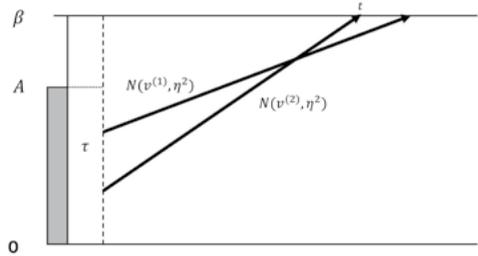


図4 LBAモデルで仮定される回答決定プロセスの概念図 (Adapted from Bunji & Okada, 2019)

答が得られる確率が高くなる一方で、回答にかかる時間は長くなる。先行研究では、境界パラメータを含むいくつかのパラメータの推定値が一对一の関係にあることが示されており (Donkin et al., 2011), こうしたパラメータの解釈自体はDiffusionモデルと同様に考えることができる。

LBAモデルにおける特定の選択肢のアキュムレータが時間 t で境界に到達する確率密度は、すでにBrown & Heathcote (2008) によって定式化されている。実際には最初に境界に到達したアキュムレータのみが関心の対象であるため、時間 t で選択肢 k を選択する確率密度は「ちょうど時間 t で選択肢 k が境界に到達する確率密度」と「時間 t までに k 以外の選択肢が境界に到達していない確率」の積で表される。

LBA IRTモデル

LBA IRTモデルは、Diffusion IRTモデルのアナロジーとして考え出された。すなわちDiffusion IRTモデルと同様に、LBAモデルをもとに尤度関数を設定し、ドリフト率と境界パラメータをそれぞれ項目と回答者の要素に分解したモデルである。Bunji & Okada (2019) では、「あてはまる」、「あてはまらない」の二肢選択式のパーソナリティ測定項目を対象としてモデルを提案したため、ドリフト率を回答者特性値と項目の閾値の差の関数で表現している。この定式化では多肢強制選択へ拡張することができないため、本稿ではドリフト率を異なる形

で定式化する (Bunji & Okada, 2020b)。

LBA モデルにおけるドリフト率は選択肢ごとに規定されるものであることから、ドリフト率を主観的効用 $\delta_i^{(k)}$ の関数とする。LBA モデルでは、モデルの識別のためにパラメータの制約を課す必要がある。制約の例としてはすべての選択肢のドリフト率の和を1に固定するといったものがよく用いられている。そこで、総和が1になる関数としてソフトマックス関数をあてはめて、ドリフト率を

$$v_i^{(k)} = \frac{\exp(\delta_i^{(k)})}{\sum_c \exp(\delta_i^{(c)})}$$

という形で表現する。境界パラメータについては、(Thurstonian) Diffusion IRT モデルと同様に回答者と項目の要素との商の形で表すこととした。

実データ分析例

以下では、Bunji & Okada (2020b) による実データ分析例を紹介する。クラウドソーシングで集めた20代から70代の日本人484名に対してオンライン調査を行い、データを収集した。調査に使用した項目は、Big-Five尺度の短縮版(Mini-IPIP; Donnellan et al., 2006) 5因子20項目を日本語化したものと、バランス型社会的望ましさ反応尺度日本語版(谷, 2008) 4項目をもとに、四肢強

制選択項目を18項目作成した。調査では、画面上に4つの文が同時に提示され、回答者は「自分に最もよくあてはまる」文に対応するキーを押して回答した。

収集したデータを、LBA IRTと多次元名義反応モデル (Revuelta, 2014) の両モデルに適用し、MCMC法によるベイズ推定によってパラメータの推定値を得た。多次元名義反応モデルは、LBA IRTモデルのドリフト率パラメータがそのまま各選択肢の選択確率となっているため、回答時間を使わないモデルとしてLBA IRTモデルの比較対象となる。

図5は、横軸に境界パラメータの推定値を、縦軸に平均回答時間をとった散布図である。2つの図はそれぞれ項目側要素(左図)、回答者側要素(右図)に着目したものである。左図では、項目側要素パラメータが大きい項目ほど回答時間の平均値が小さく、このパラメータが確かに回答時間に影響を与える項目側要因を反映していることが分かる。実際に、回答時間に影響を与える要因の一つと思われる「文のリーダビリティ」(Hasebe & Lee, 2015) の項目平均(各項目に含まれる4文のリーダビリティスコアの平均値)と項目側要素パラメータの推定値の相関係数は0.256であり、読みやすい文で構成された項目ほど平均回答時間が短い傾向が示された。同様に、

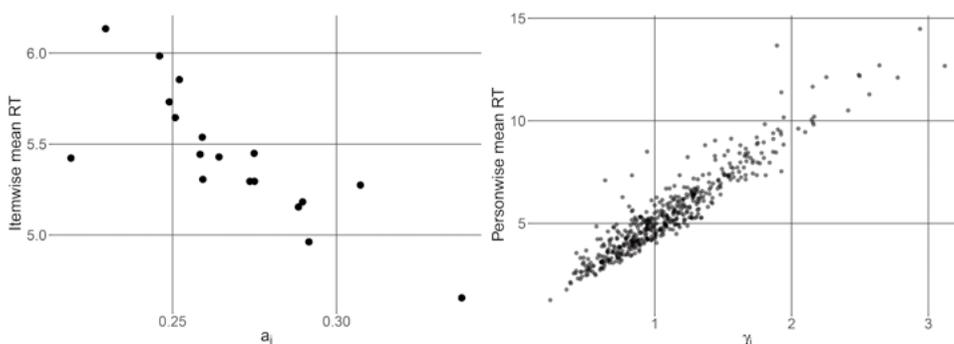


図5 (左図) 縦軸: 項目ごとの平均回答時間, 横軸: 境界の項目側要素パラメータの推定値
(右図) 縦軸: 回答者ごとの平均回答時間, 横軸: 境界の回答者側要素パラメータの推定値
(Adapted from Bunji & Okada, 2020b)

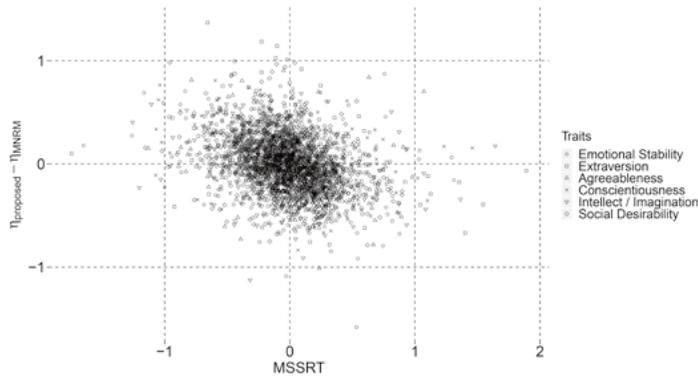


図6 特性値の推定値の差とMSSRTの散布図
(Adapted from Bunji & Okada, 2020b)

回答者側要素パラメータと平均回答時間の間にも高い相関が見られており、このパラメータが回答時間に影響を与える心理特性など (e.g., 慎重さ) を反映したパラメータである可能性を示している。

続いて、回答時間を用いたことによる具体的な推定値の変化を確認した。図6の散布図は、縦軸に両モデルによる特性値の推定値の差を、横軸にMSSRTを示している。因子ごとに算出した相関係数は-0.48から-0.39であった。この結果から、LBA IRTモデルでは回答時間データを使用することによって、推定値が回答時間と整合的な方向に変動していることが示された。

4 おわりに

本稿では、多肢強制選択型に対する回答データと回答時間データとを同時に利用して回答者

の特性を推定するためのモデルを紹介した。パーソナリティ測定の商品への回答行動の背後にある認知過程にDiffusionモデルやLBAモデルのような連続サンプリングを想定することには、ドリフト率のような心理学的に解釈できるパラメータが得られ、認知心理学の知見と教育・心理測定における知見を接続できるといった利点がある。

現代では、本稿で紹介したように、データの生成メカニズムを反映した複雑なモデルであっても、尤度関数を構成できればMCMC法などを用いてそのパラメータの推定値を得ることが比較的容易になっている。その一方で、「パラメータを推定できる」と「モデルが妥当である」ことは無関係である。そのため、作成したモデルの背後にある仮定および推定されたパラメータの妥当性については、今後も検証を重ねる必要がある。

注

1) 現所属：ベネッセ教育総合研究所

文献

Apple, Matthew T. and Peter Neff, 2012, "Using Rasch Measurement to Validate the Big Five Factor Marker Questionnaire for a Japanese University Population", *Journal of Applied Measurement*, 13(3): 276-296.

- Brown, Anna and Alberto Maydeu-Olivares, 2011, "Item Response Modeling of Forced-Choice Questionnaires," *Educational and Psychological Measurement*, 71(3): 460-502.
- Brown, Scott D. and Andrew Heathcote, 2008, "The Simplest Complete Model of Choice Response Time: Linear Ballistic Accumulation," *Cognitive Psychology*, 57(3): 153-178.
- Bunji, Kyosuke and Kensuke Okada, 2019, "Item Response and Response Time Model for Personality Assessment via Linear Ballistic Accumulation", *Japanese Journal of Statistics and Data Science*, 2(1): 263-297.
- Bunji, Kyosuke and Kensuke Okada, 2020a, "Joint Modeling of the Two-Alternative Multidimensional Forced-Choice Personality Measurement and Its Response Time by a Thurstonian D-Diffusion Item Response Model", *Behavior Research Methods*, 52(3): 1091-1107.
- Bunji, Kyosuke and Kensuke Okada, 2020b, "Linear Ballistic Accumulator Item Response Theory Model for Multidimensional Multiple-Alternative Forced-Choice Measurement of Personality [Manuscript submitted for publication]" .
- Cao, Mengyang and Fritz Drasgow, 2019, "Does Forcing Reduce Faking? A Meta-Analytic Review of Forced-Choice Personality Measures in High-Stakes Situations." *Journal of Applied Psychology*, 104(11): 1347-1368.
- Christiansen, Neil, Gary Burns and George Montgomery, 2005, "Reconsidering the Use of Forced-Choice Formats for Applicant Personality Assessment", *Human Performance*, 18(3): 267-307.
- Donkin, Chris, Scott Brown, Andrew Heathcote, and Eric Jan Wagenmakers, 2011, "Diffusion versus Linear Ballistic Accumulation: Different Models but the Same Conclusions about Psychological Processes?", *Psychonomic Bulletin and Review*, 18(1): 61-69.
- Donnellan, M. Brent, Frederick L. Oswald, Brendan M. Baird and Richard E. Lucas, 2006, "The Mini-IPIP Scales: Tiny-yet-Effective Measures of the Big Five Factors of Personality." *Psychological Assessment*, 18(2): 192-203.
- Dunette, Marvin D., Jean McCartney, Howard C. Carlson and Wayne K. Kirchner, 1962, "A Study of Faking Behavior on a Forced-Choice Self-Description Checklist", *Personnel Psychology*, 15(1): 13-24.
- Ferrando, Pere J. and Urbano Lorenzo-Seva, 2007, "An Item Response Theory Model for Incorporating Response Time Data in Binary Personality Items", *Applied Psychological Measurement*, 31(6): 525-543.
- Hasebe, Yoichiro and Jae-Ho Lee, 2015, "Introducing a Readability Evaluation System for Japanese Language Education", *Proceedings of the 6th International Conference on Computer Assisted Systems for Teaching & Learning Japanese (CASTEL/J)*, 19-22.
- King, Gary, Christopher J. L. Murray, Joshua A. Salomon and Ajay Tandon, 2004, "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research", *American Political Science Review*, 98(1): 191-207.
- de Leeuw, Joshua R, 2015, "jsPsych: A JavaScript Library for Creating Behavioral Experiments in a Web Browser", *Behavior Research Methods*, 47(1): 1-12.
- 日本マーケティング・リサーチ協会, 2019, 「第44回経営業務実態調査」(2020年6月23日取得, <http://www.jmra-net.or.jp/Portals/0/trend/investigation/gyoumujitai2019.pdf>)。
- Paulhus, Delroy L., P. D. Harms, M. Nadine Bruce and Daria C. Lysy, 2003, "The Over-Claiming Technique: Measuring Self-Enhancement Independent of Ability", *Journal of Personality and Social Psychology*, 84(4): 890-904.
- Ratcliff, Roger, 1978, "A Theory of Memory Retrieval", *Psychological Review*, 85(2): 59-108.
- Revuelta, Javier, 2014, "Multidimensional Item Response Model for Nominal Variables", *Applied Psychological Measurement*, 38(7): 549-562.
- 谷伊織, 2008, 「バランス型社会的望ましき反応尺度日本語版(BIDR-J)の作成と信頼性・妥当性の検討」『パーソナリティ研究』17(1) : 18-28.
- Thurstone, L. L., 1927, "A Law of Comparative Judgment", *Psychological Review*, 34(4): 273-286.



特集論文

4

ベイズ・モデリングを利用したマーケティングの高度化

宮崎 慧

関西大学商学部 教授

1 マーケティング・マネジメント・プロセスと現代のデータ環境

現代のマーケティング環境では市場の動向に関する情報が多く存在し、売上・販促データや顧客データの普及にともないマーケティング・データ分析手法も発展している。そこで本稿では上記テーマに倣い、まずマーケティングの基本的なプロセスであるマーケティング・マネジメント・プロセスを概説し、データ分析手法の一つであるベイズ・モデリングがこのプロセスの中でどのように活用されてきたかを述べる。

マーケティング・マネジメント・プロセスとはマーケティング活動を進める上での一般的な手順のことであり、図1のようにまとめられる。まずビジネス・チャンス（事業機会）の発見から始まり（リサーチ）、次にリサーチ結果に基づきマーケティング戦略を決定する。マーケティング戦略は消費者の特徴を把握するセグメンテーション、標的とする消費者を決めるターゲティング、消費者が製品やサービスをどのように認知しているかを探るポジショニングの3段階に分けられる。その後にはマーケティング戦術（マーケティング・ミックス）を決定する。価格の決定、チャネルの決定、製品やサービスの仕様の決定、プロモーションの決定の4つに分けられる。次にテスト販売（導入・実行）に入り、新製品や新サービスの導入結果のフィードバック・マーケティング・ミックスの変更（コントロール）と続く。

1960年代以前の日本は需要過多で供給不足の状況であり、すべての消費者に均一なマーケティング活動を行う「マス・マーケティング」が主流であった。その後一般消費財が各世帯に一通り行き渡って供給過剰にシフトしたため、消費者の異質性をより正確に捉えたマーケティング活動が必要となった。そこで、年齢・性別・住所などの特徴の類似した同質な消費者に細分化、つまりセグメンテーションを行い、それぞれの消費者グループに対して適切なマーケティング活動が展開された。

さらに現代ではセグメンテーションを推し進め、一人ひとりの消費者に適切なマーケティング活動を行うワン・トゥ・ワン・マーケティングの時代に入ったと言われている。現代の市場取引においては機械的にデータを収集するシステムを用いて、各消費者がいつ、どのような製品・サー

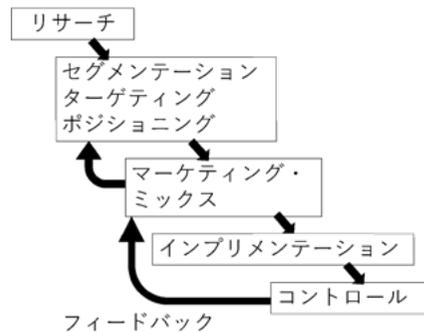


図1 マーケティング・マネジメント・プロセス

ビスを購入したかに関する大規模なデータが手に入る環境にあり、集まったデータを用いて個々の消費者に最も効果的なクーポンの種類や配布時期を決定するなどの、さらに細分化されたマーケティング施策が実行されている。

POSデータやスキャナー・パネル・データはワン・トゥ・ワン・マーケティングに頻繁に活用されてきた代表的なデータである。POSデータとは、消費者が買った商品の情報をレジにて記録し、集計した結果を在庫管理などに役立てるシステム（POSシステム）によって記録されたデータを指す。それぞれの店舗で買われた商品の種類、買われた時間、価格や数量、天気（気温、湿度など）やプロモーション（店頭陳列・チラシ配布など）に関する情報が記録される。マーケティング・ミックスは原因に、売り上げは結果に対応するため、コーザル（因果）データとも呼ばれることもある。

スキャナー・パネル・データとは商品のバーコードをスキャナーで読み取り、その情報をオンラインで記録した消費者の購買履歴データのことであり、商品に購買した消費者が紐づいたデータである。消費者のデモグラフィック情報などが記録される場合もある。これらのデータには個々の消費者の購買行動と一つ一つの製品のマーケティング・ミックスに関する情報が記

録されているため、ワン・トゥ・ワン・マーケティングを行う際に有用とみなされている。具体的なイメージ例としてアメリカの大手リサーチ会社 Information Resources Inc. が提供しているPOSデータとスキャナー・パネル・データを図2に示している。

POSデータやスキャナー・パネル・データなどの、商品一つ一つについての各消費者の購買履歴情報が記録されたデータを用いれば、最も細分化された在庫管理単位レベルでの購買予測や管理が可能となるはずである。加えて現在の成熟したマーケットでは、既存顧客の2割が売上の8割を構成するというパレート法則ののっとり、既存顧客の価値を最大化するために既存顧客との関係性をきめ細かく管理する「顧客関係管理」の考え方が一般的となっており、この方向性を実現するための指標として、企業から見た顧客の生涯価値の指標である顧客生涯価値の利用が重要視されている（照井、2008）。つまり現在は顧客ごとに算出される利益の期待値などのメトリクスを活用し、各顧客との関係性を管理すべき状況となっている。

現代のマーケティングでは、優良顧客を獲得し、かつ顧客一人ひとりとの良好な関係を長く維持することで競合優位に立てると考えられている。

IRI_KEY	WEEK	SY	GE	VEND	ITEM	UNITS	DOLLARS	F	D	PR	PANID	WEEK	UNITS	OUTLET	DOLLARS	IRI_KEY	COLOR	COLUPC
234212	1374	0	1	18200	100	1	8.49	NONE	0	0+	1127985	1376	1	GR	6.99	213290	0011820000614	↓
234212	1374	0	1	18200	8983	1	5.99	NONE	0	0+	1311035	1386	1	GR	6.99	213290	0011820000614	↓
234212	1374	0	1	18200	10985	2	11.98	NONE	0	0+	1107581	1426	1	GR	4.49	213290	0011820000784	↓
234212	1374	0	1	18200	468	13	19.37	NONE	0	0+	1168070	1389	1	GR	4.99	213290	0011820000784	↓
234212	1374	0	1	18200	53047	4	39.16	NONE	0	0+	1168070	1395	1	GR	4.99	213290	0011820000784	↓
234212	1374	0	1	18200	53168	40	522.80	A	0	1+	1201103	1380	1	GR	4.99	213290	0011820000784	↓
234212	1374	0	1	18200	769	12	114.48	NONE	1	0+	1228056	1401	1	GR	4.99	213290	0011820002992	↓
234212	1374	0	1	18200	833	12	57.48	NONE	0	0+	1311035	1390	1	GR	4.99	213290	0011820002992	↓
234212	1374	0	1	18200	53308	41	491.59	NONE	0	1+	1193516	1398	1	GR	9.99	213290	0011820003982	↓
234212	1374	0	1	18200	466	44	65.56	NONE	0	0+	1193516	1422	1	GR	8.99	213290	0011820003982	↓
234212	1374	0	3	18200	42	1	5.29	NONE	0	0+	1311035	1412	1	GR	10.49	213290	0011820003982	↓
234212	1374	0	1	18200	16	12	63.48	NONE	0	0+	1311035	1414	1	GR	10.49	213290	0011820003982	↓
234212	1374	0	1	18200	11047	4	39.26	NONE	0	0+	1127985	1374	1	GR	8.99	213290	0011820079047	↓
234212	1374	0	1	18200	11168	23	298.77	A	0	1+	1127985	1374	1	GR	8.99	213290	0011820079047	↓
234212	1374	0	1	18200	771	13	120.87	NONE	0	0+	1127985	1376	1	GR	9.99	213290	0011820079047	↓
234212	1374	0	1	18200	834	4	19.96	NONE	0	0+	1127985	1380	1	GR	9.99	213290	0011820079047	↓
234212	1374	0	1	18200	11308	14	171.26	NONE	1	0+	1127985	1381	1	GR	8.99	213290	0011820079047	↓
234212	1374	0	1	18200	96671	3	38.97	A	0	0+	1127985	1383	1	GR	9.99	213290	0011820079047	↓
234212	1374	0	1	18200	7986	7	64.53	NONE	1	0+	1127985	1384	1	GR	9.99	213290	0011820079047	↓
234212	1374	0	1	18200	96665	9	107.91	NONE	1	0+	1127985	1386	1	GR	9.99	213290	0011820079047	↓
234212	1374	0	1	18200	7985	6	28.14	NONE	0	0+	1127985	1387	1	GR	9.99	213290	0011820079047	↓
234212	1374	0	1	18200	51	12	45.48	NONE	0	0+	1127985	1388	1	GR	9.99	213290	0011820079047	↓
234212	1374	0	1	18200	61047	6	37.74	NONE	0	0+	1127985	1389	1	GR	9.99	213290	0011820079047	↓
234212	1374	0	1	18200	61030	6	71.94	B	0	1+	1127985	1389	1	GR	9.99	213290	0011820079047	↓

図2 POSデータ(左)とスキャナーパネルデータ(右)例



そのためには、消費者一人ひとりの購買行動の情報を有効に活用するきめの細かいマーケティング活動が必須である。したがって、企業は非集計データの分析を通して全消費者の平均的な特徴や似た特徴を持つ消費者セグメントをさらに詳しく細分化し、各消費者の個々の製品・サービスの購買行動をより正確に把握するリサーチ課題に取り組むことになる。

この高度なりサーチ課題を解決するためには、多くの変数を持つ新しいタイプのデータに対しても、各消費者の特徴、マーケティング・ミックスおよび需要の間の複雑な因果関係を捉えることができ、かつ得られた知見を用いて逐次購買行動の予測精度を高められる統合的なモデリング・アプローチが必要となる。そこでデモグラフィック特性などの消費者一人ひとりの特徴を記録したデータと購買履歴データの両方の情報を柔軟に取り込むことのできるベイズ・モデリングが注目され、活用されている。

2 階層ベイズ・モデリングの活用

今日のマーケティング環境では、市場取引で機械的に集まる消費者の購買行動データに対応でき、同時に実務的な示唆を導き出せるデータ分析アプローチが求められる。特に1節で述べたように、個々の消費者に対する適切なマーケティング対応、つまりワン・トゥ・ワン・マー

ケティングを可能とするアプローチが求められている。消費者の異質性の推定結果を用いて、個々の消費者に適切なマーケティング活動を行うことが求められているのである。

しかし、市場取引に関する膨大なデータを入力できたとしても、消費者ごとに個別の安定した推定を行うことは統計学的に不可能である。そこで、階層構造を加えることで消費者の異質性と購買行動の平均構造も同時にモデル化できる「階層ベイズ・モデリング」が利用されている。

図3はマーケティング・データ分析における階層ベイズ・モデルの一例を視覚化した図である。マーケティング・ミックス（価格、チャネル、製品やサービスの仕様、プロモーション）に対する各世帯の反応を、変量効果を用いてパラメータとして表現し、さらに各世帯のパラメータに対して分布を仮定する。これにより階層構造を持つモデル（階層モデル）となる。各世帯のパラメータが従う分布は世帯全体の特徴を表現している。

さらにベイズ統計学の考え方を用いて構築された階層モデルのことを階層ベイズ・モデリングという。ベイズ統計学ではパラメータを確率変数と考えるため、世帯全体の特徴をあらわすパラメータも確率変数として扱われる。階層ベイズ・モデリングにより、世帯全体の特徴を表す分布を仮定した上で各世帯の異質性をもパラメータとして推定でき、世帯全体と各世帯のパラメータを同時に安定的に推定できる。

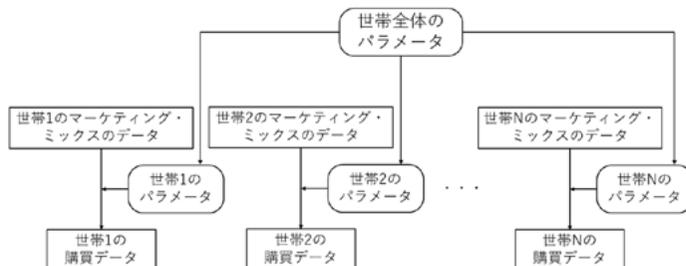


図3 マーケティング・データ分析における階層ベイズ・モデリング例の図解

Rossi, McCulloch & Allenby (1996) では階層ベイズ・モデリングを用い、シミュレーションによって個人ごとでの最適な価格戦略を示しており、様々な製品カテゴリーやサービスについて市場全体の一般化可能な予測を行う際に、階層ベイズ・モデリングが有効であると主張している。

マーケティングにおける分析手法と、取得されるデータの変遷を図4に示した。この図が示すようにPOSデータやスキャナー・パネル・データ、インターネット・データの取得が可能となることで、マーケティング変数と購買行動を結び付けると同時に、消費者異質性を表現できる階層ベイズ・モデリングが盛んに利用されてきた。前者にはロジット／プロビット・モデルなど離散選択モデルが、後者は潜在クラスや変量効果によって表現されることが多い。

マーケティング施策に対する反応は個々の消費者で異なるため個々の消費者ごとに推定するのは不可能であるが、階層ベイズ・モデリングの利用により、背後に異質性を表現する分布を仮定し(階層モデル)、そのパラメータを求めることにより消費者の異質性を取り込める。また、消費者の購買行動に関するマーケティング変数の効果も平均構造として取り込み処理することができる。購買予測はマーケティングにおいて

重要な要素となるが、ベイズ・モデリングではパラメータ推定と目的変数の予測についてベイズの定理に基づいて一貫した手続きが取られるため、ワン・トゥ・ワン・マーケティングを行う際に有効な分析体系を備える。

また同じく図4に示されているように、分析対象のマーケティング・データも取得環境が複雑化かつ多様化している。ユーザーがウェブ・サイト閲覧中に表示したページのシーケンスまたはパスに関する情報を提供するクリック・ストリーム・データや、ユーザーから発信される製品やサービスのレビューやブログなどのSNS情報、映像などの出現によりデータ量と種類はさらに増加している。2004年のフェイスブックの立ち上げによりソーシャル・ネットワーク・データの時代が開かれ、また2005年のYouTubeの登場によりユーザーがアップロードした膨大なテキストや映像記録データが行動ターゲティングに用いられる情報の源泉となっている。現代のマーケティング環境では、シンプルな構造や異質性を表さない少数のパラメータのみしか持たないモデルをビッグ・データに適用しても求める解は得られない。階層ベイズ・モデリングはこの問題に対応し、新しい知見を提供する有効なモデルと見なされている。

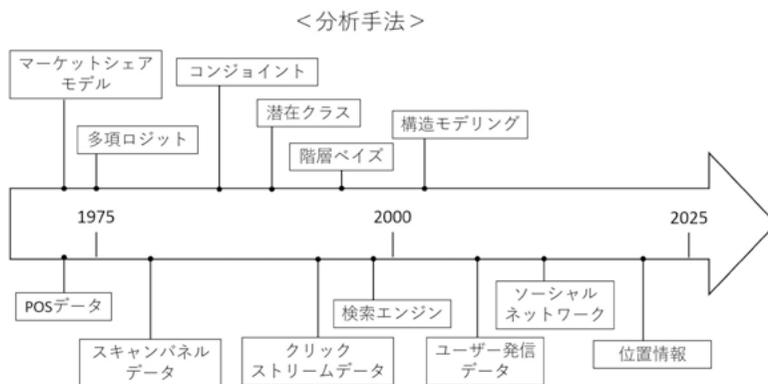


図4 マーケティングにおける分析手法とデータの変遷 (Wedel & Kannan (2016) の図を一部改変)



3 ベイズ・モデリングを用いた 代表的な先行研究

ベイズ・モデリングはマーケティングにおける強力なアプローチの一つであり、消費者行動の統合モデルを開発するために有用なフレームワークを提供できると言える (Rossi & Allenby, 2003)。そこで本節ではベイズ・モデリングを用いたマーケティング・データ分析の代表的な先行研究のレビューを通して、今日までベイズ・モデリングがどのように応用されてきたかを説明する。

POSデータ、スキャナー・パネル・データに対して階層ベイズ・モデリングを適用した研究は非常に多く報告されている。Allenby & Lenk (1994) は標準的なロジスティック回帰モデルを拡張し、消費者の嗜好を表現するために誤差効用部分に一次の自己回帰モデルを仮定し、状態依存性を考慮した変量効果を含む階層ベイズ・モデリングを構築した。モデル推定にはマルコフ連鎖モンテカルロ (以下、MCMC) アルゴリズムを用いている。ケチャップ購買のスキャナー・パネル・データに適用し、世帯の嗜好やブランド・スイッチング、および自己相関に反映される購買行動の特徴に関して考察している。

次にRossi et al. (1996) によって、回帰係数パラメータに消費者に固有の情報を組み込み、消費者異質性を表現する階層ベイズ・モデリングが提案された。この研究でもMCMCアルゴリズムがパラメータ推定に用いられている。それまでは多項ロジット/プロビット・モデルなどの離散選択モデルと階層モデルとを複合させたモデルはパラメータ推定が困難であったが、MCMCアルゴリズムの利用により解決された。

2節の図3で説明したように、階層ベイズ・モデリングの利点は独立に推定不可能である消費者異質性を、変量効果を用いてパラメータとして表現し推定できることである。従来では消費者異質性のモデリングには潜在クラスが頻繁に用いられ、当時研究者の間ではこれらの2つの方法

のどちらがより消費者の異質性を正しく表現できるかが問題となっていたが、それぞれのアプローチが特定のマーケティング問題の解決に役立ち、殆ど優劣に差は無いことが後に示されている (Andrews, Ainslie & Currim, 2002)。

Chiang, Chib & Narasimhan (1998) は考慮集合およびブランド選択モデルのパラメータで消費者異質性を考慮した階層ベイズ・モデリングを提案している。この研究もパラメータ推定にはMCMCアルゴリズムを用いている。実データ解析から、考慮集合の異質性を無視するとパラメータの異質性がモデル化されてもマーケティング変数の影響が過小評価され、嗜好や過去の購買履歴の影響が過大評価されることを明らかにしている。

実験データに対して階層ベイズ・モデリングを適用した研究も報告されている。Wedel & Pieters (2000) は階層ベイズ・モデリングを用いて広告に対する視線の追跡 (アイ・トラッキング) データを分析している。広告に対する注意がブランドの記憶につながる過程をモデル化しているが、階層ベイズ・モデリングを単なるデータ分析の手法として設定せずに、注意と記憶に関する心理学的な理論を背景に、より実データに当てはまりの良いモデルを導き出している。

Gilbride & Allenby (2004) は多くの消費者行動理論には閾値と不連続性が含まれるとした上で、離散選択モデルを用いて、消費者が製品属性と価格の検討において適用するスクリーニングのルールを調査している。選択肢を外す基準としては補償型、連結型と分離型を考慮している。スクリーニングのルールを満たす選択肢は効用最大化理論に合致する方法で評価している。階層ベイズ・モデリングを応用しデータ拡大法を用いてパラメータ推定を行っている。APSカメラのコンジョイント選択実験から得られたデータ分析により、回答者の92%が一つ以上の属性を元に選択肢を選別していることを明らかにしている。

同じ著者のGilbride & Allenby (2006) では、特

定の属性レベルを用いて選択肢をスクリーニングするモデルである多属性意思決定モデルとコスト・ベネフィット関数に基づくスクリーニング・モデルの2つのモデルを示し、消費者異質性を考慮した階層ベイズ・モデリングを用いて推定している。多属性意思決定モデルは属性レベルごとに選択肢が一つ残るまで順次排除するルールを前提としたモデルであり、コスト・ベネフィット関数に基づくスクリーニング・モデルは特定の属性レベルの選択肢を選び、すべての属性の補償としての機能を考慮した上で残りの選択肢から選ぶモデルである。実証分析では連結型および階層ベイズ多項プロビット・モデルを含めた4モデルの比較を行い、コスト・ベネフィット関数に基づくスクリーニング・モデルが多属性意思決定モデルおよび他の標準的な選択モデルよりも優れた結果を示したと述べている。

ウェブ・アクセス・ログ・データに対し階層ベイズ・モデリングを用いた先行研究も何件か報告されている。Ansari & Mela (2003) では電子メールを用いて製品・サービスの消費者の必要に応じて設定を変更する推薦システム、また時間別にプロモーション活動を実行するためのモデル構築を行い、階層ベイズ・モデリングを活用することで消費者一人ひとりに最適なメールを送る推薦システムを構築している。Montgomery, Li, Srinivasan & Liechty (2004) は階層ベイズ動的多項プロビット・モデルを用いてウェブ閲覧のパス情報を分類およびモデル化している。実データには主要なオンラインのブックストアのデータを用いており、従来の多項プロビット・モデルと1次のマルコフ・モデルとを比較し、提案モデルの優位性を確認した上で、ウェブ閲覧行動の予測に有用であると主張している。

社会的相互作用に関するベイズ・モデルの応用としてはMoe & Trusov (2011) が代表的である。消費者は自身の製品使用経験と他者の製品評価の影響の両方を踏まえた上で、オンライン上で製品評価を行っているが、この研究では製品

評価環境における社会的相互作用が、製品評価を行ったその後の製品評価行動と、製品販売の両方に対して与える影響を測定している。

図4に示した構造モデリングに関してベイズ推定を用いた研究もいくつか報告されている。構造モデリング自体の解説はBronnenberg, Rossi & Vilcassim (2005) や Chintagunta, Erdem, Rossi & Wedel (2006) が詳しく、近年の先行研究は加藤 (2016) に詳細にレビューされているため、ここでは直接効用関数の最大化に関する研究を2件紹介する。Kim, Allenby & Rossi (2002) では加法的な効用構造に基づく需要モデルを提案している。経済学における標準的な需要モデルは内点解のみを示しているが、提案モデルは線形の効用構造を入れ子型にし、一部の製品のみ選択される端点解と内点解が混在する状況を考慮している。また階層ベイズ・モデリングにより世帯ごとで効用パラメータを推定している。実データ解析において継続して一つの製品を購入する消費者が確認できたことを踏まえ、いくつかの小売業では顧客レベルでの品揃えのカスタマイズにより、多様な品揃えが少なくなることによる効用の損失が事実上生じないことを示している。

Lee, Kim & Allenby (2013) では一方の商品が他方の商品により依存し効用を生み出す一方で、両方が消費されることにより実用性が向上する商品（非対称補完財）の購買データ分析にベイズ推定を利用している。消費者一人ひとりの製品の購入数量のみが分かっている場合には非対称効果の測定は一般に困難であるが、この研究では非対称効果を測定するための一連の購買決定を対象とした直接効用モデルを活用することにより、購入数量と家庭内在庫に差異のある複数の製品カテゴリーにわたる購買行動を捕捉している。牛乳と穀物のスキャナー・パネル・データを使用して非対称性を調査し、交差価格弾力性と販売方法に関わる変数のスピルオーバー効果を探るための反事実モデルを仮定し分析を行っている。



4 まとめ

現代のマーケティング環境では、様々なマーケティング活動に関するデータが取得され、それぞれに対して適切なデータ分析方法が異なる。すべてに適切かつ効果的に用いられるデータ分析方法は無く、一方で現代のマーケティング課題は複雑になっている (Wedel & Kannan, 2016)。したがってマーケティング分析者には統計学や

計量経済学、コンピューター・サイエンスなどの分野を背景に、マーケティング反応、マーケティング・ミックスの最適化、およびベイズ・モデリングを含む、消費者への適切な個別対応の予測のためのマーケティング・モデリングに関する幅広いスキルが期待されるだろう。ベイズ推定法に限らず、機械学習手法などの推定手法や、またオペレーションズ・リサーチの最適化手法などの修得も重要となると思われる。

文献

- Allenby, G. M. & P. J. Lenk, 1994, "Modeling household purchase behavior with logistic normal regression", *Journal of the American Statistical Association*, 89(428): 1218-1231.
- Andrews, R. L., A. Ainslie & I. S. Currim, 2002, "An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity", *Journal of Marketing Research*, 39(4): 479-487.
- Ansari, A. & C. F. Mela, 2003, "E-customization", *Journal of marketing research*, 40(2): 131-145.
- Bronnenberg, B. J., P. E. Rossi & N. J. Vilcassim, 2005, "Structural modeling and policy simulation", *Journal of Marketing Research*, 42(1): 22-26.
- Chiang, J., S. Chib & C. Narasimhan, 1998, "Markov chain Monte Carlo and models of consideration set and parameter heterogeneity", *Journal of Econometrics*, 89(1-2): 223-248.
- Chintagunta, P. K., T. Erdem, P. E. Rossi & M. Wedel, 2006, "Structural modeling in marketing Review and assessment", *Marketing Science*, 25(6): 604-616.
- Gilbride, T. J. & G. M. Allenby, 2004, "A choice model with conjunctive, disjunctive, and compensatory screening rules", *Marketing Science*, 23(3): 391-406.
- Gilbride, T. J. & G. M. Allenby, 2006, "Estimating heterogeneous EBA and economic screening rule choice models", *Marketing Science*, 25(5): 494-509.
- 加藤諒, 2016, 「Forward-looking な消費者の習慣形成・飽和・学習を考慮したブランド選択モデルの構造推定」『マーケティング・サイエンス』24(1) : 53-76.
- Kim, J., G. M. Allenby & P. E. Rossi, 2002, "Modeling consumer demand for variety", *Marketing Science*, 21(3): 229-250.
- Lee, S., J. Kim & G. M. Allenby, 2013, "A direct utility model for asymmetric complements", *Marketing Science*, 32(3): 454-470.
- Moe, W. W. & M. Trusov, 2011, "The value of social dynamics in online product ratings forums", *Journal of Marketing Research*, 48(3): 444-456.
- Montgomery, A. L., S. Li, K. Srinivasan & J. C. Liechty, 2004, "Modeling online browsing and path analysis using clickstream data", *Marketing science*, 23(4): 579-595.
- Rossi, P. E. & G. M. Allenby, 2003, "Bayesian statistics and marketing", *Marketing Science*, 22(3): 304-328.
- Rossi, P. E., R. E. McCulloch & G. M. Allenby, 1996, "The value of purchase history data in target marketing", *Marketing Science*, 15(4): 321-340.
- 照井伸彦, 2008, 『ベイズモデリングによるマーケティング分析』東京電機大学出版局。
- Wedel, M. & P. K. Kannan, 2016, "Marketing analytics for data-rich environments", *Journal of Marketing*, 80(6): 97-121.
- Wedel, M. & R. Pieters, 2000, "Eye fixations on advertisements and memory for brands: A model and findings", *Marketing science*, 19(4): 297-312.

5

ベイズ統計学を利用した 欠測データの分析

伴 正隆

筑波大学ビジネスサイエンス系 准教授

1 はじめに

どのようなデータセットにも観測値の欠測は起こりうる。分析者が自身で研究デザインを組み、調査計画を立て、分析に必要なデータを集めるならば、できうる限り欠測が発生しないようなデータ収集を考えることができる。しかし、近年のビッグデータ然り、自身の目的以外の用途で収集されたデータについては、分析者によるそのようなコントロールは不可能であり、欠測値について何らかの処理をして目的とする分析を行う必要がある。しかし迂闊にもデータを削除したり、欠測値に系列の平均値を代入したり、観測値からのランダムサンプリングで補完することは、必ずしも適切な処理ではない。欠測データについて不適切な処理を行ったデータセットによる分析には、当然ながら推定結果にバイアスを含む可能性が高い。たとえば単純に欠測のあるデータを削除した場合、その欠測の発生になんらかのパターンや傾向があるときには、いわゆる選択バイアスの問題が生じる。欠測データを用いた統計解析に関する代表的な書籍は Little and Rubin (1987, 2002) であるように、その歴史は古く、欠測の扱いに関する手法の精緻化がなされている。とくに本稿で紹介する多重代入法は Rubin (1978) によって提唱された手法であり、ベイズ統計学の事後分布と予測分布の

技術を利用して欠測値の補完を行っている。

一方、ベイズ統計学は近年の計算機技術の発達と推定方法の工夫によって、それまでは単純なモデルしか扱えなかったが、より複雑な統計モデルの推定が可能となり、多くの分野で応用がなされている。欠測データ解析における多重代入法は当初よりベイズ流の手法であり、近年のベイズモデルによる現象のモデル化との親和性は高い。

本稿ではこのようなベイズ統計学を利用した欠測データの分析について、ベイジアンモデリングの概念的説明から推定の基礎を紹介し、そのあとで多重代入法の概略を示す。次章では、ベイズ統計学を利用した欠測データの分析に関連するベイズ統計学に関するトピックを紹介する。3章にて欠測データ解析の基本的な枠組みを、4章にて多重補完法について、終章ではNMARの欠測データの補完に関する話題に触れて本稿のまとめとする。

2 ベイズ統計学のおさらい

条件付き確率とベイズの定理

ここでは後の章で必要となる知識を事前に紹介することから始める。まず、ベイズ統計学が推測の基礎とするベイズの定理を導出する際に用いる条件付き確率の定義についてである。いま A, B をそれぞれ、標本空間における互いに独立で



はない事象とする。ベン図で表現するなら図1のとおりであり、同時確率 $\Pr(A \cap B) = \Pr(A, B)$ が $\Pr(A, B) = \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$ と分解できるというのが条件付き確率の定義である。ここで $\Pr(A|B)$ は、 B を条件付きとしたときに A が起こる確率。少し表現を変えると、事象 B について何らかの（数値など）結果を得たときの事象 A が起こる確率である。これは同時分布を分解する際や、欠測データ分析において欠測のパターンを紹介する際に頻出する。

つぎにベイズの定理は前述の条件付き確率の定義をアレンジし、

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} \quad (1)$$

で定義されるものである。ここで B は何らかの結果として得たもの、 A がその原因となりうるような事象という因果関係を考えると、右辺の $\Pr(A)$ は原因となりうる事象 A が起こる確率、 $\Pr(B|A)$ は原因となる事象 A のもとで結果 B が起こる確率となり、左辺 $\Pr(A|B)$ は結果 B が与えられたときに原因が A である確率となる。身近な話として例えば、あるビルを出たら玄関先のアスファルトが少し濡れていたという結果 (B) を得たときに、その原因が雨 (A) である確率を考える。すると、 $\Pr(A)$ はその日の朝に見た天気予報の降水確率、 $\Pr(B|A)$ は「雨が降ったら玄関先がこのくらい濡れるだろう」と玄関先の様子から得る蓋然性となるだろうか。この2つの値から $\Pr(A|B)$ が示す「玄関先が少し濡れているのは、雨が降ったから」の因果関係の確からしさを得ることになる。このようにベイズの

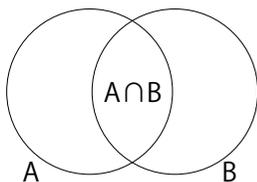


図1 同時確率

定理では、右辺が示す原因から結果が起こる確率によって、左辺が示す何らかの結果に対する原因の妥当性を示す確率を得ることができる。

定理の解釈を一般化し、 A を何らかのパラメータをもつ確率密度関数によって表現される統計モデル、 B を分析者が得たデータとすると、 $\Pr(A)$ は分析者がもつモデルに対する確信の度合いであり「事前確率」と呼ばれる。 $\Pr(B|A)$ はモデルの下で観測されたデータが起こり得る確率、つまり「尤度」である。 $\Pr(A|B)$ はモデルに対する確信の度合いがデータを観測したことでどのように変化したかを示し、「事後確率」と呼ばれる。

さらに新しい記号を導入し、結果 B に対応する観測されたデータを y 、原因 A に対応する統計モデルがもつパラメータを θ とする。 $p(\cdot)$ は連続確率分布の密度関数である。すると $\Pr(A) = \int_{\theta \in A} p(\theta) d\theta$ であり、事後確率は $\Pr(B|A) = \int_{y \in B} p(y|\theta) dy$ となる。 $p(\theta)$ は事前分布であり、パラメータに関する情報である。 $p(y|\theta)$ は観測データから計算されるモデルの尤もらしさ、 $p(\theta|y)$ は事後分布であり、データを得たことで更新されたモデルの確からしさである。 $p(y)$ は確率の公理を満たすための基準化定数である。事象が起こる確率ではなく確率分布においてベイズの定理を再構成すると、

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (2)$$

となる。推測の対象は分析者が考えるモデルであり、これを推定するには基準化定数 $p(y)$ は省略され、(2) 式は $p(\theta|y) \propto p(y|\theta)p(\theta)$ と表現される。

マルコフ連鎖モンテカルロ法

ベイズ統計学では、(2) 式が示すようにあらゆる推測が事後分布によってなされ、パラメータに関するほとんどの統計的推測は、事後分布に関する期待値によって表現できる。いま $g(\theta)$ を

パラメータ θ に関するなんらかの統計量とすると、

$$E[g(\theta)|y]=\int g(\theta)p(\theta|y)d\theta\propto\int g(\theta)p(y|\theta)p(\theta)d\theta \quad (3)$$

である。推定したいものが θ の期待値であるならば $g(\theta)=\theta$ である。つまり、パラメータの推定値(事後分布の期待値)を求めるにも積分計算が必要となり、極々単純なモデルならば解析的に解ける可能性もあるが、多重積分になるとモンテカルロ積分を用いた推定方法の一種であるマルコフ連鎖モンテカルロ(MCMC: Markov chain Monte Carlo)法が利用される。MCMC法について詳説は避けるが、最終的に推定値を得たい事後分布 $p(\theta|y)$ を定常分布にもつようなマルコフ連鎖を構成し、モンテカルロシミュレーションを行う方法である。マルコフ連鎖とは、 t を時間として確率変数の系列 $(\Theta_0, \dots, \Theta_T)$ に対する実現値 $(\theta^{(0)}, \dots, \theta^{(T)})$ が与えられたとき、 $\Pr(\Theta_{T+1}=\theta^{(T+1)}|\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(T)})=\Pr(\Theta_{T+1}=\theta^{(T+1)}|\theta^{(T)})$ が成立する確率過程である。つまり、今回の実現値は前回の値にのみ依存するという構成である。 t を順次進めてゆくと $\Pr(\theta)=\Pr(\theta^{(t)}|\theta^{(t-1)})\Pr(\theta)$ となり、 t に依存しない定常状態となる。 t をモンテカルロシミュレーションの繰り返し回数とし、事後分布についてマルコフ連鎖を構成したシミュレーション、つまり事後分布からの乱数のサンプリングを繰り返して t が増えると、いずれその事後分布の定常分布に収束する。その際の実現値のシミュレーションを

$$\theta^{(t+1)}\sim p(\Theta|\theta^{(t)}), \quad t=1, \dots, T$$

と表現する。(2)式で使用している記号に立ち返ると $\theta^{(t+1)}\sim p(\theta|y, \theta=\theta^{(t)})$ となり、既に得ている θ の実現値 $\theta^{(t)}$ を用いて事後分布から乱数発生によって $\theta^{(t+1)}$ を得るものになる。例えば $p(\theta|y)$ が正規分布で平均 μ と分散 σ が未知、つまり $p(\theta=\{\mu, \sigma\}|y)$ の状況を考える。すると、同時事後分布は

$p(\mu, \sigma|y)=p(y|\mu, \sigma)p(\mu, \sigma)$ と正規分布による尤度とパラメータの同時事前分布に分解し、さらに同時事前分布を条件付き確率の定義から $p(\mu, \sigma)=p(\mu|\sigma)p(\sigma)$ とし、 $p(\mu|\sigma)$ と $p(\sigma)$ それぞれに何らかのハイパーパラメータをもつ確率分布を指定し、 $t=1, \dots, T$ について

$$\begin{aligned} \mu^{(t)} &\sim p(\mu|\sigma^{(t-1)}, y), \\ \sigma^{(t)} &\sim p(\sigma|\mu^{(t)}, y) \end{aligned} \quad (4)$$

のように繰り返しサンプリングを行う。事後分布は尤度と事前分布の積によって表現されるので $p(\mu|\sigma^{(t-1)}, y)\propto p(y|\mu, \sigma^{(t-1)})p(\mu)$ であり、 $p(\sigma|\mu^{(t)}, y)\propto p(y|\sigma, \mu^{(t)})p(\sigma)$ である。十分な大きさの T だけをサンプリングすることで、大数の法則から(3)式に示したパラメータの期待値を得ることができる。定常分布に収束するまでに必要な繰り返し回数に加え、大数の法則が十分に働くだけの十分な T が必要である。

エルゴード性や自然共役事前分布、無情報事前分布などのベイズ推定に関する概念は他誌を参考にしてほしい。

予測分布とデータ拡大

欠測データの補完を考えるにあたり、ベイズ統計学の予測分布を用いたデータ拡大法について紹介する。ベイズ統計学では前述のとおりパラメータについても確率変数として扱い、モデルの推測においてはパラメータの事後分布をMCMC法によりシミュレートしている。そしてこれはパラメータに限らず予測値についても同様である。これまでの y を既に観測されているデータとし、 y^* を新たに得たい予測値とすると、予測分布は

$$\begin{aligned} p(y^*|y) &= \int p(y^*, \theta|y)d\theta \\ &= \int p(y^*|y, \theta)p(\theta|y)d\theta \\ &= \int p(y^*|\theta)p(\theta|y)d\theta \end{aligned} \quad (5)$$

となる。(5)式の3行目へのアレンジには、 $\{y, y^*\}$



が標本間で互いに独立であることから成立する条件付き独立性の性質を利用している。パラメータのみならず、予測値についても値ではなく分布として評価するため、予測の不確実性を含めて値を得ることができる。言い換えると予測精度についても推定することができる。また、積分計算が必要になるが、MCMC法でモデルを推定する際に、予測分布の評価もMCMCの手続きのなかで同時に行うことが可能である。つまり(4)式を例にするのであれば、(4)式にある $\{\mu^{(i)}, \sigma^{(i)}\}$ のサンプリングのあとで、

$$y^{*(i)} \sim p(y^* | \mu^{(i)}, \sigma^{(i)}) \quad (6)$$

によって $y^{*(i)}$ を平均分散がそれぞれ既知の正規分布からサンプリングすればよい。このように予測値を得るメカニズムは、単に予測をする目的のみならず、欠測値の補完、モデル推定の簡略化、さらには分析(学習)用データの増しという目的でも利用されることがあり、そのような枠組みを総称してデータ拡大法(Data Augmentation method)と呼ばれている。

以上、本章では欠測データ解析、とくにベイズ統計学を用いた多重代入法の紹介をするために必要なトピックとして、条件付き確率と条件付き独立性、ベイズの定理の解釈とベイズ統計学における推測の基礎、そしてデータ拡大法の一環である予測分布による予測について紹介した。

3 欠測データの解析

欠測データの解析法について、分析しようとしているデータセットの次元の大きさやデータの尺度といったデータそのものがもつ性質だけでなく、分析の目的の違いによる欠測の捉え方や、そもそも欠測を最小限にするような調査方法に関する検討など、議論の広がりや多岐にわたる。

本稿は紙幅の制限もあるので、 I だけの数の主体について、 J 種類の変数を取得している状況を考える。つまり行列サイズ $I \times J$ のクロスセ

クションデータであり、この行列を Y とし、各変数ベクトルを Y_j と表記する。また、 Y の要素ごとに観測された値か欠測かを示す変数、指示変数 M を導入し、 Y_{ij} が欠測であれば $M_{ij} = 1$ 、そうでなければ $M_{ij} = 0$ と定義する。これら変数は何らかのメカニズムをもって発生していると考え、 Y を生成する統計モデルを $p(Y | \theta)$ 、 M を生成する統計モデルを $p(M | \phi)$ とする。 θ と ϕ は互いに素とし、それぞれデータに対応する次元をもつパラメータベクトルである。さらに、欠測データ解析に関するテキスト等によく見られる記号付けに倣い、 Y の要素のなかで観測された集合を Y_{obs} 、欠測である集合を Y_{mis} と表記する。

欠測のメカニズムと無視可能性

欠測を含むデータについて解析を行う際に、その欠測はどの程度無視できるものなのであるうか。欠測データ解析の重要な概念のひとつに、そのような無視可能性がある。1章で触れたように、欠測を含むデータ解析は欠測の処理の仕方によっては、解析によって得た推定値に大きなバイアスを含むことがある。いま欠測を含むデータについて、欠測が無かった場合の完全データ Y と M の同時分布を考えると

$$p(Y, M | \theta, \phi) = p(Y | \theta) p(M | Y, \phi) \quad (7)$$

となり、関心があるのは完全データでの母数の推定 $p(Y | \theta)$ である。観測されていない Y_{mis} を積分消去すると

$$p(Y_{obs}, M | \theta, \phi) = \int p(Y_{obs}, Y_{mis} | \theta) p(M | Y_{obs}, Y_{mis}, \phi) dY_{mis} \quad (8)$$

となる。いま欠測が完全に無視可能であるならば、完全データの同時分布が観測値だけのそれに等しくなり、これは

$$p(Y_{obs}, M | \theta, \phi) = \int p(Y_{obs}, Y_{mis} | \theta) dY_{mis} = p(Y_{obs} | \theta) \quad (9)$$

で表せる。これは分析の目的が完全データでの θ の推定であれば、観測値だけの確率密度関数

から尤度関数を構成し、最尤法により推定値を得ればよいことを表している。

すると、無視可能性とは $p(M | Y_{obs}, Y_{mis}, \phi)$ の項が示す、欠測の発生メカニズム、あるいは欠測のデータ $Y = \{Y_{obs}, Y_{mis}\}$ との依存関係に依ることがわかる。欠測データ解析の分野では、このメカニズムについて、標準的にはつぎの3つの場合分けを考える。

- 1) $p(M | Y_{obs}, Y_{mis}, \phi) = p(M | \phi)$
...MCAR: Missing Completely at Random
- 2) $p(M | Y_{obs}, Y_{mis}, \phi) = p(M | Y_{obs}, \phi)$
...MAR: Missing at Random
- 3) $p(M | Y_{obs}, Y_{mis}, \phi) = p(M | Y_{obs}, Y_{mis}, \phi)$
...NMAR: Missing Not at Random

ひとつひとつ確認すると、MCARは欠測の発生が Y に依存せず、完全にランダムに発生している。少し表現を変えると、欠測のメカニズムが Y_{obs} でも Y_{mis} でも説明できないともいえる。このような状況であれば (9) 式が成り立ち、観測されたデータだけで解析を行うことができる。

MARの場合、欠測の有無 M は欠測値そのもの Y_{mis} との関係が無い場合、(8) 式は

$$p(Y_{obs}, M | \theta, \phi) = p(Y_{obs} | \theta) p(M | Y_{obs}, \phi) \quad (10)$$

となり、同時分布は観測値の確率分布と欠測の有無に関する確率分布の積に分解できる。従って

$$p(Y_{obs}, M | \theta, \phi) \propto p(Y_{obs} | \theta) \quad (11)$$

から、ベイズ法も含めて尤度に基づく方法で θ を推定するのであれば、欠測値の無視は可能になる。

最後にNMARでは、観測されたデータだけでは欠測の発生を説明できず、そもそも観測されない Y_{mis} との関係をモデル化する必要がある。NMARの場合のモデル化は、完全データと M の同時分布の分解の仕方により2つに大別される。ひとつは「選択モデル」と呼ばれるものであり、欠測のメカニズムであ

る $p(M | Y_{obs}, Y_{mis}, \phi)$ を離散選択モデルによってモデル化するアイデアである。もうひとつは (7) 式を $p(Y, M | \theta, \phi) = p(Y | \theta, M) p(M | \phi)$ と分解する「パターン混合モデル」と呼ばれるものである。概略としては、欠測のなかに何らかの K 種類のパターンを見出して $M = \{M_1, \dots, M_K\}$ とし、パターンを条件付きとしてモデルを評価するものである。

単調／非単調

また、欠測データ解析において、分析方法選択の分岐点となる概念のひとつに欠測パターンの単調／非単調がある。欠測パターンが単調であるとは、 Y に含まれる変数ベクトルを適切に並べ替えると、図2 (a) のように欠測の少ない変数順に並べ替えることができるものをいう。クロスセクションデータではあまり想定されないが、特定の対象について経時観測したデータについては、観測対象の脱落 (dropout) が最たる例である。一方、非単調については、図2 (b) のように欠測の少ない順での並べ替えが上手くできない状態である。経時観測データであれば、ある観測対象についてある時期に欠測があり、その後またある時に値が観測される状況になる。

このような分類を採用する理由は、欠測パターンが単調であれば観測データの同時分布を簡潔に表現することができ、一定の条件 (MCAR) のもとで欠測データの補完が相対的に容易になるためである。変数ベクトル $Y_{obs,j}$ について実現値を y_j とし、欠測の少ない順に $j = 1, 2, 3$ とすると、条件付き確率の定義から

$$p(y_1, y_2, y_3) = p(y_2, y_3 | y_1) p(y_1) = p(y_3 | y_1, y_2) p(y_2 | y_1) p(y_1)$$

$$p(y_1, y_2) = p(y_2 | y_1) p(y_1)$$

$$p(y_1)$$

となる。ここで例えば $p(y_2 | y_1)$ にパラメータ $\gamma_2 = \{\alpha, \beta, \sigma^2\}$ を導入し、 $p(y_2 | y_1, \gamma_2)$ を誤差項を $\varepsilon \sim N(0, \sigma^2)$ とする回帰モデルとして



考えると、 $y_2 = \alpha + \beta y_1 + \varepsilon$ として表現できる。つまり、このような回帰式で $p(y_1)$ から順次パラメータを推定することで、 Y_{obs} の同時分布を得ることができる。

一方で、非単調の場合にはこのようなベクトルごとの同時分布の分解はできず、場合によっては変数ベクトルごとにモデルを構築するなどより細かい処理が必要になる。

4 多重補完法

推定結果に含まれるバイアスを問題にせず、欠測データ解析における欠測の取り扱いを大別すると、1) 欠測しているデータを、本稿で言えば主体ごと削除し解析に使用しない。2) 欠測部分に何らかの予測値を補完し、疑似完全データで解析を行う。3) 不完全データとして解析を行う3種類が挙げられる。さらに2)の予測値の補完については、予測値がもつ不確実性を考慮するか否かの分類があろう。本章で扱う多重補完法(Rubin, 1987)の概略は、名前のおりで、いくつかのパターンの予測値を与え(補完)、それぞれの予測値での解析結果を得て(解析)、最終的にそれを統合するもの(統合)である。ここでははじめに2章で紹介したMCMC法を利用する、フルベイズでの解析方法の紹介から始める。

フルベイズでの欠測データ解析

2章で紹介したように、ベイズ統計学の推測の

基礎はベイズの定理であり、パラメータに限らず欠測値(予測値)は確率分布によって表現し、モデルの推定は事後分布の評価によって行う。多重補完法は、MCMC法によって事後分布から Y に関するモデルのパラメータ推定値を得ると同時に予測分布から Y に含まれる欠測値の予測値を得るものである。

いま(11)式にあるMARの状況を考える。(2)式にあるパラメータ θ の事後分布を考えると、積分消去のテクニックで欠測値 Y_{mis} を導入し、条件付き確率の定義を用いて整理すると、

$$p(\theta | Y_{obs}) = \int p(\theta, Y_{mis} | Y_{obs}) dY_{mis} \quad (12)$$

$$= \int p(\theta | Y_{obs}, Y_{mis}) p(Y_{mis} | Y_{obs}) dY_{mis}$$

になる。この最終項は完全データによるパラメータの事後分布 $p(\theta | Y_{obs}, Y_{mis})$ と、予測分布 $p(Y_{mis} | Y_{obs})$ の積になっている。さらに予測分布について(5)式をもとに整理すると、

$$p(Y_{mis} | Y_{obs}) = \int p(Y_{mis} | \theta) p(\theta | Y_{obs}) d\theta \quad (13)$$

となる。これら(12)(13)式の積分評価をMCMC法によって行うと、(4)式の表現を参考に

$$\theta^{(t)} \sim p(\theta | Y_{obs}, Y_{mis}^{(t-1)}),$$

$$\tilde{Y}_{mis}^{(t)} \sim p(Y_{mis} | \theta^{(t)}) \quad (14)$$

となる。具体的な手順としては、MCMC法の繰り返し乱数生成を始める前に、パラメータの初期値 $\theta^{(0)}$ と欠測値の初期値 $\tilde{Y}_{mis}^{(0)}$ を用意してお

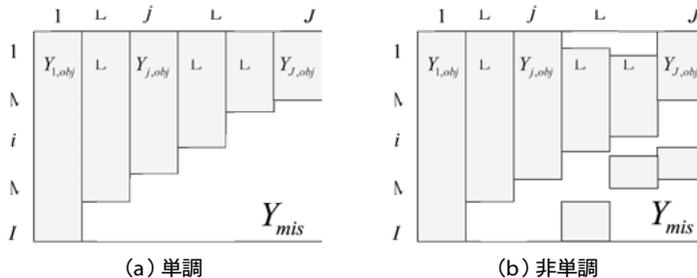


図2 欠測パターンの単調/非単調

く。次いで、(14)式にあるように初期値を用いた疑似完全データ $\tilde{Y} = \{Y_{obj}, \tilde{Y}_{mis}\}$ でのパラメータの事後分布から最初のパラメータ推定値 $\theta^{(1)}$ を乱数生成し、その推定値を利用して欠測値の予測分布から新しい欠測値、つまり予測値を得る。事後分布から得られる乱数の系列が定常過程になって以降の、 t について十分な数の乱数の期待値をとることで、事後分布に関する様々な統計量を得ることができる。

(14)式は欠測値の補完に関するモデルとその推定だが、 Y に含まれる変数を用いた何かしらの解析(主解析)が存在し、それについてもベイジアンモデリングによってモデルが構成されているのであれば、(14)式に主解析のMCMCステップを加えればよい。このように欠測値の補完と主解析とを1つのMCMC戦略において実行することの利点として、それぞれのモデルがもつ不確実性を同時に評価できる点が挙げられる。欠測値の補完に関する分析モデルを R_1 、主解析の統計モデルを R_2 とすると、これは2つのモデルの同時分布 $p(R_1, R_2) = p(R_2 | R_1)p(R_1)$ を評価していることになる。一方で、2つのモデルを別々に推定し、単一値を欠測値に補完して主解析のモデルに利用した場合、主解析については $p(R_2 | E(R_1))$ を評価していることになる。ひいては、 R_2 の推定結果に関する要約統計量には表れない欠測データの補完の仕方に伴う不確実性が含まれることになる。

また、最も単純な問題設定として、完全データ $Y = \{Y_{obs}, Y_{mis}\}$ が各変数の平均ベクトル μ と分散共分散行列 Σ を、パラメータ $\theta = \{\mu, \Sigma\}$ にもつ多変量正規分布に従う確率変数とする。多変量正規分布からの乱数生成が容易なプログラミング言語であれば、欠測のパターンが単調/非単調によらず、どちらも同程度の計算的労力で欠測値の補完を行うことが可能である。多変量回帰モデルのベイジアンモデリングに関する細かい式の展開は、照井(2008)などのテキストが参考になる。

MICE

フルベイズ以外の有力な多重補完法の1つに「MICE (multiple imputation by chained equation)」(van Buuren and Oudshoorn, 2000)がある。前節のフルベイズでは1つのMCMC戦略のなかで2つのモデルを推定することになる。そのため、すでに R_1 で乱数生成しているパラメータ推定値のMCMC系列が定常状態に収束していたとしても、主解析モデル R_2 のMCMC系列が収束するまで、 R_1 の乱数生成を止めることができない。かつてと比べて近年の一般消費者向けPCの計算処理能力は高いのでそれほど問題にならないかもしれないが、 R_2 が多数のパラメータをもつ複雑なモデルである場合、計算的な負担は大きくなる。また、補完モデルについて $Y = \{Y_j\}$ の各変数の同時分布を考えるため、多様な尺度の変数を含む場合は変数間の相関をモデル化するために変数変換など何らかの工夫が必要になる。

一方MICEでは、補完モデル R_1 を変数ごとに構築することで多様な尺度に対応する。具体的には、まず何らかの方法で Y_{mis} に初期値を与えて疑似完全データを構成する。変数 Y_j について自身の観測値 $Y_{j,obs}$ と j 以外の疑似完全データ $\tilde{Y}_{-j} = \{Y_{-j,obj}, \tilde{Y}_{-j,mis}\}$ から変数 j の補完モデルを構築し、事後分布 $p(\theta_j | Y_{j,obs}, \tilde{Y}_{-j}, M)$ とする。さらに予測分布を $p(Y_{j,mis} | Y_{j,obs}, \tilde{Y}_{-j}, M, \theta_j)$ とし、 $j = 1, \dots, J$ についてこれら事後分布と予測分布から乱数生成する。これをMCMCの1サイクルとして、初期値の影響が無くなるまで10~20回程度回し、その後十分な回数 T の乱数生成を行い、 T 組の疑似完全データを得る。変数ごとに観測されている分のデータと、それに対応する疑似完全データについて、変数ごとに個別にモデルを構築している点が特徴であり、明示的に変数 Y の同時分布を評価しない構成になっているが、White et al. (2011) ではフルベイズ法に比して性能に大差はない報告がなされている。

これら欠測値に誤差を含む予測値を補完するパラメトリックな方法以外にも予測平均マッチ



ング法 (Heitjan and Little, 1991) や傾向スコア法 (Lavori et al., 1995) といったノンパラメトリックな補完モデルも提案されている。

5 おわりに

本稿では触れなかったが、NMARの欠測データ解析に関する研究も近年盛んに行われている。NMARは(8)式の $p(M | Y_{obs}, Y_{mis}, \phi)$ が示すように、欠測の発生メカニズムが欠測値 Y_{mis} 自身に依存しているため、そもそも観測できていないものからこのメカニズムを明らかにすることは難しい。そこでデータ融合(data fusion)のテクニックを用い、外部データから欠測値を予測する方法が提案されている。データ融合の基本的なアイデアは、情報源の異なるデータセットについて、それぞれの共通情報をのりしろとして統合するものであり、マルチソースデータのシングルソース化なども表現される。するとここでは、補完したい欠測を含む変数 $Y_j = \{Y_{j,obs}, Y_{j,mis}\}$ について、 Y_j に関連すると考えられ、 $Y_{j,obs}$ と共通する情報(のりしろ)をもつ外部データから $Y_{j,mis}$ を予測することになる。ここでもベイズ統計学、とくにノンパラメトリックベイズと呼ばれるモデルのクラスが有効であると言われている(星野, 2009など)。この代表的なものはディリクレ過程という確率過程を事前分布に設定するモデリングである。ディリクレ過程からはデータの数だけの確率分布を生成することが可能であるため、データに適合した柔軟な確率分布、正確には無限混合分布を生成することができる。(2)式にあるようにベイズモデルではパラメータに事

前分布を設定するが、標準的なベイズ統計学ではこれは単一の確率分布である。ディリクレ過程ではたとえばここに幾つもの正規分布が混合した混合正規分布が設定され、複数の山をもつ柔軟な形状の確率分布からパラメータ推定や欠測値の予測が可能となる。同じクラスのモデルで、他にもガウス過程を利用したNMARへのデータ融合による対応(光廣・星野, 2020)も開発されている。

また、多重補完法による欠測データの補完では、欠測メカニズムをいかにMARに近づけ妥当性の高い補完にするかが重要である。そのためには、MARの定義である(10)式の回帰式 $p(M | Y_{obs}, \phi)$ を考えると、欠測のメカニズムについて観測値 Y_{obs} による説明力の高さが必要であり、その方法のひとつに補助変数の導入がある。補助変数とは研究デザインで使用する変数とは無関係で、対象に関する特性を表す、欠測値の予測に有益な変数である。ここで仮に研究デザインに関わる変数を補助変数とすると、研究デザイン由来でNMARになってしまっている場合に、その補助変数ではランダム性を高めることにはならない。

以上、ベイズ統計学と欠測データ解析の概説から始まり、MCMC法による多重補完法について紹介したが、そもそも観測されないデータの妥当性の検証は困難であり、ひとつの補完方法によって補完を完結するのではなく、種類の異なる複数の方法で補完し結果を確認することや、データのMARからの乖離の影響を調べるなどをする必要がある。

文献

- 阿部貴行, 2016, 『欠測データの統計解析 (統計解析スタンダード)』朝倉書店。
- Heitjan, F. and R. J. A. Little, 1991, "Multiple imputation for the fatal accident reporting system," *Applied Statistics*, 40: 13-29.
- 星野崇, 2009, 『調査観察データの統計科学—因果推論・選択バイアス・データ融合 (シリーズ確率と情報の科学)』岩波書店。
- Lavori, P. W., R. Dawson and D. Shera, 1995, "A multiple imputation strategy for clinical trials with truncation of patient data," *Statistics in Medicine*, 14: 1913-1925.
- Little, R. J. A. and D. B. Rubin, 1987, *Statistical Analysis with Missing Data*, New York: Wiley.
- , 2002, *Statistical Analysis with Missing Data*, 2nd ed., New York: Wiley.
- 光廣正基・星野崇宏, 2020, 「NMARな欠測でのガウス過程によるデータ融合法」人工知能学会全国大会論文集(第34回全国大会)。
- Rubin, D. B., 1978, "Multiple Imputations in Sample Surveys: A Phenomenological Bayesian Approach to Nonresponse," *Proceedings of the Survey Research Methods Section*: 20-34.
- , 1987, *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley and Sons.
- 照井伸彦, 2008, 『ベイズモデリングによるマーケティング分析』東京電機大学出版局。
- van Buuren, S. and C. G. M. Oudshoorn, 2000, *Multivariate imputation by chained equations: MICE V1.0 user's manual*, TNO Prevention and Health, Leiden.
- White, I. R., P. Royston and A. M. Wood, 2011, "Multiple imputation using chained equations: Issues and guidance for practice," *Statistics in Medicine*, 30: 377-399.



特集論文

6

ベイズ統計学のための ソフトウェア

JASPとbrmsによる分析

清水裕士

関西学院大学社会学部 教授

近年、ベイズ統計学を用いた分析が注目を集めている。心理学の分野では2010年代に、Lee & Wagenmakers (2013) やKruschke (2015) といったベイズ統計学についての書籍が立て続けに出版され、また日本においてもそれらの訳本、豊田 (2015, 2016, 2017) のベイズ統計学についてシリーズ本、生態学者の久保 (2012) による統計モデリングの入門書、そしてデータサイエンスの分野においては松浦 (2016) が出版している。書店の統計学の棚には、ベイズ統計学に関する本がたくさん平積みされていることに気づかれた読者も多いだろう。

このようにベイズ統計学が注目されているなか、これを研究に実際に取り入れた論文も増えてきている。だが一方で、それらの論文は、統計学に明るい研究者によるものが多く、まだ一般の社会科学者に広がっているとは言い難い。その背景として、ベイズ統計分析を実行するためのソフトウェアが十分に整備されてこなかったこと、そしてソフトウェアが充実したあと、その存在や使い方について周知されてこなかったことも、普及を遅らせた要因として挙げられるだろう。

そこで本論文は、ベイズ統計分析を実行するための代表的なソフトウェアやパッケージを紹介することを目的とする。実はすでに、ベイズ統計分析はかなり簡単に利用できるようになっていることが、本論文を読むことで理解できるだろう。もしベイズ統計分析を自身の研究に取り入

れたいと考えている読者がいれば、本論文がその導入のきっかけになることを期待する。

1 二種類のベイズ統計分析

ベイズ統計分析といっても、さまざまな方法がある。これを大雑把に二種類に分けると、1つは仮説検定の方法をベイズ統計的に行う分析であり、もう1つはモデルの推定にベイズ統計学を利用する分析である (清水, 2018)。

まず、前者の仮説検定にベイズ統計分析を用いる方法について簡単に解説する。これは、頻度主義統計学の帰無仮説検定の代わりにベイズ統計学による仮説検定を使うものである。頻度主義統計学では、仮説検定の指標として p 値が用いられる。しかし、帰無仮説検定における p 値が研究者であっても正確に理解されていないこと、また p 値を不当に小さく見せる研究実践がこれまで多く用いられてきたことから、批判も多くなってきた (大久保, 2016)。それらの背景から、帰無仮説検定ではなく、ベイズ統計学を用いた仮説検定の有用性が主張されるようになってきた (e.g., Kruschke, 2015)。ベイズ統計学を用いた仮説検定は、 p 値ではなくベイズファクターという指標を用いる。ベイズファクターを用いた仮説検定では、帰無仮説と対立仮説とを対等に扱うため、帰無仮説を積極的に採択することができるようになるとされている (岡田, 2018)。

ベイズファクターの解釈や計算方法については岡田 (2018) に詳しい。

続いて、後者のベイズ推定を用いた分析について解説する。t 検定や回帰分析といったシンプルな統計分析であればパラメータ推定は簡単に行える。しかし、階層線形モデル、因子分析、項目反応理論、一般化線形混合モデルといった複雑な統計分析となると、最尤法でも推定が難しい、あるいはできても不安定になることがある (久保, 2012)。そこで複雑なモデルでも推定が安定して行えるマルコフ連鎖モンテカルロ法 (Markov Chain Monte Carlo method: 以下, MCMC法) によるベイズ推定が注目された。もともとベイズ推定そのものは最尤法では解きにくい階層モデル (階層線形モデルや因子分析のように、パラメータが更に確率分布に従うようなモデル) を比較的安定して解くことができる方法であるが、事後分布の計算そのものに限界点があった。しかしMCMC法は、ベイズ推定における事後分布を解析的に計算できない問題と、パラメータごとの周辺事後分布の計算が難しい問題を同時に解決することができる方法である。MCMC法の登場によって、複雑なモデルを安定して解くことができるようになり、自然科学、社会科学、データサイエンスなど、さまざまな領域で活用されている。

2つのベイズ統計学の活用方法について概説したが、どちらのベイズ統計分析を使いたいかによって、より便利なソフトウェアが変わってくる。ベイズ統計学による仮説検定を行いたいのであれば、Wagenmakersのグループが作っているJASPがとても簡単で便利である。一方で、複雑なモデルをベイズ統計学によってパラメータ推定したい場合は、StanやJAGSといったMCMC法を用いた汎用的な確率的プログラミング言語を用いる研究が多い。しかし、これらの確率的プログラミング言語は習得のためのハードルがやや高い。そこで、Stanを統計ソフトR上で簡単なコードを書くだけで動かすことができ

るbrmsパッケージ¹⁾を紹介し、ベイズ統計モデリングの敷居が大きく下がっていることを示す。

2 JASPを用いたベイズ統計分析

それではまず、JASPを用いたベイズ統計分析の方法を解説する。

JASPはアムステルダム大学のWagenmakersを筆頭としたJASPチームが開発している、無償で提供されている統計ソフトウェアである。JASPは帰無仮説検定に基づく分析と、ベイズ統計学による仮説検定の両方を行えるのが特徴である。JASPはユーザーインターフェイスにグラフィカルなものを採用しており、マウスのクリックによる操作で分析ができる。一部の分析では内部で統計ソフトRが動いており、解析がどういいうプログラムで動いているのかを確認することができる。

JASPは、Webサイト²⁾にアクセスすれば、ダウンロードすることができる。なお、本稿で紹介しているJASPはver0.13を用いている。また、JASPについての詳しい使い方については、JASPのWebサイトに詳しく書かれている。ただ、すべて英語で書かれているので、学生向けには清水・山本 (2020) がわかりやすいので、こちらも併せて参照するとよいだろう。これ以外にも、難波・由比・神原・宮谷・中尾 (2016) の解説もある。

3 データの読み込み

JASPで分析するには、まずデータを読み込む必要がある。データは左上の青い三本線があるボタン (図1の丸印で囲まれたところ) を押し、「Open」を選択するとデータの読み込み方法について選択肢が4つ表示される。「Recent Files」はこれまで読み込んだことがあるファイルを選択でき、「Computer」はPCのフォルダから選択でき、「OSF」はOpen Science Frameworkにアッ



図1 JASPの画面

プされているデータセットを読み込むことができる。また「Data Library」からはJASPに最初から含まれているデータセットを活用できる。最初は「Computer」を選び、「Browse」を選択すればよいだろう。データファイル形式はcsvファイルで保存するのが一番わかりやすい。

またJASPの特徴として、もとのデータファイルを更新すると、すぐにJASP内のデータも更新されるという点がある。データファイルを変更した場合にもう一度読み込み直す必要はない。

データを読み込むと、図2のようにデータセットの中身が表示される。また、上部に選択可能なデータ分析のボタンが表示される。最初に確認すべきなのは、それぞれの変数の型である。図2の変数aやbは丸が3つ重なったようなマークがついているのに対し、time1やtime2はモノサシのようなマークがついている。またageは棒が3つのマークが付いている。丸が3つのマークは名義尺度であることを意味し、棒が3つのマークは順序尺度、モノサシのマークは間隔尺度であることを意味している。このマークは利用者が自由に切り替えることができるが、変数がどの尺度であるかによって使える分析が変わってくるので注意が必要である。

JASPでは他の統計ソフトと同様、データハンドリングや要約統計量の計算など、基礎的な分析ももちろん実行可能である。しかし、本原稿はベイズ統計分析に関する部分のみの紹介に留める。

4 ベイジアン t 検定

JASPの特徴はすでに述べたように、頻度主義に基づく帰無仮説検定だけでなく、ベイズ統計学に基づく仮説検定を行える点にある。ベイズ統計学による平均値の差についての仮説検定は、ベイジアン t 検定などと呼ばれるが、JASPでもそれが可能である。

ベイジアン t 検定は、 p 値の代わりにベイズファクターを計算して帰無仮説と対立仮説とを比較する点が頻度主義統計学の帰無仮説検定と異なる。ベイズファクターは2つの仮説の周辺尤度の比で計算される。周辺尤度とは、事前分布を含めたモデル（ベイジアン t 検定の場合は帰無仮説か対立仮説）が正しいときにデータが得られる確率を意味している。その比は、相対的にどちらのモデルがデータから見て確からしいかを表していると言える。

ベイジアン t 検定におけるベイズファクターは、帰無仮説を基準としたときの対立仮説の確からしさ、つまり対立仮説の周辺尤度を帰無仮説の周辺尤度で割ったもので計算される。ベイズファクターは一般に、おおよそ3を超える（すなわち3倍対立仮説が確からしい）と、差があると判断することが多い。ただし、それも明確な基準ではなく、あくまで目安であることに留意しておこう。

なお、JASPのベイジアン t 検定はサヴェッジ=ディッキー法と呼ばれる方法で直接ベイズファクターを計算しており、それぞれの仮説の周辺尤度そのものは計算されない。

ベイジアン t 検定を実行するには、図2のT-Testsというボタンを押すとメニューが表示され、その下側にある「Baysian」の中の、「Bayesian Independent Sample T-Test」を選択する。すると、図3のような画面が表示される（ただし、図3は分析実行後の図）。

ここでは、time4の平均値がaという2群を識

別する変数によって差が見られるかどうかを検定してみよう。Dependent Variablesにtime4を指定し、Grouping Variableにaを指定する。このとき、Dependent Variablesに入れることができるのは間隔尺度だけ、Grouping Variableには名義尺度と順序尺度だけであることに注意する(それぞれのボックスには入力可能な変数の型がマークで表示されている)。変数を入力すると、即座に結果が表示される。しかし、最初に表示されるのはベイズファクターだけなので、PlotsにあるPrior and posteriorにチェックを入れるとよい。すると、図3のように事前分布と事後分布が表示され、ベイズファクターも同時に表示されるのでわかりやすい。今回の分析では、ベイズファクターが6を超えていたので、差があると言えるだろう。また、差の95%信用区間も表示されており、-2.94から-0.12程度であることがわかる。

ベイズian t検定で注意が必要なのは、事前分

布によって大きく結果が変わることがある点である。サヴェッジ=ディッキー法によるベイズian t検定では、効果量 δ の事前分布として尺度パラメータが0.707のコーシー分布を用いることが多い。ベイズ統計学に明るくないユーザーはその設定を積極的に変更する必要はないが、事前分布を変えることでどのように結果が変わるのかは確認しておいたほうがよいだろう。JASPには事前分布の設定を変えるとどのようにベイズファクターが変化するかを图示する機能がある。PlotsのBayes factor robustness checkをチェックすると、図4のように事前分布の尺度パラメータを変えたときにベイズファクターの変化を图示してくれる。今回の結果では、事前分布の設定では大きくベイズファクターが変わらないことを確認できた。

5 ベイズian相関分析

相関係数もベイズ統計学の枠組みで推定し、また仮説検定をすることができる。相関係数をベイズ推定することのメリットは、1つは無相関のモデルと比較した場合のベイズファクターが計算できること、もう1つは相関係数の事後分布を計算することで、信頼区間が視覚的にわかりやすいことである。本特集の岡田論文で説明されているように、ベイズ統計学では信頼区間で

ID	sex	time1	time2	time3	time4	age		
1	OB581	1	A	5.82	7.81	8.81	9.38	21
2	OB582	8	A	6.87	4.84	5.86	6.31	22
3	OB583	1	A	2	4.87	7.95	9.31	18
4	OB584	8	A	7.85	5.88	4.87	7.38	21
5	OB585	1	B	2.81	4.1	6.88	8.35	18
6	OB586	8	B	2.88	3.87	3.83	3.38	18
7	OB587	1	B	3.83	5.83	4.85	7.38	18
8	OB588	8	B	4.85	5.88	6.82	4.38	28
9	OB589	1	C	3.86	4.87	6.88	5.14	18
10	OB510	8	C	3.88	2.88	9.1	2.32	21
11	OB511	1	C	3.82	3.85	6.1	7.31	28
12	OB512	8	C	5.87	6.87	2.85	9.35	28

図2 データを読み込んだJASP

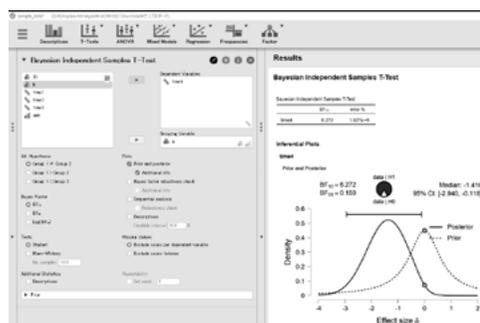


図3 JASPによるベイズian t検定

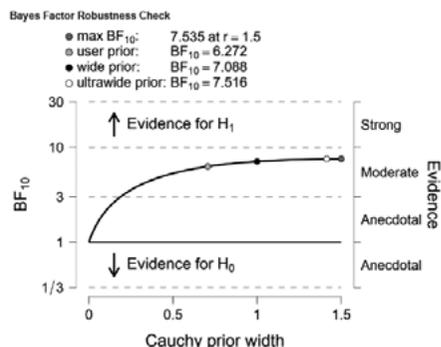


図4 ベイズファクターの頑健性チェック



はなく信用区間が出力される。ベイズ統計分析における相関係数は、2つの観測変数に二変量正規分布を仮定したときの相関係数パラメータの事後分布を計算することで得られる。そのとき、ベイズファクターは相関係数パラメータが0のときのモデルと、自由推定したときのモデルの周辺尤度とを比較することで計算される。

JASPでは、平均値の差の比較と同様、相関係数のベイズ統計分析も容易に実行できる。Regressionを選び、Bayesian Correlationを選択すると、ベイズ統計分析による相関分析が可能である。

JASPのベイジアン相関分析は、事前分布として-1~1の範囲ですべて確率密度が等しい一様分布を用いて解析的に最大事後確率推定値 (MAP) を計算している。よって、相関係数の値自体は通常のピアソンの積率相関係数と一致している。異なる点は、*p* 値の代わりにベイズファクターが

出力されている点や、事後分布についてのプロットを得ることができる点である。(図5、図6)。

6 JASPによるベイズ統計分析まとめ

これまで述べてきたように、JASPはこれまで頻度主義統計学で分析されてきた手法を、そのままベイズ統計分析に適用できる。また、それらの操作はとても直感的で、出力も見やすい。その特徴から、これからベイズ統計分析を行ってみたい、というユーザーにとってはとてもありがたいソフトウェアであると言える。

JASPは本稿でとりあげた平均値の差の比較や相関係数以外にも、分散分析、混合モデル、回帰分析、クロス表分析についてベイズ統計分析が可能である。JASPでのベイズ統計分析は、ほとんどの結果にベイズファクターが自動的に出力されるので、頻度主義以外の仮説検定を行いたい場合には非常に便利である。

ただ、いくつかの複雑な統計分析、たとえば因子分析や構造方程式モデルなどについてはまだベイズ推定は実行できない。また、一般化線形混合モデルについては、いくつかの制約があり、また解析的ではなくMCMC法を用いて推定していることから、次に挙げるbrmsパッケージによる分析のほうが、融通が効いて、使い勝手がよい。

Bayesian Correlation

Bayesian Pearson Correlations					
Variable		time1	time2	time3	time4
1. time1	Pearson's r	—			
	BF ₁₀	—			
2. time2	Pearson's r	0.496	—		
	BF ₁₀	1.194	—		
3. time3	Pearson's r	-0.136	0.264	—	
	BF ₁₀	0.384	0.484	—	
4. time4	Pearson's r	0.020	0.312	0.754	—
	BF ₁₀	0.355	0.552	12.594	—

図5 JASPによるベイジアン相関分析

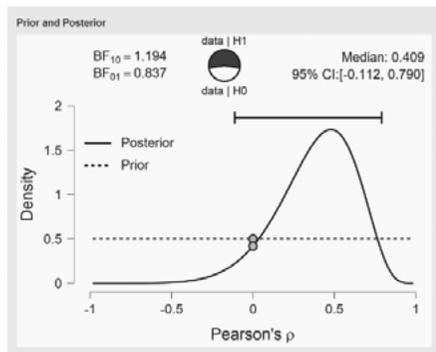


図6 相関係数の事後分布

7 Stanを用いたベイズ統計分析

JASPには簡単な統計分析については、ベイズ統計分析を解析的に実行する機能がなかった。しかし、複雑なモデルになると解析的には解けないので、MCMC法による推定が必要である。そのとき、決まった分析を選択するタイプのソフトウェアよりは、モデルを自由に設定できる確率的プログラミング言語を用いたほうが使い勝手がよくなる。

確率的プログラミング言語とは、確率モデルを特定のルールに基づいて記述するだけで、あ

とはMCMC法などの汎用的な求解法によって自動的にパラメータを推定してくれるプログラミング言語を指す。ベイズ統計分析でよく使われる確率的プログラミング言語には、古くから使われていたがいまは開発が止まっているWinBUGS, WinBUGSの使い勝手を引き継ぎつつアルゴリズムを改良したJAGS, いまでも開発が進んでユーザー数を増やしているStan, データサイエンスの分野でよく使われているTensorFlow Probabilityなどがある。ここで紹介した確率的プログラミング言語は、主としてMCMC法によってベイズ推定を行う。よって、パラメータ推定は最尤法のように点推定値を数値計算によって求めるのではなく、乱数によるサンプリングを繰り返して求める。MCMC法についてはKruschke (2015,2017) に詳しい。

8 Stanによる確率的プログラミング

次に、心理学や社会学の分野で比較的良好に使われている、Stanを紹介する。Stanは統計学者Gelmanが筆頭となって開発している確率的プログラミング言語であり、またそれを搭載したソフトウェアである。StanではGelman自身が開発したNUTS (No-U-Turn Sampler) というMCMC法のアルゴリズムを採用している。NUTSはBUGSやJAGSが採用しているギブスサンプラーに比べて、複雑なモデルにおいてより効率的に事後分布を推定できるため、近年はStanの利用者が増えている。

Stanは、Stan独自の文法に基づいてプログラミングを行い、確率モデルを記述する。しかし、基本的には統計ソフトRに似た文法であるのでそれほど複雑なわけではない。Stanの詳しい使い方について本稿で説明する余裕はないが、例として回帰分析をStanで実行したい場合にどのようなコードを書くのかを示しておく(図7)。

まず回帰分析を確率モデルで表記すると、

$$Y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta X_i$$

となる。このモデルでは、目的変数Yが残差標準偏差 σ をもつ正規分布に従い、その平均パラメータが切片 α と、回帰係数 β による線形モデルで表現されることが示されている。

この確率モデルをStanで記述すると以下のようになる。dataブロックにはStan外部から読み込む変数を宣言し、parametersブロックでは推定するパラメータを宣言する。正の値をとるパラメータにはlower=0という制約をつけている。それらで宣言した変数を使って確率モデルをmodelブロックの中で記述する。

Stanは、多くの統計ソフトによって実行できる。社会科学では統計ソフトRから操作するrstanパッケージが使われることが多いが、PythonやJuliaといったさまざまなプログラミング言語によっても実行できる。

Stanをはじめとする確率的プログラミング言語は、多様な確率モデルに対応することができるという汎用性の高さが特徴である。しかしその分、プログラミング言語の習得には時間がゆかり、分析初心者にはやや敷居が高くなってしまいう問題点もある。本稿は、ベイズ統計分析をこれから

```

1 data{
2   int N;
3   real Y[N];
4   real X[N];
5 }
6
7 parameters{
8   real alpha;
9   real beta;
10  real<lower=0> sigma;
11 }
12
13 model{
14   real mu[N];
15   for(i in 1:N){
16     mu[i] = alpha + beta*x[i];
17     Y[i] ~ normal(mu[i], sigma);
18   }
19 }
20

```

図7 Stanによる回帰分析のコード



はじめたい読者に向けているので、Stanそのものの利用ではなく、Stanコードを指定した確率モデルに応じて自動的に生成してくれるパッケージを次節で紹介したい。

9 brmsパッケージによるベイズ統計分析

brmsパッケージは、Rの一般化線形混合モデルを最尤法で解くためのlme4パッケージと同じ文法でRコードを書くだけで、それに対応したStanコードを内部で自動的に生成し、MCMC法による推定を行ってくれる、いわゆるラッパーパッケージである。当初のlme4は実行可能な確率モデルをStanに翻訳するためのパッケージであったが、その後かなりの機能が拡張され、一般化線形混合モデルだけではなく、項目反応理論、構造方程式モデル、Drift Diffusionモデルなど、心理学や社会学で用いられる多くの統計モデルに対応するようになった。

brmsパッケージの詳しい使い方については、パッケージのHELPドキュメントを見るか、いくつかのWebの記事)で紹介されているので、そちらを参照してほしい。本稿では、簡単な紹介にとどめる。

10 brmsパッケージによる回帰分析

それでは、brmsパッケージを使って上記と同様、回帰分析をベイズ推定してみよう。なお、Rそのものの操作方法についてはここでは解説しない。必要に応じてRに関するテキストを参照してほしい。

まず、Rにデータを読み込み、その後に、インストールしておいたbrmsパッケージを読み込む。そして下記のようにbrm()関数に、モデル式とデータオブジェクトを指定するだけでよい。あとは以下のコードを実行すると、自動的にStanコードが生成され、MCMCがはじまる。

```
dat<-read.csv("data.csv")
library(brms)
fit.reg<-brm(sat~talk1+talk2,data=dat)
summary(fit.reg)
```

summary関数を使うと、推定した結果の要約を見ることができる(図8)。要約では、期待事後確率推定値(EAP)と事後分布の標準偏差(頻度主義における標準誤差に対応する)、95%信用区間が表記される。また、右側にあるBult_ESSとTail_ESSは、MCMCサンプルの自己相関の低さを表しており、高いほど事後分布の結果が信頼できることを意味している。ESSはMCMCサンプルの大きさに応じて大きくなるので、もしESSが小さいときは、MCMCサンプルを大きくするとよい。もし信用区間によって何かを推論しようとするときは、特にTail_ESSが低い場合には注意が必要である。2000以上あれば十分である。

MCMC法を使う場合は、事後分布を要約指標だけで解釈するのではなく、事後分布のカーネル密度推定や、トレースプロットなどを見て、その形状や収束の程度を視覚的にも判断するほうがよい。そのようなときは以下の、

```
plot(fit.reg)
```

というコードによって、事後分布について視覚的な情報が得られる。図9の左側が事後分布のカーネル密度推定、右側がトレースプロットである。

```
> summary(fit.reg)
Family: gaussian
Links: mu = identity; sigma = identity
Formula: sat ~ talk1 + talk2
Data: dat (Number of observations: 300)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept    2.58      0.18    2.23    2.93 1.00    4666    3491
talk1       -0.19      0.06   -0.20    0.01 1.00    3748    2770
talk2        0.36      0.06    0.24    0.48 1.00    3538    2765

Family Specific Parameters:
      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    0.95      0.04    0.88    1.03 1.00    3994    3060
```

図8 brmsによる回帰分析の結果

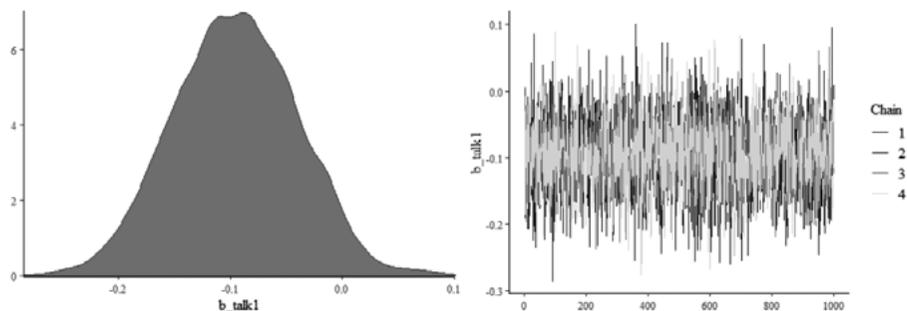


図9 brmsによる事後分布のプロット

11 brmsによる一般化線形混合モデル

brmsは、階層線形モデルや一般化線形混合モデルを簡単に推定できる。以下のようなコードを実行することで、目的変数に正規分布を仮定しつつ、グループで切片の変量効果が正規分布に従うという階層線形モデルを構築できる。

```
fit.hlm <- brm (sat~talk1+talk2+(1|group), data=dat)
summary (fit.hlm)
```

結果を要約すると、図10の出力を得る。回帰係数に加えて、切片の変量効果の標準偏差の推定結果が出力されていることがわかる。変量効果の標準偏差の95%信用区間は、0.40~0.68となっており、切片の平均が2.65であることを考えると、グループ間変動はかなり大きいことがわかる。

```
> summary(fit.hlm)
Family: gaussian
Links: mu = identity; sigma = identity
Formula: sat ~ talk1 + talk2 + (1 | group)
Data: dat (Number of observations: 300)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Group-Level Effects:
~group (Number of levels: 100)
              Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)  0.54      0.07    0.40    0.68 1.00    1497    2277

Population-Level Effects:
              Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept    2.65      0.19    2.28    3.01 1.00    4303    3141
talk1        -0.07      0.05   -0.18    0.04 1.00    5396    3270
talk2         0.31      0.07    0.18    0.45 1.00    4164    2974

Family Specific Parameters:
              Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma         0.79      0.04    0.71    0.87 1.00    3117    2843
```

図10 brmsによる階層線形モデルの結果

また、一般化線形混合モデルについては、familyを指定することで実行可能である。たとえばポアソン分布を仮定したい場合は、

```
fit.glm <- brm (sat~talk1+talk2+(1|group), data=dat, family=poisson)
```

とすればよい。

brmsパッケージが扱える確率分布は多岐にわたり、正規分布、二項分布、ポアソン分布、負の二項分布、幾何分布、対数正規分布、移動対数正規分布、指数分布、ワイブル分布、ベータ分布、ガンマ分布、ディリクレ分布、フォンミーゼス分布、ラプラス分布などがある。他にも、ゼロ過剰ポアソン分布といった混合分布や、順序カテゴリカル分布などの順序尺度を扱うための分布なども用意されている。リンク関数も標準的なものはすべて対応しているので、線形モデルの範囲ではかなり柔軟なモデリングが可能である。

12 おわりに

これまで、JASPとRのbrmsパッケージを使ったベイズ統計分析について紹介した。本稿で述べてきたように、ベイズ統計分析はかなり敷居が低くなっていることがわかるだろう。これまで心理的なハードルの高さを感じてきた読者には、ぜひ一度これらのソフトウェアを使ってベイズ統計分析を試してみしてほしい。



注

- 1) <https://github.com/paul-buerkner/brms>
 2) <https://jasp-stats.org/>
 3) 静岡理工科大学の紀ノ定保礼氏の記事がわかりやすい。 <https://das-kino.hatenablog.com/entry/2018/12/15/230938>

文献

- Kruschke, J. K., 2015, *Doing Bayesian Data Analysis: A tutorial with R, JAGS, and Stan.* (2 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。
- 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。)
- 松浦健太郎, 2016, 『StanとRでベイズ統計モデリング』 共立出版。
- 難波修史・由比匠・神原広平・宮谷真人・中尾敬, 2016, 「JASPによる心理学者のためのベイズ統計」『広島大学心理学研究』 16 : 97-108.
- 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.
- 清水裕士, 2018, 「心理学におけるベイズ統計モデリング」『心理学評論』 61 : 22-41.
- 清水優菜・山本光, 2020, 『研究に役立つJASPによるデータ分析—頻度論的統計とベイズ統計を用いて』 コロナ社。
- 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。
- , 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。
- , 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。