

実験心理学者のための階層ベイズモデリング入門 —RとStanによるチュートリアル—

武 藤 拓 之

京都大学こころの未来研究センター

Introduction to hierarchical Bayesian modeling for experimental psychologists: A tutorial using R and Stan

Hiroyuki MUTO

Kokoro Research Center, Kyoto University

Hierarchical Bayesian modeling is a powerful and promising tool that aids experimental psychologists to flexibly build and evaluate interpretable statistical models that consider inter-individual and inter-trial variability. This article offers several examples of hierarchical Bayesian modeling to introduce the idea and to show its implementation with R and Stan. As a tutorial, it uses data from well-known experimental paradigms in perceptual and cognitive psychology. Specifically, I present linear models for correct response time data from a mental rotation task, probit models for binary choice data from two psychophysical tasks, and drift diffusion models for both response time and binary choice data from an Eriksen flanker task. The R and Stan scripts and data are available on the Open Science Framework repository at <https://doi.org/10.17605/osf.io/2zxs6>. The importance of model selection and the potential functions of open data practices in statistical modeling are also briefly discussed.

Key words: Bayesian statistical modeling, ex-Gaussian distribution, psychophysical measurement, drift diffusion model, open data

データの生成過程を確率モデルで表現し、そのモデルをデータに当てはめることで現象の理解と予測を目指す手法は統計モデリングと呼ばれる。広義の統計モデリングには回帰分析や t 検定、分散分析といったお馴染みの分析も含まれるが、近年、マルコフ連鎖モンテカルロ(MCMC)法と呼ばれる汎用的なアルゴリズムを利用してベイズ推定を実行できるソフトウェアが普及し、複雑な統計モデリングを柔軟かつ容易に実行できる環境が整ってきた。このようなベイズ統計モデリングを実験心理学に適用することで、厳密に統制された刺激から観察可能な反応が生成されるときの情報処理モデルを実証的かつ数理的に検証することが可能となる。この状況に鑑み、本稿では知覚・認知心理学における有名な3つの実験パラダイム（心的回転課題・心理物理課題・フラン

カー課題）を例に、ベイズ統計モデリングの実践法に関する解説とチュートリアルを提供する。特に本稿では、拡張性が高く幅広い現象に適用可能な階層ベイズモデリングに焦点を当てる。

階層ベイズモデリングとは文字通り、データが持つ階層性を考慮したモデルを使ってパラメータをベイズ推定する手法のことである。階層性を仮定した分析法はマルチレベル分析と呼ばれることもある。例えば50試行×2条件から成る反応時間計測実験を30人の参加者に課して計3,000試行分のデータを得たとき、このデータは参加者と試行という2つのレベルから成る階層構造を有している。通常の分析では、外れ値や誤答試行などを除外してから参加者と条件の組み合わせごとに平均反応時間を計算し、得られた集計データを対応のある t 検定などで分析することが多い。このような分析では、試行レベルにおけるデータのばらつきの情報は捨てられることになる。これに対し、階層モデリングではデータが持つ階層構造をモデルに反映させることにより、個人差と試行

Corresponding address: Kokoro Research Center, Kyoto University, 46 Shimoadachi-cho, Yoshida, Sakyo-ku, Kyoto 606-8501 Japan. E-mail: h.muto@zm.commufa.jp
J-STAGE First published online: April 1, 2021

間変動の両方を考慮した分析を行うことができる¹。外れ値や誤答試行の除外などにより参加者ごとの有効データ数が異なっているとしてもそれを考慮できるのが階層モデルの利点の1つである。

簡単な階層モデルであれば線形混合モデル (linear mixed model; LMM) や一般化線形混合モデル (generalized linear mixed model; GLMM) を利用して最尤法で解くこともできるが、こういった既成のモデルでは型にはまった分析しかできないため柔軟性に欠ける。また、複雑なモデルは最尤法ではうまく解けないことも多い。これに対し、MCMC法を利用したベイズ推定であれば、分析者の目的に合わせて確率モデルを柔軟に構築することができるし、確率モデルさえきちんと記述できれば比較的安定してパラメータを推定できる。推定量の漸近正規性を利用する最尤推定と比較して、MCMC法によるベイズ推定ではサンプルサイズが小さいときでも推定のバイアスが小さくなりやすいという利点もある (総説として McNeish & Stapleton, 2016)。また、ベイズ推定ではパラメータの不確かさを事後分布という確率分布で表すため区間推定が行いやすく、さらに事前知識をパラメータの事前分布として反映させることもできる。加えて、2つのパラメータの差や比などといった任意の生成量の事後分布を得たり、将来のデータが従うと期待される事後予測分布を推定して予測に役立てたりすることもできる。こういったベイズ統計モデリングの利点に関しては清水 (2018) に詳しくまとめられている。

本稿では、ベイズ統計モデリングの世界で現在広く利用されている Stan (Stan Development Team, 2021) とその R 用のパッケージである rstan を使ったチュートリアルを提供する。Stan のインストール方法や基本的な使い方などは馬場 (2019) や松浦 (2016), Stan Advent Calendar² 等に譲り、本稿では解説しない。基本的には Stan の記法に従って確率モデルをそのままスクリプトとして記述するだけなので、本稿ではそれぞれの確率モデルの意味について重点的に解説しつつ、Stan で分析するときのテクニックや間違いやすい点についても随時説明する。

確率モデルと Stan スクリプトの対応をわかりやすくす

るために、本稿では Stan の記法になるべく近い表記法で数式を表し、計算効率よりも可読性に配慮して Stan スクリプトを記述した (したがって、本稿の数式は必ずしも一般的な記法とは一致しない)。また、慣例に倣ってパラメータをギリシャ文字、観測変数を英文字で表した。事前分布の決め方については様々な議論があるが (e.g., van de Schoot et al., 2021), 本稿ではすべての分析において、パラメータが取りうる範囲内で十分に幅の広い一様分布を事前分布として採用した (Stan では事前分布の指定を省略すると自動的にこのような一様分布が使われる)。

R と Stan のスクリプトとデータは筆者の Open Science Framework (OSF) のリポジトリからダウンロードできる (<https://doi.org/10.17605/osf.io/2zxs6>)。リポジトリの最上位ディレクトリには R スクリプト (.R) と7つのフォルダが設置されている。data フォルダには分析に使用する4種類のデータセット (.csv), stanmodel フォルダには11個の Stan スクリプト (.stan) が格納されている。Stan スクリプトは同名の R スクリプトから呼び出して分析を実行することができる。estimate フォルダには各モデルによる推定結果の要約 (.csv), plot から始まる3つのフォルダには結果を要約したグラフ (.png), stanfitRDS フォルダにはMCMCの結果を格納したファイル (.rds) が含まれている。これらは筆者の分析環境 (Windows 10 Pro, R 4.0.3, rstan 2.21.2, Rtools40) で推定した結果である。

本稿の流れは以下の通りである。次の節では、統計モデリングを行ううえで扱いやすいデータセットの作り方について簡単に解説する。続く3つの節では、3種類の実験パラダイムから得られた実データを使った階層ベイズモデリングのチュートリアルを提供する。1つ目の例として、心的回転課題の正反応時間データに対する様々な階層線形モデリングを紹介し、階層ベイズモデリングの基本的な考え方を丁寧に解説する。2つ目の例として、2つの心理物理実験 (恒常法と上下法による錯視量の測定) で得られた二値反応データに対する階層プロビットモデルを紹介する。最後の例として、フランカー課題で得られた反応時間と二値反応を同時に説明できる拡散過程モデル (drift diffusion model) の例を示す。最後の節では本稿の内容を総括しつつ、モデル選択やオープンデータの役割など、統計モデリングを实践するうえで重要と思われるトピックについて簡潔に議論する。

データセットの作り方

モデリングの具体例を示す前に、扱いやすいデータセットの形式について解説しておく。伝統的な t 検定や

¹ より一般には、パラメータに確率分布を仮定したモデリングを広く階層モデリングと呼ぶ。したがって、階層構造を持たないデータに対して階層モデリングを適用することも可能であるが、本稿ではそのような例は扱わない。

² Stan Advent Calendar 2020 では有志達が Stan Advent Boot Camp と称して初心者向けの Stan の解説を行っている (<https://qiita.com/advent-calendar/2020/stan>)。

Table 1.
Data structure commonly used for conventional analysis.
(mental rotation data as an example).

participant	rt_0deg	rt_40deg	rt_80deg	rt_120deg	rt_160deg
1	0.809	0.776	0.757	0.972	1.220
2	0.663	0.979	1.104	1.868	1.866
3	0.665	0.885	1.109	1.250	1.518
⋮	⋮	⋮	⋮	⋮	⋮
24	0.702	0.868	0.996	1.325	1.595

Table 2.
Structure of trial-by-trial data from a mental rotation experiment.
(mental rotation.csv).

serial	participant	trial	angle	is.correct	rt
1	1	1	80	1	0.838
2	1	6	40	1	0.660
3	1	10	80	1	0.845
⋮	⋮	⋮	⋮	⋮	⋮
1503	11	573	160	1	3.554
1504	12	21	160	1	3.800
1505	12	22	120	1	2.586
⋮	⋮	⋮	⋮	⋮	⋮
3200	24	572	40	1	0.814
3201	24	574	0	1	0.563
3202	24	575	80	1	0.973

分散分析を用いて実験データを分析するときには、参加者と条件の組み合わせごとに反応時間や正否などを集計し、縦方向に参加者、横方向に条件を並べたデータセットを作るのが一般的である。Table 1は心的回転実験で得られた正反応時間（単位は秒）の平均値を参加者と条件（角度差）の組み合わせごとに計算して作ったデータの例である。例えばこのデータを使って1要因5水準の反復測定分散分析を実行することにより平均反応時間に対する角度差の主効果を検定できる。

一方、階層モデリングでは試行レベルのデータを潰さずにそのまま利用するため、Table 2のように試行を縦方向に、変数を横方向に並べたデータセット（c.f., mental-rotation.csv）を分析に使用する。この表のserialは行番号、participantは参加者番号、trialは参加者ごとの試行番号（分析に含めない試行が除外されているため飛び飛びになっている）、angleは各試行の条件（角度差）、is.correctは反応の正否（誤答試行を除外しているためこのデータでは全て1）、rtは各試行の反応時間（単位は秒）を表す。実験プログラムから出力されるローデータはこの形式であることが多いと思われる。別の例として、心理物理学的測定法の恒常法を使って得られたミュラー・

リヤー錯視実験の試行ごとのデータをTable 3に示す。conditionは各試行で提示された標準刺激（内向図形か外向図形か）、c_lengthは比較刺激の長さ（標準刺激の物理的な長さを1としたときの比）、responseは参加者が標準刺激と比較刺激のどちらを選択したかを表す変数である。このように、各観測が各行、各変数が各列に対応し、1つの表に異なる型の観測単位が含まれない（e.g., 心的回転データと錯視データを混ぜない）といった条件を満たすデータセットを整然データ³（tidy data; Wickham, 2014）と呼ぶ。階層モデリングに限らず、整然データは分析をするときに一般的に扱いやすい形式なのでぜひ覚えておいて欲しい。

本節の最後に、データセットを作るときのコツを3つ紹介する。1つ目に、反応時間はミリ秒ではなく秒にしてStanに渡した方がよい（c.f., Table 2）。これは、データのスケールが大きいとパラメータの推定効率が低下して

³ 整然データという訳は西原史暁氏のブログ記事を参考にした（<https://id.fnshr.info/2017/01/09/tidy-data-sei-zen/>）。同氏による整然データの解説も参考になる（<https://id.fnshr.info/2017/01/09/tidy-data-intro/>）。

Table 3.
Structure of trial-by-trial data collected using the method of constant stimuli (constant.csv).

serial	participant	trial	condition	c_length	response
1	1	1	inwards	1.2	comparison
2	1	2	inwards	0.9	comparison
3	1	3	outwards	1.1	standard
⋮	⋮	⋮	⋮	⋮	⋮
2288	13	176	outwards	1.5	comparison
2289	14	1	outwards	0.5	standard
2290	14	2	inwards	1.2	comparison
⋮	⋮	⋮	⋮	⋮	⋮
4398	25	174	outwards	1.0	standard
4399	25	175	outwards	1.0	standard
4400	25	176	inwards	0.6	standard

しまう場合が多いからである。モデルによってはデータを標準化することによって推定効率を高められることもあり、事前分布を設定するうえで便利な場合もあるが、前処理や後処理が少し面倒であり、混乱を招く恐れもあるため本チュートリアルでは採用しなかった。2つ目に、名義変数の値（ラベル）はなるべくわかりやすく命名しよう。名義変数は分析時には数値に置き換える必要があるが（e.g., 内向図形条件=1, 外向図形条件=2）、最初から数値化してしまうとどの値が何を表すのかが分からなくなり分析や解釈を誤ってしまうリスクが増える。筆者はデータセット上では名義変数のラベルを英単語で表し（c.f., Table 3）、データをStanに渡すときにRで数値化することになっている。Stanで名義変数を宣言するときには必ずコメントで数値とラベルの対応を書いておこう。最後に、データセットを作るときにはなるべく情報量を削らないように気を付けるとよい。本チュートリアルでは説明の便宜と混乱の防止のために不要な行と列をあらかじめ除外したデータを用意したが、前処理の過程もスクリプトとして残しておいたほうが間違いに気づきやすく、後で同じデータセットを使って別の分析を行うときにも便利である。

心的回転データのモデリング

最初の分析例として、本節では心的回転のデータを使って階層ベイズモデリングの基本的な考え方を解説する。典型的な心的回転課題（Shepard & Metzler, 1971）では、参加者は向きが異なる2つの物体を見せられ、それらが同一の物体であるか互いに鏡像の関係にあるのかをなるべく速く正確に判断することが求められる。このような実験を行うと、2物体が同一であると回答するのに要する正反応時間の平均値は2物体の角度差の線形関数になることが知られている。この結果は、人が物体のイ

メージを等角速度で回転させた結果であると解釈される。このような心的回転課題のデータに階層ベイズモデリングを適用してみよう。この目的のために、本節ではMuto & Nagai (2020)の実験データ（ $N=24$ ）から誤答試行と外れ値を除外し、統制条件（元の論文における人型・顔なし条件）かつ同一物体のペアが提示された試行のみを抜き出した計3,202試行から成るデータセット（mentalrotation.csv; Table 2はその抜粋）を使用する⁴。

個人差を無視したモデル

階層モデルを考える前に、まずは個人差を無視した単純な非階層モデルを考えると分かりやすい。個人差を無視するというのは、心的回転のいかなる処理過程においても個人差が全く存在しないと仮定するという意味である。別の言い方をすれば、データをプールしてあたかも1人の参加者から全3,202試行分のデータを収集したかのようにみなして分析をするのと同義である。このような想定の下で、

$$\theta_i = \alpha + \beta x_i \quad (1)$$

$$Y_i \sim \text{Normal}(\theta_i, \eta) \quad (2)$$

というシンプルな単回帰モデルを考えてみよう（mr1_nonhier.stan）。添え字の i は行番号（1-3,202）であり、式（1）の x_i は i 行目の試行で提示された刺激の角度差（0度・40度・80度・120度・160度のいずれか）を表す。式（1）は角度差 x_i の線形関数を表しており、切片 α は符号化やキー押しといった心的回転以外の処理に要する時間、傾き β は1度の心的回転に要する時間として解釈で

⁴ 除外前の試行と変数を含む元のデータは<https://doi.org/10.17605/osf.io/mncde>からダウンロードできる。

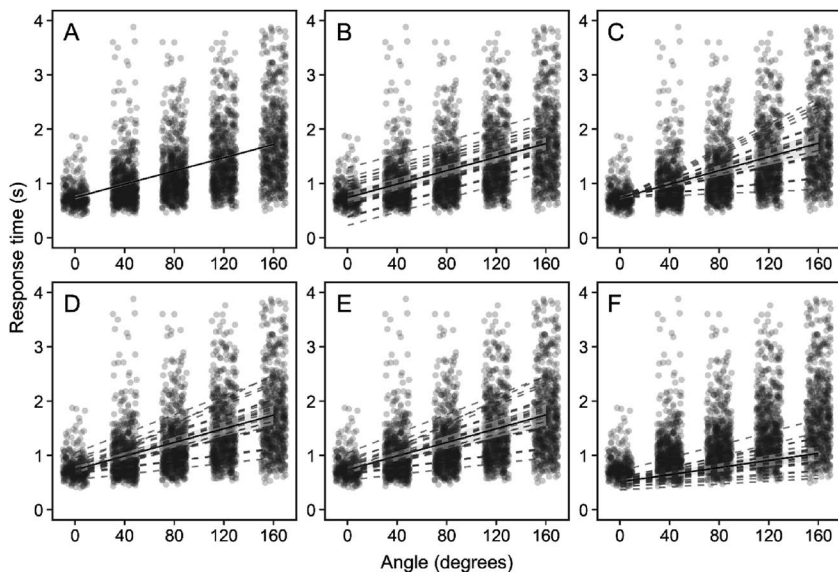


Figure 1. Plots of mental rotation data and linear functions estimated by six models described in the main text (corresponding to Panel A-F). Each point shows correct response times for “same” pairs (a total of 3,202 trials). Solid lines and error bands indicate mean regression lines and their 95% credible intervals. Dashed lines are regression lines estimated per participant ($N=24$).

きる。 α と β に添え字が付いていないのは、切片と傾きが試行や参加者によらず常に一定であるというモデルの仮定を反映している。 θ_i は*i*行目の試行における反応時間の予測値（このモデルの場合は平均値）であり、例えば*i*=3とき、Table 2より $\theta_3=\alpha+80\beta$ となる。式(2)の Y_i は*i*行目の試行で観測された反応時間を表しており⁵、平均 θ_i ・標準偏差 η （分散は η^2 ）の正規分布から生成されると仮定する。標準偏差 η は試行間変動の大きさを表すパラメータとして解釈できる。ちなみに、式(1)を式(2)に代入して

$$Y_i \sim \text{Normal}(\alpha + \beta x_i, \eta) \quad (3)$$

という1つの式にまとめてもモデルとしては等価であるが、本チュートリアルでは分かりやすさを優先し、式(1)、(2)のような冗長な表現を採用した。

このモデルに含まれるパラメータは α ・ β ・ η の3つである。これらを3,202試行分の角度差と反応時間のペアのデータから推定することを試みる。Stanで推定を行うためには、dataブロックとparametersブロックでデータとパラメータを宣言したうえでtransformed parametersブ

ロックとmodelブロックにモデル式を記述すればよい。標準偏差を表す η は負の値にはなり得ないので、parametersブロックで宣言するときには`real<lower=0>eta;`と下限を指定しておく。 α と β も理論上は正の値になることが期待されるが、推定上は負の値になっても問題がないので今回の分析では範囲に制限を設けないことにした。transformed parametersブロックとmodelブロックの区別は初心者には迷いやすいが、確率分布からのサンプリング(~)を含む式はmodelブロック、それ以外はtransformed parametersブロックに書くことで覚えておけば基本的には問題ない⁶。例えば、式(1)は等式であり確率分布からのサンプリングではないのでtransformed parametersブロックに、式(2)は正規分布からのサンプリングを表すのでmodelブロックに記述する。式(3)を使って記述する場合にはtransformed parametersブロックは不要である。

このモデルによって推定された回帰直線をローデータに重ねたものがFigure 1Aである。実線で示した回帰直線は、角度差ごとに推定された平均反応時間 θ_i の事後分布の期待値（expected a posteriori; 以下EAP）を結んだも

⁵ 正確に言えば、 Y_i は*i*行目の反応時間の観測値 y_i を実現値として返す確率変数であるが、本稿では混乱を避けるために両者を厳密には区別しないことにする。

⁶ modelブロックで変数を宣言して等式を記述することも可能ではあるが、Stanに慣れるまではお勧めしない。

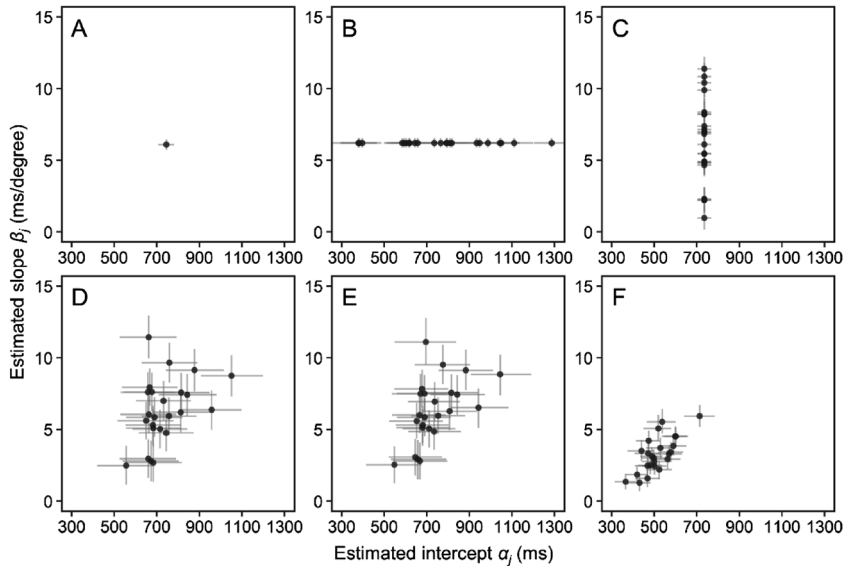


Figure 2. Each model's estimates of the intercept and slope parameters for each participant ($N=24$) from mental rotation data. Points and error bars show expected a posteriori (EAP) estimates and 95% credible intervals, respectively.

のである。今回のモデルでは個人差が一切存在しないという仮定を置いたため、全参加者の予測値がこの1本の直線で表されている。切片 α のEAPは745 ms (95%確信区間は708–781)、傾き β のEAPは6.1 ms/度 (95%確信区間は5.7–6.5)であった (Figure 2A)。これらの推定値も全参加者で共通であるため、どの参加者も心的回転以外の処理に約745 ms、イメージを1度回転するのに約6.1 msを要すると解釈されることになる。しかし、心的回転の処理過程に個人差が存在しないという仮定は現実的ではない。そこで、以降の項ではこの単純な単回帰モデルを叩き台に、切片 α と傾き β の個人差を表現するための階層モデルを構築していく⁷。

切片の個人差を含むモデル

前項で紹介したモデルを修正し、切片 α に個人差があることを仮定してみよう。傾き β については参加者間で共通であると考え。このような仮定は

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \quad (4)$$

$$\theta_i = \alpha_{j(i)} + \beta x_i \quad (5)$$

$$Y_i \sim \text{Normal}(\theta_i, \eta) \quad (6)$$

⁷ 試行間変動を表すパラメータ η の個人差をモデル化することも可能であるが、簡単のために本節では考えないことにする。

というモデル式で表せる (mr2_hier_intercept.stan)。前項のモデルとの最も重要な相違点は、切片 α に添え字が付いたことである (式 (5) を式 (1) と比較せよ)。添え字の j は参加者番号 (1–24) を、 $j(i)$ は i 行目の試行の参加者番号を表している。例えば $i=1,504$ のとき、Table 2より1,504行目の参加者番号は12であるから $\alpha_{j(i)} = \alpha_{j(1504)} = \alpha_{12}$ である。式 (4) は、このように添え字を使って表された24名分の $\{\alpha_1, \alpha_2, \dots, \alpha_{24}\}$ がそれぞれ独立に平均 μ_α 、標準偏差 σ_α の正規分布に従うという仮定である⁸。簡単に言えば、切片 α_j の個人差に正規分布を仮定したということである。 μ_α は α_j の参加者間平均、 σ_α は個人差の大きさとして解釈できる。個人差を表現した正規分布から α_j が得られ (式(4))、さらに α_j の式 (式(5)) をパラメータとして含む正規分布から試行ごとの反応時間の観測値が得られる (式(6)) という意味で階層的である。一方、添え字のない β と η に関しては参加者間で共通であることが仮定されている。

式(5)の $\alpha_{j(i)}$ をStanで表現するには、参加者番号を表す変数をparとでも置き、alpha[par[i]]と記述すれば

⁸ ベイズ推定の枠組みでは、未知のパラメータ (μ_α と σ_α) を含む正規分布を α_j の事前分布として仮定したと考えるができる。また、 α_j をローカルパラメータ、 μ_α や σ_α をグローバルパラメータと呼ぶことがある。本稿では混乱を避けるためにローカルパラメータに対しては事前分布という用語を用いないことにした。

よい。このparはTable 2のparticipant列に相当する変数であるため、要素の数は参加者の人数(24)ではなくデータの行数(3,202)であることに注意しよう。par[i]はi行目の参加者番号(1-24)となる。例えばi=1,504のとき、alpha[par[1504]]はalpha[12]と同じ意味である。なお、 α_j の個人差を表す式(4)は確率分布からのサンプリングであるため式(6)と同様modelブロックに記述する。

このモデルによって推定された回帰直線をFigure 1Bに示す。グラフ中の実線は参加者間平均 μ_α に基づいてプロットした平均的な回帰直線、破線は参加者ごとの α_j に基づく24名分の回帰直線である。このモデルでは β に個人差がないことを仮定したので回帰直線の傾きはどの参加者でも等しいが、切片 α_j の個人差によって上下方向にばらついていることが分かる。Figure 2Bには切片と傾きの推定値の個人差をプロットしている。傾きは全員同じ値(EAPは6.2 ms/度)であるが、切片の参加者間平均のEAPは741 ms、標準偏差のEAPは277 msであり、参加者ごとの切片の推定値はおおよそ400-1300 msの範囲に散らばっている。

傾きの個人差を含むモデル

次は切片 α を参加者間で共通とし、傾き β のみに個人差を仮定した階層モデルを考えてみよう。考え方は先ほどと同様で、 β に添え字jを付けて、 β_j の個人差に正規分布を仮定すればよい。モデル式は

$$\beta_j \sim \text{Normal}(\mu_\beta, \sigma_\beta) \quad (7)$$

$$\theta_i = \alpha + \beta_{j(i)} x_i \quad (8)$$

$$Y_i \sim \text{Normal}(\theta_i, \eta) \quad (9)$$

と表せる(mr3_hier_slope.stan)。 μ_β は傾きの参加者間平均、 σ_β は傾きの個人差の大きさを表す。添え字のない α は参加者間で共通である。

このモデルによる回帰直線の推定結果をFigure 1Cに、参加者ごとの切片と傾きの推定値をFigure 2Cに示す。切片すなわち0度のときの反応時間の予測値は参加者間で共通であるが(EAPは735 ms)、傾きの参加者間平均のEAPは6.3 ms/度、標準偏差のEAPは3.0 ms/度であり、参加者ごとの傾きの推定値はおおよそ1-11 ms/度の範囲に散らばっている。

切片と傾きの個人差を含むモデル

前2項のモデルでは切片と傾きの片方にのみ個人差を

仮定したが、両方に個人差を仮定することももちろん可能である。最も単純な表現としては、 α と β の両方に添え字jを付けて、それらが別々の正規分布に従うと仮定することであり、

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \quad (10)$$

$$\beta_j \sim \text{Normal}(\mu_\beta, \sigma_\beta) \quad (11)$$

$$\theta_i = \alpha_{j(i)} + \beta_{j(i)} x_i \quad (12)$$

$$Y_i \sim \text{Normal}(\theta_i, \eta) \quad (13)$$

とモデル化することができる(mr4_hier_both_nocorr.stan)。

ただし、このモデルでは α_j と β_j が独立した正規分布から得られることが仮定されているため、 α_j と β_j は無相関である。心的回転のスピードが速い人(β_j が小さい)ほど符号化やキー押しの時間も短い(α_j が小さい)といった相関関係に関心がある場合には、多変量正規分布を仮定し、

$$\Sigma = \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta}^2 \\ \sigma_{\alpha\beta}^2 & \sigma_\beta^2 \end{pmatrix} \quad (14)$$

$$\alpha_j, \beta_j \sim \text{MultiNormal}((\mu_\alpha, \mu_\beta), \Sigma) \quad (15)$$

$$\theta_i = \alpha_{j(i)} + \beta_{j(i)} x_i \quad (16)$$

$$Y_i \sim \text{Normal}(\theta_i, \eta) \quad (17)$$

とモデル化する必要がある(mr5_hier_both_corr.stan)。ここで、式(15)の右辺の第1引数 (μ_α, μ_β) は切片と傾きの参加者間平均を要素として持つベクトルである。また第2引数の Σ は分散共分散行列であり、対角成分の σ_α^2 と σ_β^2 がそれぞれ α_j と β_j の個人差の分散を、残りの $\sigma_{\alpha\beta}^2$ が α_j と β_j の共分散を表す。これらの値から

$$\rho = \frac{\sigma_{\alpha\beta}^2}{\sigma_\alpha \sigma_\beta} \quad (18)$$

という式を利用して相関係数 ρ を求めることができる。このようにパラメータの推定値を使って生成量を計算するときにはgenerated quantitiesブロックを使うとよい。変数が3つ以上ある場合も同様に多変量正規分布を使って表現できる。

相関を仮定しないモデル(式(10)-(13))による推定結果をFigure 1DおよびFigure 2Dに相関を仮定したモデル(式(14)-(17))による推定結果をFigure 1EおよびFigure 2Eに示す。どちらのモデルにおいても切片と傾き

の両方の個人差を推定できていることが分かる。推定結果はモデル間でほとんど変わらないが、相関を仮定したモデルでは相関係数 ρ のEAPが 0.338 、95% 確信区間が -0.026 – 0.788 と推定された。

指数-正規分布モデル

これまでのモデルでは反応時間の試行間変動が従う確率分布として標準偏差 η の正規分布を仮定してきたが、反応時間分布は一般には正の方向に歪んだ（右側の裾が長い）分布になることが知られている。実際、Figure 3に示したローデータのヒストグラムからも、今回の反応時間データが正の方向に歪んでいることが確認できる。そこで本項では、正に歪んだ反応時間分布を表現するために、試行間変動として正規分布の代わりに指数-正規分布（exponentially modified Gaussian distribution または ex-Gaussian distribution）を仮定したモデルを考えてみよう。指数-正規分布は正規分布に従う確率変数と指数分布に従う確率変数の和の分布であり、右側の裾の長さを指数分布のパラメータ λ によって表すことができる（ λ が小さいほど裾が長い）。前項で紹介した相関を仮定したモデル（式（14）–（17））の Y_i が従う確率分布（式（17））を正規分布から指数-正規分布に変更し、

$$\Sigma = \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha,\beta}^2 \\ \sigma_{\alpha,\beta}^2 & \sigma_\beta^2 \end{pmatrix} \quad (19)$$

$$\alpha_j, \beta_j \sim \text{MultiNormal}((\mu_\alpha, \mu_\beta), \Sigma) \quad (20)$$

$$\theta_i = \alpha_{j(i)} + \beta_{j(i)} x_i \quad (21)$$

$$Y_i \sim \text{ExpModNormal}(\theta_i, \eta, \lambda) \quad (22)$$

とモデル化するだけでよい（mr6_hier_exgaus.stan）。

Stanで指数-正規分布を実装するにはexp_mod_normal という関数を使って $Y[i] \sim \text{exp_mod_normal}(\mu[i], \eta, \lambda)$ と書けばよい。指数-正規分布のパラメータをStanで推定しようとすると“Rejecting initial value”というエラーメッセージが大量に表示されてサンプリングが開始しないことがよくあるが、これは、指数-正規分布の確率密度関数に含まれる相補誤差関数 $\text{erfc}((\theta_i + \lambda\eta^2 - x)/(\eta\sqrt{2}))$ の引数の値が大きすぎると計算上0の値を返し、その結果として確率密度が0となってしまう対数尤度が負の無限大に発散してしまうことが原因であると考えられる。これを防ぐために、指数-正規分布のパラメータの初期値としてなるべく小さな値をR側で指定してやるとよい（mr6_hier_exgaus.Rを参照）。

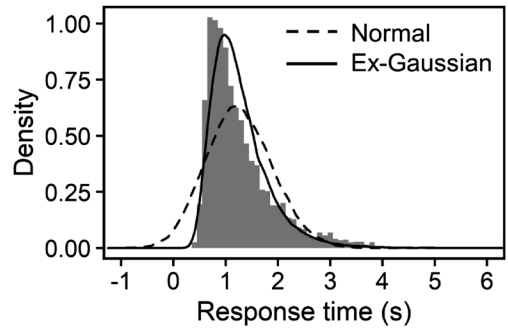


Figure 3. A histogram of correct response time data of the mental rotation task (a total of 3,202 trials) and posterior predictive distributions derived from the normal distribution model (the dashed line) and the ex-Gaussian distribution model (the solid line).

Figure 3は、全データのヒストグラムの上に破線で正規分布モデル（式（14）–（17））の事後予測分布を、実線で指数-正規分布モデル（式（19）–（22））の事後予測分布を重ねたものである。やはり指数-正規分布を仮定したほうがデータへの当てはまりはよいようである。Figure 1Fは指数-正規分布モデルによる回帰直線（式（21））の推定結果を示している。正規分布を仮定したこれまでのモデルと比べて回帰直線の位置がずいぶん低くなっているように見えるが、これは、指数-正規分布の平均が θ_i ではなく $\theta_i + 1/\lambda$ であることに由来する。平均値の回帰直線を引きたければ θ_i の代わりに $\theta_i + 1/\lambda$ のEAPをプロットすればよい。どちらの回帰直線を使うべきかは分析の目的次第である。場合によっては中央値や最頻値に基づく回帰直線を引くこともあり得るだろう（正規分布を仮定した場合は平均値と中央値と最頻値は一致する）。Figure 2Fには指数-正規分布モデルによって推定された参加者ごとの α_j と β_j の推定値を示している。これまでのモデルよりも推定値が小さい理由は回帰直線のときと同様で、平均値ではなく θ_i を構造化したからである（式（21）、（22））。興味深いことに、 α_j と β_j の相関係数は正規分布モデルを用いたときよりも大きな値となった（EAPは 0.578 、95% 確信区間は 0.110 – 0.886 ）。

本節のまとめ

本節では心的回転課題の正反応時間データを例に様々な階層線形モデルを紹介した。伝統的な分析法との対応で言えば、最初に紹介した個人差を無視したモデルは単回帰分析、その後で紹介した切片と傾きに個人差を仮定した正規分布モデルはLMMに相当する。また正規分布で表現できないデータを分析するときにはGLMMがよ

く使われる。最後に紹介した指数-正規分布モデルも適切なパッケージを使えば最尤法で解くこともできる。こういった、名前の異なる分析を統一的な方法で柔軟に行うことができるのがベイズ統計モデリングの利点の1つである。相関をゼロに固定する、パラメータに等値制約を与える、目的に適した確率分布を選択する、といった工夫も単に確率モデルを書き換えるだけで実装できる。本節では最も単純な線形モデルを使って階層ベイズモデリングの基礎を解説したが、この手法がより複雑なモデルにも適用可能であることを以降の節で示したい。

心理物理学的測定データのモデリング

本節では心理物理学的測定法を用いた2つの錯視実験の二値反応データに対する階層モデリングの例を紹介する。1つ目のデータセットは武藤(2017)がオンラインで恒常法を用いて収集したミュラー・リヤー錯視実験のデータ(constant.csv)である。この実験では、2種類の錯視図形(内向図形と外向図形)の主線の長さの主観的等価点(物理長に対する比)が測定された。実験の各試行では標準刺激(内向図形と外向図形)と比較刺激(50-150%の長さの線分11種類)が1つずつ左右に提示され、参加者はどちらの線分がより長く見えるかをキー押して回答した。全ての組み合わせ(2×11通り)が8回ずつランダムな順に提示され、計4,400試行(176試行×25名)分のデータが得られた。

2つ目のデータセットはMuto, Ide, Tomita, & Morikawa (2019)が上下法(階段法とも呼ばれる)を用いて収集した、アイシャドウによって目が大きく見える錯視に関する実験データ(updown.csv)である。この実験では女性の顔の3Dモデルを使い、上瞼にアイシャドウが塗られた顔と塗られていない顔の目の大きさの主観的等価点(物理的な直径に対する拡大率)を測定した。実験の各試行では標準刺激(アイシャドウあり・なし)と比較刺激(目の直径が92-108%のアイシャドウなし画像17種類)が1つずつ左右に提示され、参加者は目がより大きく見える方をキー押して選択した。比較刺激の目の大きさが最大(108%)の試行から開始する下降系列と最小(92%)の試行から開始する上昇系列の2種類が織り交ぜられ、参加者の反応に応じて1ポイントずつ比較刺激の目の大きさが変化した。各系列は参加者の選択が6回反転した時点で打ち切りとなったため、参加者によって総試行数は異なっていた。元の実験では顔の向きが様々な操作されたが、本節では実験1の正面顔条件のデータ(計1,312試行, $N=20$)のみを分析対象とする。

個人差を無視したモデル

まずは恒常法で収集されたミュラー・リヤー錯視実験のデータ(constant.csv)のみを使い、個人差を仮定しないモデルを考えよう。恒常法実験で得られる二値反応の生成過程として、

$$Y_i = \begin{cases} 0, & \text{if } x_i < \phi_{k(i)} + \varepsilon_i \\ 1, & \text{if } x_i \geq \phi_{k(i)} + \varepsilon_i \end{cases} \quad (23)$$

$$\varepsilon_i \sim \text{Normal}(0, \eta) \quad (24)$$

というモデルを想定することができる。ここで、 Y_i は*i*行目の試行において比較刺激(線分)が選択されたとき1、標準刺激(錯視図形)が選択されたとき0を返す従属変数である。また、 x_i は*i*行目の試行で提示された比較刺激(線分)の物理的な長さを表す。標準刺激は内向図形と外向図形の2種類であったので、それぞれの主観的な長さ(主観的等価点)を ϕ_1, ϕ_2 というパラメータで表す。式(23)の $\phi_{k(i)}$ の添え字 $k(i)$ は*i*行目の試行で提示された錯視図形の種類(内向図形のとき1, 外向図形のとき2)を表す。したがって、 $\phi_{k(i)}$ は*i*行目の試行で提示された錯視図形の主観的な長さ(ϕ_1 または ϕ_2)を表す。参加者は、提示された比較刺激の物理的な長さ x_i と標準刺激の主観的な長さ $\phi_{k(i)}$ を比較して、より長い方を選択すると考えられる。しかし、実際には刺激や生理状態などに由来するノイズによって、全く同じ刺激のペアに対して異なる反応を示すこともあるため、平均0・標準偏差 η の正規分布に従う攪乱項 ε_i が加えられている。このモデル式のままだと従属変数 Y_i の式が確率分布で表現されておらず、尤度の計算が難しいので、等価な表現である

$$\pi_i = p(x_i \geq \phi_{k(i)} + \varepsilon_i) = \text{Normal_cdf}(x_i, \phi_{k(i)}, \eta) \quad (25)$$

$$Y_i \sim \text{Bernoulli}(\pi_i) \quad (26)$$

というモデル式を実装する(pp1_cm_nonhier.stan)。Normal_cdfは正規累積分布関数(その逆関数はプロビット関数と呼ばれる)であり、 π_i は*i*行目の試行で比較刺激が選択される確率を表す。ベルヌーイ分布に従い、確率 π_i で $Y_i=1$ 、確率 $1-\pi_i$ で $Y_i=0$ となる。ちなみにこのモデル式はプロビット回帰分析のモデルと等価である。

Stanを使ってこのモデルを当てはめてみよう。攪乱項の標準偏差を表す η は非負のパラメータなので、parametersブロックで宣言する際に下限を0に指定しておく。標準刺激の主観的な長さを表す ϕ_1 と ϕ_2 は負の値になっても推定上は問題ないが、手続き上0%未満になること

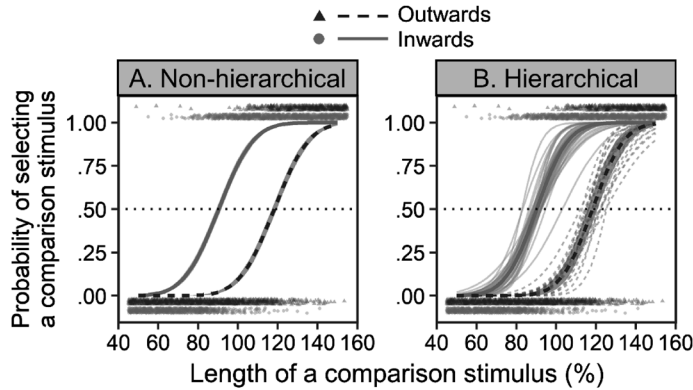


Figure 4. Psychometric curves fitted to the binary data from the Müller-Lyer illusion experiment with the method of constant stimuli. Thick lines and error bands indicate mean psychometric curves and their 95% credible interval. Thin lines shown in Panel B represent psychometric curves estimated per participant.

はあり得ないので下限を0に指定した⁹。確率分布を含まない式 (25) は transformed parameters ブロックに¹⁰、確率分布を含む式 (26) は model ブロックに記述する。式 (25) の $\phi_{k(i)}$ を Stan で書くときは、各試行で提示された標準刺激の種類を表す二値変数（内向図形=1, 外向図形=2）を cond とでも名付け、phi[cond[i]] と記述すればよい。このモデルを使って推定された心理測定関数が Figure 4A である。主観的等価点の EAP は内向図形で 90.3%（95% 確信区間は 89.1–91.2%）、外向図形で 118.2%（95% 確信区間は 117.0–119.4%）であった。

個人差を考慮したモデル

前項の非階層モデル（式 (25), (26)）を階層モデルに拡張してみよう。ここでは内向図形と外向図形の主観的な長さ ϕ_1, ϕ_2 に加え、攪乱項の標準偏差 η にも個人差があると仮定し、これらに参加者番号を表す添え字の j を加え、さらにこれらのパラメータ間に相関があると考えてみよう。前節の心的回転データの例で示したのと同様のやり方で、これらのパラメータが多変量正規分布に従

うと仮定し、

$$\phi_{j,1}, \phi_{j,2}, \eta_j \sim \text{MultiNormal}((\mu_{\phi_1}, \mu_{\phi_2}, \mu_{\eta}), \Sigma) \quad (27)$$

$$\pi_i = \text{Normal_cdf}(x_i, \phi_{j(i), k(i)}, \eta_{j(i)}) \quad (28)$$

$$Y_i \sim \text{Bernoulli}(\pi_i) \quad (29)$$

とモデル化する（pp2_cm_hier.stan）。式 (27) の右辺の第1引数 $(\mu_{\phi_1}, \mu_{\phi_2}, \mu_{\eta})$ は各パラメータの参加者間平均を要素として持つベクトル、第2引数 Σ は 3×3 の分散共分散行列を表す。ただし、 η_j は非負のパラメータであり、また $\phi_{j,1}$ と $\phi_{j,2}$ も非負であると宣言しているため、正確に言えば式 (27) の右辺は多変量正規分布ではなく、実現値の取りうる範囲が0以上に制限されているゼロ切断（zero-truncated）多変量正規分布である。Stan では parameters ブロックで範囲を指定するだけで自動的に切断分布からのサンプリングを行ってくれるので、事後分布のみに興味がある場合には特に意識する必要はない¹¹。

この階層モデルによって推定された心理測定関数を Figure 4B に示す。先ほどの非階層モデルとは異なり、参加者の人数分だけ心理測定関数が推定されている（太線は参加者間平均）。曲線の横位置（ $\phi_{j,1}$ と $\phi_{j,2}$ に対応）と形（ η_j に対応）の両方が参加者によって異なることが確認できる。主観的等価点の参加者間平均の EAP は内向図形（ μ_{ϕ_1} ）で 90.2%（95% 確信区間は 87.7–92.8%）、外向図形（ μ_{ϕ_2} ）で 118.2%（95% 確信区間は 115.8–120.6%）

⁹ パラメータの取りうる範囲が自明であるときには範囲を指定したほうが推定効率が良くなる場合があり、またデータに不備があったときにエラーが出て気付きやすいといった利点もある。ただし、範囲が自明でない場合は要注意である。例えば、心的回転の効率（回帰直線の傾き）やストループ効果の大きさ（不一致条件と一致条件の反応時間の差分）は正であることが理論的に期待されるが、負の値を示すケースが全くないとは断言できない。

¹⁰ 式 (25) の Normal_cdf は累積確率を返す関数であり確率分布（確率密度関数）ではないためチルダ（～）ではなく等号（＝）が使われている。

¹¹ ただし、WAIC や周辺尤度を求めるときなど尤度を正確に計算する必要がある場合には切断分布であることを考慮した記法を用いなければならない。

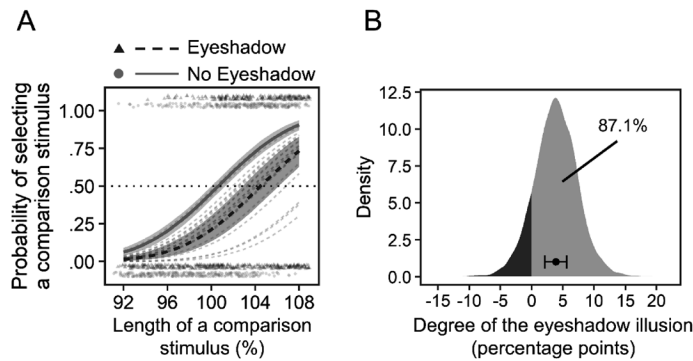


Figure 5. Summary of results from hierarchical modeling of the eyeshadow illusion experiment with the up-down method.

(A) Psychometric curves fitted to the binary data from the eyeshadow illusion experiment with the up-down method. Thick lines and error bands indicate mean psychometric curves and their 95% credible interval. Thin lines represent psychometric curves estimated per participant. (B) A posterior predictive distribution showing the inter-individual variability in the degree of the eyeshadow illusion (the difference in the point of subjective equality between faces with and without eyeshadow). The point and the error bar show the expected a posteriori (EAP) estimate and 95% credible interval (CI) of the mean degree of the illusion. The light area corresponds to the estimated proportion of participants who show positive values of the illusion.

であった。また、 $\phi_{i,1}$ と $\phi_{i,2}$ の相関係数のEAPは.030 (95%確信区間は-.565-.596)であったため、内向図形と外向図形の主観的等価点に相関があるとは言えそうにない。

次はアイシャドウによる目の過大視錯視の実験データを階層モデルで分析してみよう。上下法を用いたこの実験では比較刺激の提示順序がランダムではなく1段階ずつ変化し、また主観的等価点に近い比較刺激がそうでない比較刺激よりも多く提示される適応的方法であるという点で恒常法と異なるが、課題の基本的な構造は恒常法と同様であるため、先ほどと同じ発想でモデル化できる。そこで、先のモデル(式(27)–(29))にいくつかの仮定を追加し、

$$\delta_j \sim \text{Normal}(\mu_\delta, \sigma_\delta) \quad (30)$$

$$\pi_i = \begin{cases} \text{Normal_cdf}(x_i, \phi, \eta), & \text{if } s_i = 1 \\ \text{Normal_cdf}(x_i, \phi + \delta_{f(i)}, \eta), & \text{if } s_i = 2 \end{cases} \quad (31)$$

$$Y_i \sim \text{Bernoulli}(\pi_i) \quad (32)$$

というモデルを考えた(pp3_ud_hier.stan)。 s_i は*i*行目の試行で提示された標準刺激がアイシャドウなし(統制刺激)のとき $s_i=1$ 、アイシャドウあり(実験刺激)のとき $s_i=2$ となる二値変数である。今回の分析では、実験刺激と統制刺激の主観的等価点の差分として定義される錯視量に関心があるため、統制刺激の主観的等価点を ϕ 、錯視量を δ_j とパラメータ化した。アイシャドウなし条件は目の大きさのみが物理的に異なる2つの顔を比較する

という非常に単純な課題であり、課題が適切に遂行されれば主観的等価点は全員同じ値(100%に近い値)に落ち着くことが予想されるため、 ϕ は階層化せず参加者間で共通とした。また、比較刺激の提示回数に偏りがある上下法では攪乱項の標準偏差を参加者と条件の組み合わせごとに高精度で推定するのが難しく、分析上の関心も薄いので、今回は η を参加者間・条件間で共通とした。以上の理由から、錯視量 δ_j のみを階層化し、その個人差が平均 μ_δ 、標準偏差 σ_δ の正規分布に従うと仮定することにした(式(30))。

心理測定関数の推定結果をFigure 5Aに示す。参加者間で共通であると仮定した統制刺激(アイシャドウなし)の主観的等価点 ϕ のEAPは100.6%(95%確信区間は100.0–101.3%)であった。これに対し、錯視量の参加者間平均 μ_δ のEAPは3.8ポイント(95%確信区間は2.1–5.6)であり、アイシャドウによる目の過大視錯視が確認された。また、錯視量の個人差の大きさを表す標準偏差 σ_δ のEAPは3.3ポイント(95%確信区間は2.2–4.9)であった。Figure 5Bは式(30)を使って求めた錯視量の個人差の事後予測分布である(点とエラーバーは参加者間平均とその95%確信区間)。ここから、例えば錯視量が正の値となる参加者の割合(薄い灰色で塗られた領域の面積)は87.1%である、といった具合に将来の参加者の錯視量に関する予測を立てることができる。

フランカー課題データのモデリング

最後の例として、本節では視空間的注意の研究でよく

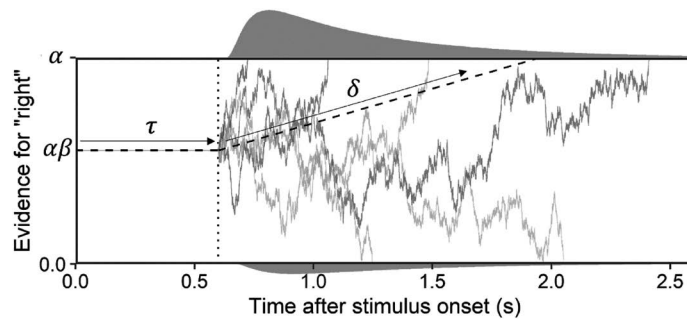


Figure 6. An illustration of evidence accumulation processes assumed by the drift diffusion model.

使われるフランカー課題 (Eriksen & Eriksen, 1974) を使ったモデリングを紹介する。一般的なフランカー課題では、視野の周辺に提示された刺激 (フランカー) を無視しながら中心に提示された刺激 (標的) の弁別を行うことが求められる。例えば標的として中心に提示された矢印が左向きか右向きかを回答する課題の場合、フランカーの情報が標的と競合する不一致条件 (e.g., “←←→←←”) では、フランカーが標的と同じ情報を持つ一致条件 (e.g., “→→→→→”) よりも反応が遅延し誤答が増加することが知られている。反応時間と誤答率は別々のモデルを使って分析されることが多いが (e.g., それぞれに対して分散分析を行う), 本節では両方を同時に分析することができる拡散過程モデルを使ってフランカー課題データの分析を試みる。

この目的のために、本節では矢印を刺激として用いた Kobayashi, Muto, Shimizu, & Ogawa (in prep.) のフランカー課題のデータ (flanker.csv) を分析に使用する。元の実験では刺激間の距離が様々な操作されたが、ここでは刺激間距離が最も短い条件で行われた 2,163 試行 ($N=28$) 分のデータを分析対象とする。反応時間が極端に短い試行はあらかじめ除外しておいたが、誤答試行は分析に使用するため除外していない。

拡散過程モデルと Wiener 分布

拡散過程モデル (Ratcliff, 1978) では、刺激提示の直後から反応のための証拠の蓄積が開始し、その証拠の量が閾値を超えた時点で反応が生じることが仮定される。このモデルを図で説明したものが Figure 6 である。横軸は刺激提示からの経過時間、縦軸はその時点での証拠の量を表している。フランカー課題の各試行において、証拠の量が上側の閾値に到達したら「右」、下側の閾値に到達したら「左」が選択され、到達に要した時間が反応時間として観測されると考えよう。縦軸の上端の α は境

界間距離 (boundary separation) と呼ばれる非負のパラメータで、文字通り上側と下側の閾値間の距離を表す。境界間距離 α が大きいほど判断により多くの証拠を必要とすることを意味するので、この α は判断の慎重さを表すパラメータとして解釈される。

刺激提示直後の証拠の量である $\alpha\beta$ は証拠の初期値を表しており、この値の大きさによって、左もしくは右を選択しやすい反応バイアスが表現される。 β は 0-1 の範囲をとり得るパラメータであり、初期値の相対位置を示す (バイアスがなければ $\beta=.5$)。この図では初期値 $\alpha\beta$ が中央よりも少し高い位置にある ($\beta>.5$) ため少しだけ右バイアスがあると解釈できる。

刺激提示直後に描かれた、証拠の量に変化が見られない期間を非決定時間 (non-decision time) と呼び、その長さを τ という非負のパラメータで表す。このパラメータは、刺激の符号化やキー押しのような、証拠の蓄積とは無関係な処理に要する時間として解釈される。キー押しは判断の後で行われるはずなので、刺激提示直後の非決定時間にキー押しが含まれるのはおかしいと思われるかもしれないが、非決定時間はどの区間に挿入されても数学的には等価であるため、説明の都合上この位置に図示されることが多い。非決定時間は証拠の量が変化しない区間の総和であると考えてもよい。

非決定期間が過ぎると証拠の蓄積が開始し、その量が Wiener 過程と呼ばれる確率過程に従って上下に揺らぎながらやがて上側 (α) もしくは下側 (0) の閾値に達する。Figure 6 には例として 6 試行分の揺らぎを示している。この揺らぎは刺激が持つ情報に応じて線形に上昇または下降する傾向を示すと仮定され、その線形の傾き (単位時間あたりの平均的な証拠の蓄積量) をドリフト率 (drift rate) と呼び、 δ というパラメータで表す (Figure 6 の破線の傾き)。例えば提示された標的が “→” のときには $\delta>0$ 、 “←” のときには $\delta<0$ となり、仮に全く無関

係な刺激が提示されたとしたら $\delta=0$ になると期待される。また、刺激が“ $\rightarrow\rightarrow\rightarrow\rightarrow$ ”のときは“ $\leftarrow\leftarrow\leftarrow\leftarrow$ ”のときよりも右反応に関する証拠の蓄積が速いはずなので、 δ はより大きくなると予想される。標的が“ $\rightarrow$ ”のときに証拠の量が上側の閾値に到達して「右」と反応できれば正解であるが、確率的な揺らぎによって下側の閾値に到達し「左」と反応してしまった場合には誤答となる。このような拡散過程モデルによって、Figure 6の上側と下側に示したような確率分布が導かれる。この確率分布は二値反応（あるいは正誤）と反応時間の同時分布であり、これをWiener分布と呼ぶ。別の言い方をすれば、データにWiener分布を当てはめて4つのパラメータ¹²を推定することで、拡散過程モデルに基づいた推論を行うことができる。ただし、拡散過程モデルの仮定に適さない現象のデータに無理やりWiener分布を当てはめてもパラメータの解釈は難しいので注意が必要である（e.g., 武藤, 2020a）。

個人差を無視したモデル

まずは個人差を仮定しない非階層モデルについて解説する。Wiener分布は二値反応とその反応時間が従う同時分布なので、 i 行目の試行の反応時間を Y_i 、二値反応を Z_i と表し、左反応のとき $Z_i=1$ 、右反応のとき $Z_i=2$ とコーディングする。また、 i 行目の試行で中心に提示された標的の向きを c_i 、それ以外のフランカーの向きを f_i と表し、“ $\leftarrow$ ”を1、“ $\rightarrow$ ”を2とコーディングして、

$$Y_i, Z_i \sim \begin{cases} \text{Wiener}(\alpha, \tau, \beta, -\delta_1), & \text{if } c_i = 1, f_i = 1 \\ \text{Wiener}(\alpha, \tau, \beta, -\delta_2), & \text{if } c_i = 1, f_i = 2 \\ \text{Wiener}(\alpha, \tau, \beta, \delta_2), & \text{if } c_i = 2, f_i = 1 \\ \text{Wiener}(\alpha, \tau, \beta, \delta_1), & \text{if } c_i = 2, f_i = 2 \end{cases} \quad (33)$$

とモデル化する。ここで、境界間距離 α 、非決定時間 τ 、反応バイアス β は試行によらず一定であると仮定している。また、一致条件（ $c_i=f_i$ ）のドリフト率を δ_1 、不一致条件（ $c_i \neq f_i$ ）のドリフト率を δ_2 と置いた。ただし、標的が“ $\leftarrow$ ”（ $c_i=1$ ）のときは“ $\rightarrow$ ”（ $c_i=2$ ）のときとは逆側の閾値（i.e., 下側）に証拠が蓄積されると期待されるので、符号を反転して $-\delta_1$ と $-\delta_2$ にしてある。

このモデルをStanで実装するときにはwienerという

関数を使えばよいのであるが、この関数は上側（右反応）の反応時間分布に対応した関数であるため、下側（左反応）の反応時間分布も表せるようにするには少し工夫する必要がある。上側の分布（ $\text{Wiener}_{\text{upper}}$ ）と下側の分布（ $\text{Wiener}_{\text{lower}}$ ）の関係は

$$\text{Wiener}_{\text{lower}}(\alpha, \tau, \beta, \delta) = \text{Wiener}_{\text{upper}}(\alpha, \tau, 1-\beta, -\delta) \quad (34)$$

と表すことができる。この関係を理解するにはFigure 6を上下反転してみるとよい。境界間距離 α と非決定時間 τ の値は図を上下反転しても変化しないが、反応バイアス β は直線 $\beta=.5$ に対して上下に反転した位置に移動するため $1-\beta$ となり、 δ の符号も反転することがわかるだろう。したがって、反応が左（ $Z_i=1$ ）か右（ $Z_i=2$ ）かに応じてwiener関数の引数を書き換えればよい。ただし、if文が多くなるとスクリプトの可読性が悪くなり間違いやすくなる。そこで筆者がお勧めするのは、

$$\text{Wiener}^*(\alpha, \tau, \beta, \delta, z) = \begin{cases} \text{Wiener}_{\text{upper}}(\alpha, \tau, 1-\beta, -\delta), & \text{if } z=1 \\ \text{Wiener}_{\text{upper}}(\alpha, \tau, \beta, \delta), & \text{if } z=2 \end{cases} \quad (35)$$

という新しい関数を自分で定義する方法である。この関数には5番目の引数として二値反応を表す z が追加されており、この値に応じて上側の分布と下側の分布が切り

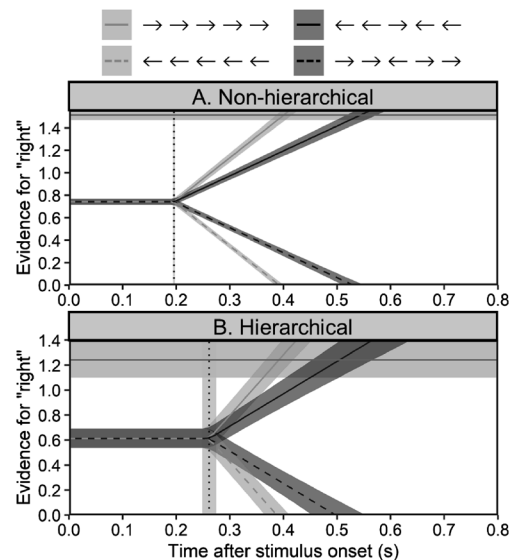


Figure 7. Evidence accumulation process for the flanker task estimated by the non-hierarchical (A) and hierarchical (B) drift diffusion model. Non-decision time and boundary separation are represented by vertical dotted lines and horizontal solid lines, respectively. Error bars show 95% credible intervals.

¹² 他に、 $\beta \cdot \tau \cdot \delta$ の試行間変動を表すパラメータも拡散過程モデルに含まれるが、これらは0に固定されることが多い。また、Wiener過程の揺らぎの大きさ s^2 はスケーリングパラメータとして任意の値に設定することができる（Stanでは1に固定されている）。

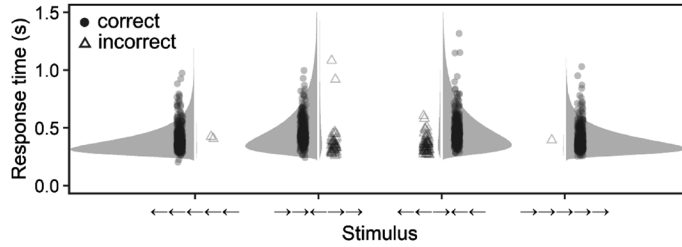


Figure 8. Raw data of the flanker task and posterior predictive distributions calculated using EAP estimates from the non-hierarchical drift diffusion model.

替わるため、 z に Z_i を代入するだけで上側か下側かを気にせずにWiener分布を当てはめることができる。このような自作関数はStanではfunctionsブロックで定義することができる（関数名はwiener2とした）。この関数を使って式(33)のモデルを書き直すと

$$Y_i \sim \begin{cases} \text{Wiener}(\alpha, \tau, \beta, -\delta_i, Z_i), & \text{if } c_i = 1, f_i = 1 \\ \text{Wiener}(\alpha, \tau, \beta, -\delta_i, Z_i), & \text{if } c_i = 1, f_i = 2 \\ \text{Wiener}(\alpha, \tau, \beta, \delta_i, Z_i), & \text{if } c_i = 2, f_i = 1 \\ \text{Wiener}(\alpha, \tau, \beta, \delta_i, Z_i), & \text{if } c_i = 2, f_i = 2 \end{cases} \quad (36)$$

と表すことができる（fl1_ddm_nonhier.stan）。この式(36)は式(33)と完全に等価であるが、Stanで実装するときにはこのような工夫が必要になる。

もう1つ注意すべき点としてはパラメータの範囲が挙げられる。 α と τ は非負の実数、 β は(0,1)の範囲の実数であるが、 τ にはもう1つの制約がある。それは、反応時間の観測値が非決定時間 τ を下回ることが絶対にないという制約である（Figure 6参照）。そこで、観測された反応時間の最小値（今回のデータでは202 ms）を τ の上限として指定する。これをしておかないとエラーが出て推定に失敗してしまうことが多いので注意しよう。

Stanによる推定結果に基づいてプロットした証拠の蓄積過程をFigure 7Aに示す。反応バイアス β のEAPは.490（95%確信区間は.480–.501）であり、左もしくは右への反応の偏りは見られなかった。境界間距離 α のEAPは1.51（95%確信区間は1.47–1.56）、非決定時間 τ のEAPは196 ms（95%確信区間は194–197）であった。ドリフト率のEAPは一致条件で3.78（95%確信区間は3.61–3.95）、不一致条件で2.22（95%確信区間は2.09–2.35）であり、一致条件のほうが不一致条件よりも証拠の蓄積が速いことが確認された。Figure 8には各パラメータのEAPを使って計算した予測分布¹³（灰色の分布）を各条件のロー

データ（円は正答試行、三角形は誤答試行）に重ねて示している。一致条件（左端と右端）では正答率が100%に近く、事後予測分布の誤答側はほとんど見えない。一方、誤答が目立つ不一致条件（中央の2条件）では事後予測分布も両側に広がっている。また、反応時間分布の特徴である正の歪みもきちんと表されている。このように、拡散過程モデル（Wiener分布）を用いることで二値反応とその反応時間を同時に説明・予測することができる。

個人差を考慮したモデル

前項のモデルを階層モデルに拡張してみよう。今回は簡単のためにパラメータ間の相関は仮定せず、5つすべてのパラメータの個人差に正規分布を仮定し、

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \quad (37)$$

$$\tau_j \sim \text{Normal}(\mu_\tau, \sigma_\tau) \quad (38)$$

$$\beta_j \sim \text{Normal}(\mu_\beta, \sigma_\beta) \quad (39)$$

$$\delta_{j,1} \sim \text{Normal}(\mu_{\delta_1}, \sigma_{\delta_1}) \quad (40)$$

$$\delta_{j,2} \sim \text{Normal}(\mu_{\delta_2}, \sigma_{\delta_2}) \quad (41)$$

$$Y_i, Z_i \sim \begin{cases} \text{Wiener}(\alpha_{j(i)}, \tau_{j(i)}, \beta_{j(i)}, -\delta_{j(i),1}), & \text{if } c_i = 1, f_i = 1 \\ \text{Wiener}(\alpha_{j(i)}, \tau_{j(i)}, \beta_{j(i)}, -\delta_{j(i),2}), & \text{if } c_i = 1, f_i = 2 \\ \text{Wiener}(\alpha_{j(i)}, \tau_{j(i)}, \beta_{j(i)}, \delta_{j(i),2}), & \text{if } c_i = 2, f_i = 1 \\ \text{Wiener}(\alpha_{j(i)}, \tau_{j(i)}, \beta_{j(i)}, \delta_{j(i),1}), & \text{if } c_i = 2, f_i = 2 \end{cases} \quad (42)$$

とモデル化した（fl2_ddm_hier.stan）。前項と同様、実装上は自作関数のWiener*（式(35)；Stanスクリプト上ではwiener2）を使用する。ただし、 α_j と τ_j と β_j は範囲に制約があるため実際には切断正規分布に従う。

¹³ Wiener分布モデルの正確な事後予測分布を計算するのは難しいため、確率モデルに各パラメータのEAP

を代入して得られた確率分布で代用した。渡辺（2012）はこれを平均プラグイン推測と呼んでいる。

前項の非階層モデルでは反応時間の観測値の最小値を τ の上限として採用したが、今回の階層モデルでは参加者の人数分だけ τ_j を推定することになるので、参加者ごとの反応時間の最小値を考慮する必要がある。しかし、Stan では τ_j の範囲を j に応じて個別に指定するのは少し面倒である。そこで今回は、 β_j の初期値が参加者ごとの最短反応時間の半分以下になるようにし、初期値が棄却されないように工夫した (fl2_ddm_hier.R を参照)。また、参加者ごとの最短反応時間のうち最大のもの (387 ms) を β_j の上限として parameters ブロックで宣言した。

このモデルによって推定された、平均的な証拠の蓄積過程を Figure 7B に示している。参加者ごとの最短反応時間を考慮したことによって、非決定時間の参加者間平均 μ_i の推定値は非階層モデルの τ の推定値よりも大きな値となった。Figure 9 には各パラメータの参加者間平均の推定結果（点とエラーバーは EAP と 95% 確信区間を表す）および個人ごとの推定値の事後予測分布を示している。右下のパネルは生成量として計算したフランク干渉の大きさ（不一致条件と一致条件のドリフト率の差分 $\delta_{j,1} - \delta_{j,2}$ ）の結果である。この結果に基づいてフランク干渉の個人差に関する推論や予測を行うこともできる。

おわりに

本稿では、実験心理学者にとってお馴染みの実験パラダイムを例に階層ベイズモデリングの基本的な考え方を解説した。本稿で紹介したモデルは比較的単純なものばかりであったが、領域固有の知識や新しい着想を組み合わせることでより発展的なモデリングへと拡張し、検証

可能な仮説の範囲を広げることができる (Lee, 2011 も参照)。Carlin & Chib (1995) がいみじくも述べているように、どのモデルを用いるのかを制限するのはユーザーの想像力だけなのである。

ベイズ統計モデリングは自由で柔軟な研究を可能にするが、これは研究者の自由度 (researcher degrees of freedom) を高めることと表裏一体であるため、都合のよい結果のみを報告するチェリー・ピッキングの横行や無意味な研究の乱立が懸念されるかもしれない。実際、モデル化の方法によってパラメータの推定結果が変わることは本チュートリアルでも繰り返し確認された。また、パラメータの構造化の仕方によって推定結果が正反対になってしまう場合があることも報告されている (e.g., Muto, in press)。事後分布や事後予測分布はモデルが正しいという仮定の下で得られる確率分布に過ぎないため、パラメータの推定結果に基づく推論を行うためにはその前提となるモデルを適切に選択することが不可欠である。

ではどのようにモデルを選択すればよいのであろうか。大前提としてすべてのモデルは間違っている (“All models are wrong”) ので、我々は研究の目的に照らして最大限合理的にモデルを構築するほかない。この観点から言えば、少なくともそのモデルを選択した理論的・経験的な根拠をきちんと説明できることは必須であろう。どうしてもモデルを一意に決められない場合には、感度分析を行い複数の設定下で結果が一貫すること（あるいは一貫しないこと）を確認する方法もある。WAIC や周辺尤度、事後モデル確率といった評価指標を利用してモデル比較を行うのも 1 つの手であるが（詳しくは浜田・石田・清水 (2019)、岡田 (2018)、渡辺 (2012) 等を参

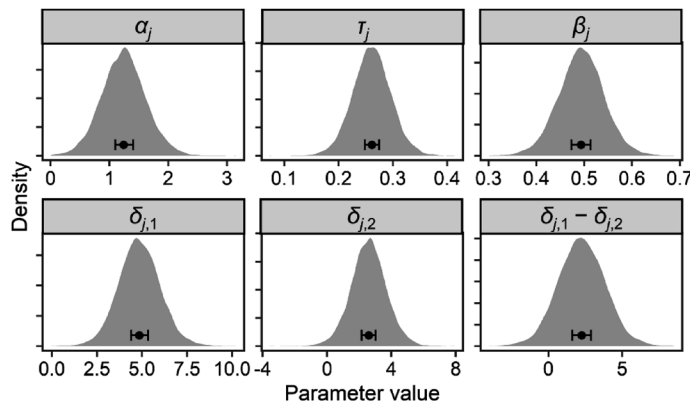


Figure 9. Posterior predictive distributions of the inter-individual variability in parameters estimated by the hierarchical drift diffusion model. Points and error bars show expected a posteriori (EAP) estimates and 95% credible intervals of parameter means.

照), これらの指標は万能でも絶対でもないので過信は禁物である。研究の透明性を高める方法としては, 分析方法の事前登録や(チュートリアルとして, 長谷川他, 印刷中), ローデータとスクリプトをOSFなどのリポジトリに公開して第三者が分析の頑健性や再生性(reproducibility)を検証できるようにすること(e.g., 武藤, 2020c)等が挙げられる。また, モデルを試行錯誤的に修正していった場合にはそのプロセスを論文中で正直に報告することが大事である。もっとも, 伝統的によく使われる t 検定や分散分析等も一種の統計モデルであることを考えると, ここでの議論はベイズ統計モデリングに限った話ではない。どのような研究であれ, 分析の根拠を明確化し透明性を担保することが肝要である。

オープンデータには研究の透明性を高めるだけではなく, 統計モデリングを含めた二次分析を促進する機能もある(武藤, 2020c)。筆者自身もこれまで, 既存のデータを利用したモデリング研究を複数行ってきた(武藤, 2020a, 2020b; Muto, in press; 武藤・水原・入戸野, 2019)。また, 本稿のチュートリアルも全て筆者が関わった研究で収集したデータの二次分析である。最近はオープンデータを推奨あるいは義務付ける学術誌も増えており, データ論文のみを扱う学術誌(e.g., *Journal of Open Psychology Data*)も存在するため, 二次分析研究を行いやすい環境が整いつつある。オープンデータを利用すれば, COVID-19のようなパンデミックによってデータの取得が困難になっても新たな研究を展開できるというのも強みである。たとえ自分自身が二次分析を行うつもりがなくても, 自分のデータを使って他の誰かが面白い研究してくれるかもしれないので, 読者の皆様にもぜひオープンデータを始めてほしい。データを公開するときは, 「データセットの作り方」の節で説明した点に留意して頂けると二次分析者としてはありがたい。オープンデータや二次分析研究に関するより詳細な議論や具体的な実践方法については北條(2020)や国里(2020), 三浦(2018), 武藤(2020c)等を参照されたい。

統計モデリングアプローチにも限界はあるが, 実験計画法に基づく従来の研究法と組み合わせる互いの弱点を補完することで更なる真価を発揮すると考えられる(武藤, 2019; Muto, in press)。従来の実験的研究法は効果の有無や関連の有無を質的に評価するのに適しているが, 関数形の推定などといった量的な説明・予測には不向きである。また, 要因を追加して小さな効果を検証する方向に舵を切りやすく, 結果として再現性の低下を招いてしまうと考えられる。一方, 統計モデリングは刺激と反応の関数関係を数理的に表現するのに向いているが, モ

デルで仮定されたパラメータの妥当性(e.g., 実在性・解釈可能性)は外的な基準なしには確認できず, 重要な変数の存在を見落としてしまうリスクもある(e.g., 疑似相関)。それに, 複雑なモデリングを行うためにはデータが十分な情報を含んでいる必要がある。こういった弱点を克服するためにも統計モデリングと従来の研究法を組み合わせることが重要である。この実例を示す規範的な研究を紹介しよう。White, Ratcliff, & Starns(2011)はフランカー課題遂行中の選択的注意のプロセスを明らかにするために, 拡散過程モデルをベースとした3つのモデルを提案し, データへの当てはまりを検証した。通常のフランカー課題を用いた実験1ではどのモデルもデータを同じくらいよく説明できたため, モデルの優劣を決められなかった。しかし, 反応バイアス(β)や境界間距離(α)などのパラメータを系統的に操作した実験2-5によってモデル間の差異が強調され, 総合的には縮小スポットライトモデルと呼ばれるモデルが最もよくデータを説明できることが示された。このように, 統計モデリングアプローチはこれまでの実験法に対するアンチテーゼではなく, むしろ両者を組み合わせることによって新たな研究の地平が拓かれるのである。

実験心理学における最近のモデリング研究の例としては, 反応時間のばらつきに着目して適合性効果の個人差(Haaf & Rouder, 2017)や文字の心的回転課題の認知方略(Muto, in press)を検証した研究(日本語による紹介として武藤(2021)も参照), 心理物理関数の形状に着目して視覚探索課題における文脈手がかり効果の継時変化を調べた研究(Kobayashi & Ogawa, 2020)等が挙げられる。また, 様々な領域におけるベイズ統計モデリングの実践例を紹介している和書・翻訳書も多数出版されている(e.g., 浜田他, 2019; Lee & Wagenmakers, 2013 井関訳 2017; 豊田, 2017, 2018, 2019)。このような実践例を参考にしながら読者の皆様にもぜひオリジナルなモデリングにチャレンジしてみてほしい。

引用文献

- 馬場真哉(2019). RとStanではじめるベイズ統計モデリングによるデータ分析入門 講談社
- 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. <https://doi.org/10.1111/j.2517-6161.1995.tb02042.x>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143–149. <https://doi.org/10.3758/BF03203267>
- Haaf, J. M., & Rouder, J. N. (2017). Developing constraint in

- Bayesian mixed models. *Psychological Methods*, 22, 779–798. <https://doi.org/10.1037/met0000156>
- 浜田 宏・石田 淳・清水裕士 (2019). 社会科学のためのベイズ統計モデリング 朝倉書店
- 長谷川龍樹・多田奏恵・米満文哉・池田鮎美・山田祐樹・高橋康介・近藤洋史 (印刷中). 実証的研究の事前登録の現状と実践——OSF事前登録チュートリアル—— 心理学研究.
- 北條大樹 (2020). 方法・結果の再現性——手続き・データ・解析共有の意義とその実際—— Retrieved from <https://osf.io/b5wdx/> (January 30, 2021)
- Kobayashi, H., Muto, H., Shimizu, H., & Ogawa, H. (in prep.). Bayesian hierarchical diffusion modeling of the inter-stimulus spacing effect in the Eriksen flanker task.
- Kobayashi, H., & Ogawa, H. (2020). Contextual cueing facilitation arises early in the time course of visual search: An investigation with the speed-accuracy tradeoff task. *Attention, Perception, & Psychophysics*, 82, 2851–2861. <https://doi.org/10.3758/s13414-020-02028-9>
- 国里愛彦 (2020). 2次分析研究 Retrieved from https://kunisatolab.github.io/main/secondary_analysis.html (February 4, 2021)
- Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, 55, 1–7. <https://doi.org/10.1016/j.jmp.2010.08.013>
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge, England: Cambridge University Press. (リー, M. D.・ワーゲンメイカーズ, E.-J. 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
- McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, 28, 295–314. <https://doi.org/10.1007/s10648-014-9287-x>
- 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. https://doi.org/10.24602/sjpr.61.1_3
- 武藤拓之 (2017). Stanでpsychophysics——階層ベイズモデルで恒常法データを分析する—— Retrieved from <https://www.slideshare.net/HiroyukiMuto/stanpsychophysics> (January 30, 2021)
- 武藤拓之 (2019). この人をたずねて (関西学院大学社会学部教授 清水裕士氏) 心理学ワールド, 86, 36–37.
- 武藤拓之 (2020a). 心的回転に対するdiffusion modelingの有効性の検証 日本行動計量学会第48回大会抄録集, 235.
- 武藤拓之 (2020b). Spatial Orientation Testの得点化方法の改善——フォン・ミーゼス分布による角度データのモデリング—— 日本心理学会第84回大会発表抄録集, 115.
- 武藤拓之 (2020c). まずはオープンデータから！——高めよう信用性, 広めよう二次分析—— Retrieved from <https://pt.slideshare.net/HiroyukiMuto/at-84> (January 30, 2021)
- 武藤拓之 (2021). ベイズ統計モデリングの有用性を示す認知心理学研究の紹介——個人間・試行間のばらつきを理解する—— 認知科学, 28, 182–187. <https://doi.org/10.11225/cs.2020.077>
- Muto, H. (in press). Evidence for mixed processes in normal/mirror discrimination of rotated letters: A Bayesian model comparison between single- and mixed-distribution models. *Japanese Psychological Research*. Advanced online publication. <https://doi.org/10.1111/jpr.12306>
- Muto, H., Ide, M., Tomita, A., & Morikawa, K. (2019). Viewpoint invariance of eye size illusion caused by eyeshadow. *Frontiers in Psychology*, 10, Article 1510. <https://doi.org/10.3389/fpsyg.2019.01510>
- 武藤拓之・水原啓太・入戸野 宏 (2019). 自由選択課題におけるポストディクション現象の認知モデル——ベイズ統計モデリングによる検証—— 日本認知心理学会第17回大会発表論文集, O1–O2. https://doi.org/10.14875/cogpsy.2019.0_2
- Muto, H., & Nagai, M. (2020). Mental rotation of cubes with a snake face: The role of the human-body analogy revisited. *Visual Cognition*, 28, 106–111. <https://doi.org/10.1080/13506285.2020.1727598>
- 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. https://doi.org/10.24602/sjpr.61.1_101
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108. <https://doi.org/10.1037/0033-295X.85.2.59>
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703. <https://doi.org/10.1126/science.171.3972.701>
- 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. https://doi.org/10.24602/sjpr.61.1_22
- Stan Development Team (2021). RStan: The R interface to Stan. Retrieved from <http://mc-stan.org/> (January 30, 2021)
- 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
- 豊田秀樹 (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
- 豊田秀樹 (2019). たのしいベイズモデリング2——事例で拓く研究のフロンティア—— 北大路書房
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1, Article 1. <https://doi.org/10.1038/s43586-020-00001-2>
- 渡辺澄夫 (2012). ベイズ統計の理論と方法 コロナ社
- White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, 63, 210–238. <https://doi.org/10.1016/j.cogpsych.2011.08.001>
- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59 (10), 1–23. <https://doi.org/10.18637/jss.v059.i10>