

# 宇佐美慧（印刷中）個人内関係の推測と統計モデル—ランダム切片交差遅延パネルモデルを巡って— 発達心理学研究

報告者：宇佐美慧

縦断的な調査・観察を通して測定された多変数間の関係を推測する目的で、SEMに基づく交差遅延パネルモデル(CLPM)の適用が心理学を中心に広くなされてきた。一方で、個人の異質性を統制して個人内の変化のプロセスである個人内関係を推測することへの関心が近年高まっており、これは縦断データに基づく因果推論の問題とも密接に関わる事柄である。個人内関係の推測上の限界から CLPM の利用を批判した Hamaker, Kuiper, & Grasman (2015)以後、時間的に安定した個人差としての特性因子(stable trait factor)を組み込んだランダム切片交差遅延パネルモデル(RI-CLPM)が、心理学領域を中心に急速に普及している。

個人内関係の推測のために利用可能な統計モデルは RI-CLPM 以外にも多くあるが、その違いは必ずしも明確に区別されておらず、また個々の統計モデルの解釈やその選択を巡る議論は現在でも活発になされている。実際のデータ発生プロセスや真のモデルが一般に研究者にとって不明な中で、統計モデル内の共通因子や個人効果（異質性）が適切に表現されているかどうかの判断は難しい。本論文では、個人内関係を推測するための一つの有効な方法として RI-CLPM を位置付けてその適用の実際にについて解説し、そして他の統計モデルとの数理的・概念的関係性およびそれらの個人内関係の推測において生じうる問題点について整理した。具体的には、過剰調整の問題、個人内関係と変化の軌跡（平均構造）に関する推測を同時にを行うことの問題、係数の解釈の複雑性の観点から、既存の統計モデル (LCM-SR, LCS, ALT, GCLM) の適用上で生じうる問題点を整理した。

RI-CLPM の特性因子は個人内変動とは無相関な量と仮定している点に異質性を表現する上での特徴があり、またそれが他の有力な統計モデルの選択肢としての動学的パネルモデル(DPM)との関係性を繋ぐ接点にもなる。想定されるデータ発生プロセスに基づき RI-CLPM や DPM 等の特定の統計モデルを支持する立場に立って推定を行いながらも、あり得る他の候補のモデルについても感度分析として適宜推定を行い、適合度指標や情報量規準の結果も参考しながらモデル選択の適切性や推定結果の頑健性について検討していくことが 1 つの現実的な手段と思われる。様々なデータ発生プロセスを踏まえた推定結果の比較検討や、また（特に、時変的な）交絡変数への対処という観点からは疫学・計量経済学・統計的因果推論等の他領域で培われた方法論も踏まえた個人内関係の推測法の開発も、モデル選択を巡る論点整理と並んで今後の重要な研究課題と言える。

Kuhfeld, M., & Soland, J. (2022). Avoiding bias from sum scores in growth estimates: An examination of IRT-based approaches to scoring longitudinal survey responses. *Psychological Methods*, 27(2), 234–260. <https://doi.org/10.1037/met0000367>

報告者：佐宗 駿

発達心理学の分野では、自己効力感や動機づけなどの構成概念に関する測定値の縦断的な発達や変化への関心が高い。例えば複数時点で同一内容の自己報告型質問紙調査を行い、関心のある構成概念に関して総合得点や項目反応理論(IRT)を用いて得点化した上で、潜在成長モデルなどの統計モデルを活用してこれらの発達や変化が議論されることがある。

このような営みが行われる一方、総合得点に基づく得点化と IRT に基づく得点化のいずれが、より正確に真の発達(e.g., 関心のある slope の平均や分散などのパラメタ)を推定可能なのか、また、各得点化がどの程度正確にパラメタを推定できるのかは十分に検討されていない。そこで本論文では、特に、IRT による得点化における項目のキャリブレーションの際に用いるデータ収集デザイン(i.e., 時点数×学年数×縦断・横断)及びモデルの種類(e.g., 一次元 IRT, 多次元 IRT )の違いに注目し、総合得点による得点化を含めた比較検討を行った。

この目的を達成するために、本研究では、3つのシミュレーション研究及び、実データ解析を行った。現実の応用場面を想定して、シミュレーション1では、時点数が3で、母集団における真の軌跡が線形である状況を、シミュレーション2では時点数が4で、軌跡が非線形である状況を想定し、パラメタリカバリ性能を比較検討した。シミュレーション3では、発達に及ぼす介入の効果に関心がある場面を想定し、真の slope parameter の異なる2群(i.e., 介入群と統制群)を設定し、両群における発達の差異に対するパラメタリカバリ性能を検討した。

実データ解析では、ある単一のコホートを対象とし、小学5年生から中学2年生の4時点での成長マインドセットに対する質問紙調査(5項目)への回答データを取り上げ、パラメタの推定結果の差異を検討した。

以上の検討の結果、時点間の共分散を考慮できる多次元IRT(MIRT)を用いた場合、総合得点やその他のIRTに基づく方法と比べて、より正確に真のslope parameterをリカバリできた。一方で、総合得点に基づく方法は、いずれのシミュレーション条件においても、真のslope parameterの平均と分散を過小推定する傾向にあった。

もちろん、MIRTは発達や変化の測定における”万能薬”とまではいかないが、対象とする時点全ての縦断データが得られる場合には、好ましい方法となり得るだろう。

Smithson, M. & Broome, S. B. (2022). Compositional data Analysis Tutorial.  
*Psychological methods*, Advance online publication.

報告者：若井大成

本論文は、各個人毎のデータが全て正であり、かつその合計値が定数になるデータである compositional data の分析法について概説したものである。Compositional data の代表例としては、割合や確率などのデータが挙げられる。データそのものはユークリッド空間において単体(simplex)上にあるものと捉えられるが、この制約は多変量解析の実行上不都合であるため、データを变形して扱いやすくする必要がある。本論文では、2種類のデータ变形方法を導入した後、0というデータに対する対処法をまとめ、実データ分析で手法の応用例を示している。

データ变形方法の1つめは対数比(log-ratio)変換である。これは、 $i$ 人目の compositional data のある  $j, k$  番目の変数（要素またはカテゴリ）を  $y_{ji}, y_{ki}$  とすると、 $\log\left(\frac{y_{ji}}{y_{ki}}\right)$  として表されるものである。この変換をすると、データは範囲無制限のユークリッド空間へと写像されるため、線形モデルをはじめとする慣習的な分析法で扱うことが可能となる。ただし、対数比変換後のデータの共分散が分散のみから求まってしまうといった特殊性を有していることには注意が必要である。対数比変換の中でも様々な種類の変換がある。基準となるような変数がある場合には、その変数の値との対数比 (additive log-ratio; ALR) を取る変換は解釈性に優れる。一方、そのような変数がない場合には、対数変換したデータを中心化する変換(centered log-ratio; CLR)や、ユークリッド空間の正規直交基底を作る変換(isometric log-ratio; ILR)などが用いられる。

データ变形方法の2つめは確率比(probability-ratio)変換である。これは  $\frac{y_j}{y_j+y_k}$  として表されるもので、対数比変換と完全に対応した変換である。対数比変換と比較すると、データを対数変換せずに扱うため解釈性に優れている、分散の不均一性をモデル化しやすいといった利点がある。一方、確率比間の持つ従属関係をモデル化する手法が未開発であるという現状がある。

上記2つの変換は、ともにデータが0を含まないことを前提としているため、0を含むデータに対しては特別な対応が必要である。一部のカテゴリに0が偏在している場合には、複数のカテゴリを結合(amalgamation)することで簡単に対処できるが、これで完全に対処できる場面は限られる。0というデータには、まさに0が真値だと考えられる本質的な0データと、真値は0ではないものの、測定の段階で0とされた丸め込みの0データがあると考えると、それに対する対処法は異なってくる。後者の0データに対しては、全データに同じ変換を施して0を非0にする再尺度化(rescaling)や、0を小さな正の数に置換する方法がある。前者の0データに対しては、0と非0のデータの生成プロセスが質的に異なるとするハードルモデルによる対処法がある。

Urban, C. J., & Bauer, D. J. (2021). A deep learning algorithm for high-dimensional exploratory item factor analysis. *psychometrika*, 86(1), 1-29.

報告者：土方啓一郎

項目因子分析は因子分析モデルにより多変量カategoriカルデータを解析するための方法である。項目因子分析のパラメータ推定はこれまで多くの場合、周辺最尤推定により行われてきた。これはサンプルサイズが無限大に近づくとき、周辺最尤推定のもたらすパラメータの推定量が、一致性・漸近正規性・漸近有効性という統計学的に良いとされる性質を持つためである。周辺最尤推定では、周辺尤度という関数を最大化するパラメータの値を推定値とするが、この周辺尤度の導出には多重積分が伴うため、サンプルサイズや因子数が大きいと計算量が爆発的に増大し、推定が困難になる。この計算負荷の問題に対してはこれまで、Metropolis-Hastings Robbins-Monro (MH-RM)法や変分推定などの計算コストを抑えた近似手法が開発されてきたが、それらの方法もサンプルサイズや因子数が非常に大きい時には計算負荷の問題を克服しきれない。そこで Urban and Bauer (2021)により高速に推定が可能な方法として、深層学習に基づいた変分推定アルゴリズムが提案された。

変分推定は、変分分布を用いて直接推定することの難しい事後分布を近似する方法であり、同論文では特に amortized variational inference (AVI) と importance-weighted variational inference (IW-VI) という二つの変分推定に関するテクニックを用いてより高速な推定と近似精度の高い変分分布を実現している。AVI は、変分分布のパラメータである変分パラメータに関して、サンプルごとに変分パラメータを推定するのではなく、観測データを引数として変分パラメータを出力するようなパラメトリックな関数(Urban & Bauer (2021)ではニューラルネット)の存在を仮定し、かつその関数がサンプル間で共通であると考えることで、サンプルごとの変分パラメータの推定を、関数を構成する有限個のパラメータの推定に落とし込む手法である。これにより変分パラメータの推定では、サンプルサイズが増大しても、関数を構成する有限個のパラメータを推定すれば良いことになり、効率的に計算することが可能となる。また IW-VI は、importance sampling というサンプリングにより、伝統的な変分推定よりも高い精度で周辺最尤推定量の近似を可能にする方法であり、この AVI と IW-VI に基づいた提案アルゴリズムを MH-RM 法などの既存のアルゴリズムと計算速度や推定バイアスという観点から比較するシミュレーションを行ったところ、サンプルサイズや因子数の多い場合でも既存手法より高速に推定が可能であり、かつ周辺最尤推定量の持つ一致性に近いような推定バイアスの結果が得られたと報告されている。提案アルゴリズムの限界や今後の発展としては、変分推定量の性質が理論的に未解明であること、因子数などのハイパーパラメータのチューニング方法の開発などが挙げられていた。

Van Geen, C. & Gerraty, RT. (2021). Hierarchical Bayesian models of reinforcement learning: Introduction and comparison to alternative methods. *Journal of Mathematical Psychology*.

報告者：森隆太郎

本論文は、強化学習のモデルにおいて、予測性能の高いパラメタを推定する方法について解説したチュートリアル論文である。強化学習は、環境との相互作用の中で大きな報酬につながる行動方策を学習する過程を定式化した、代表的なフレームワークである。強化学習のモデルは、学習率や逆温度などの基本的なパラメタを持ち、それらを行動データから推定することができる。推定されたパラメタは、個人差を縮約したものと見なすことができ、心理学や神経科学では、しばしば個人ごとの推定値に基づいて個人差に関する推論を行う。

個人差について意味のある推論を行うためには、サンプリングノイズではなくデータにある真の構造をよく抜き出したパラメタを推定する必要があり、これは新しいデータに対する予測精度 (out-of-sample predictive accuracy) から測ることができる。しかし、予測性能の高いパラメタを推定することは、一般に難しい問題である。

本論文は、階層ベイズモデルを用いて、うまく正則化されたパラメタ推定を行う方法を紹介する。具体的には、①興味のあるパラメタについてグループレベルの事前分布を仮定し、②それを規定するハイパーパラメタにさらに事前分布を置きつつ、③これら二層の事後分布を同時にベイズ推定する方法である。グループレベルの事前分布を課すことで、個別の不確かな推定値が集団平均に近づく shrinkage が起こることになる。この階層ベイズモデルによる推定方法の優位性を、二碗バンディット課題における Q 学習モデルのパラメタ推定を題材に、三つの代替方法と比較しながら実証する。

はじめ二つの代替案は、個人差を対照的に極端な形で扱う。第一の方法は、個人差を仮定せず、グループ全体で一つの値を推定する方法である。第二の方法は、グループレベルの制約を設けず、個人ごとのデータから個別にパラメタを推定する方法である。先行研究 (Gershman, 2015) の実データを再解析した結果から、これら二つの代替案によって推定されたパラメタより、階層ベイズモデルによって推定されたパラメタの方が、予測精度が有意に高いことを確認できる。これは、事前分布が、個人差の推定に対するグループレベルの制約をちょうど良い強さで実現していることを示唆している。

三つ目の代替案は、データセットの一部を事前分布の推定だけに使う方法である。具体的には、一部のデータからパラメタを個人ごとに推定して分布を算出し、それをグループレベルの事前分布として用いて残りのデータの個人パラメタを推定する。しかし、この方法はデータの効率性が悪く、さらに実データの再解析から、予測精度の面でも階層ベイズを用いた方法にわずかに劣ることが確認できる。

Heck et al. (2022) A Review of Applications of the Bayes Factor in Psychological Research. *Psychological methods*, 10.1037/met0000454. Advance online publication.

報告者：古川 結唯

本論文では仮説検定やモデル選択の手段として近年関心が高まっているベイズファクター (Bayes Factor: BF)について、心理学研究における応用に焦点を当て、計 6 つのトピック (帰無仮説検定、区間仮説、情報仮説、Bayesian Evidence Synthesis (BES)、ベイズ的変数選択とベイズ的モデル平均化、認知モデルの評価) を概観し、利点と注意点について概説する。各トピックはそれぞれ分析事例を通して具体的な手続きの説明がなされ、BF を計算するための R パッケージ等のツールが紹介されている。なお各トピックで取り上げる事例の分析コードは OSF にて参照できる (<https://osf.io/k9c5q/>)。

前半 3 トピックは仮説検定を扱う。ベイズ的な帰無仮説検定は、帰無仮説を表現する統計モデルと、対立仮説を表現する統計モデルを BF によって比較することで行われる。頻度論的な有意性検定では帰無仮説を棄却することはできても積極的に支持することはできないのに対し、BF を用いることで『データが特定の対立仮説と比較し帰無仮説をどの程度支持するか』を定量的に検討できる。研究仮説に応じた帰無仮説および対立仮説の設定と、適切な事前分布の設定により、標準的な点帰無仮説 (例えば  $H_0: \mu_1 - \mu_2 = 0$ ) の検定や、優越性デザイン、非劣性デザイン、同等性デザインといった区間仮説検定を可能にする。また、複数のパラメータの大小関係を仮説に含める情報仮説 (Informative Hypotheses) について概説し、心理学研究において帰無仮説が研究者の研究疑問 (リサーチ・クエスチョン) を正しく反映しない場面が多く見受けられるという問題に対処する方法を提示している。

4 番目のトピックは複数の研究結果の統合を行う BES の枠組みである。BES は、同じ理論的概念を扱っていれば、研究デザインの異なる多様な研究結果を統合できる柔軟な枠組みである。事例紹介では BES の手法の一つとして、BF の逐次的更新によって、複数の研究結果を全て統合した BF および事後モデル確率を算出する。

心理学研究ではしばしば、ある結果変数Yをうまく予測できる共変量はなにか、そしてその共変量が結果変数Yにどの程度影響を与えるかについて関心がもたれる。5 番目のトピックではベイズ的変数選択およびベイズ的モデル平均化を扱う。具体的な手順を説明とともに、事前モデル確率の設定方法に関する実践的な議論を展開している。

最後のトピックでは、実際の行動や認知過程等を統計的・数理的に説明する方法の開発と検証を目的とする認知モデリング研究における BF の利用を議論する。ステレオタイプを扱う課題への実験データに対する多項処理木モデル(multinomial process tree model)の構築と BF によるモデル選択について説明している。

Seelig, S. A., Rabe, M. M., Malem-Shnitski, N., Risze, S., Reich, S., & Engbert, R. (2020). Bayesian parameter estimation for the SWIFT model of eye-movement control during reading. *Journal of Mathematical Psychology*, 95, 102313.

報告者：若井大成

本論文は、文章読解中の眼球運動の数理モデルの一種である SWIFT モデルについて、データから一貫した枠組みでパラメタ推定を行うための拡張を提案したものである。読解中の眼球運動についての数理モデルはこれまで数多く提案されてきたが、それらの比較は定性的なものにとどまっていること、実データからのパラメタ推定が分析者が恣意的に設定した目的関数に依存していること、モデルを通してデータの個人差を説明できないなど多くの限界があった。そこで本研究では、SWIFT モデルに尤度を加えることでベイズ的に一貫した枠組みでデータからパラメタを推定できるようにした。

1 章では主に読解中に見られる眼球運動のパターンが概説されている。文章中の単語認識では、単語を視野の中心に入れる必要があるため、高速な視線運動であるサッケード (saccade) が生じる。視線を移動させた後は単語認識のため視線を固定してその単語を固視 (fixation) する。読解では 1 つ 1 つ単語を順に認識しているわけではなく、1 つの単語を二度以上固視したり (refixation)，ある単語を固視せず飛ばしたり (skipping)，飛ばした単語に視線を戻したり (regression) というように複雑な眼球運動が行われている。SWIFT はこれらの眼球運動を一貫した枠組みで説明する点で有用なモデルとされている。

2 章では SWIFT モデルにおける仮定について説明されている。サッケードで視線を移動させる目標となる単語は、単語毎の活性化変数という時変的な潜在変数によって決定される。そのサッケードが起きるタイミングは、サッケードタイマーと 3 つのサッケードプログラムという時変的な潜在変数によって決定される。サッケード後にその単語を処理する速さは、固定した視線の位置と単語内の各文字との距離によって決定される。また、サッケードの到達位置は目標とする単語の中心から系統的および偶発的にずれるとする。

3 章では SWIFT モデルを下に、観測された各固視の位置および固視時間の尤度を定義する。位置の尤度については、目標とする単語を所与としたときの尤度が SWIFT モデルで与えられているが、目標とする単語の選択確率は活性化変数という複雑な潜在変数に依存する。この周辺化は計算コストが大きいため、擬似周辺尤度によって近似する。固視時間の尤度については、サッケードの種類に応じてサッケードタイマー やサッケードプログラムの実行のパターンを数理的に定義した上で、個々のプログラムの実行時間が指數分布に従う事から求まるが、解析的な計算を行う代わりにシミュレーションで疑似データを生成して確率密度を近似する。

4 章では実データ分析の方法と結果が紹介された。尤度ベースの SWIFT モデルで推定したパラメタを所与として生成した疑似データと実データを比較した結果、一部の指標で良い適合が確認され、モデルが読解の個人差を部分的に説明できる可能性が示された。

Zhan, P., & Qiao, X. (2022). Diagnostic Classification Analysis of Problem-Solving Competence using Process Data: An Item Expansion Method. *Psychometrika*, 87, 1529–1547. <https://doi.org/10.1007/s11336-022-09855-9>

報告者：佐宗 駿

CBAs (Computer-Based Assessments)の発展に伴い、解答者の問題解決過程に関するプロセスデータの入手が容易となった。しかし、CBAs で得られるプロセスデータは、多くの場合、解釈・分析上の困難を極める。その要因の一つとして、データの形式が解答者ごとに長さの異なるロング形式であることが挙げられる。

本研究では、この問題に対して、単一の CBA 項目に複数の Phantom Items を設定する項目拡大法(Item Expansion Methods)という手法を提案し、解決を試みた。具体的には、プロセスデータから得られる可能な action sequences の集合から、当該問題の解決に関連する要素を Phantom Items として抽出する。さらに、各 Phantom Items の実行に必要な問題解決スキルを複数のアトリビュートとして設定することで、診断分類モデル(Diagnostic Classification Models)による各スキルの習得状況の推定も可能にする。

従来、問題解決力は、項目反応理論を用いた一次元的な診断が主であり、問題解決力を構成する下位スキルの習得状況の診断は十分になされてこなかった。しかし、項目拡大法とともに、診断分類モデル、とりわけ高次因子モデル(e.g., HO-GDINA モデル)を適用することで、問題解決力を潜在連続変数として推定可能になるだけでなく、その下位スキルの習得状況についても潜在離散変数として推定が可能となる。

CBAs によって問題解決力を測定した PISA2012 の課題 1 項目に対する本手法の応用例が提示された。分析の結果、推定されたアトリビュート習得パターンから、問題解決に関わる 5 つの下位スキルそれぞれの習得・未習得状況を明らかにすることができた。この結果は、従来の素点カテゴリ(i.e., 0 点, 1 点, 2 点)と比べて、より詳細な診断情報を得られるという提案手法の有用性を示すものである。

本研究は、Practitioner-friendly な手法の提案を目指したものもある。そのため、本研究の分析のほぼ全てが統計ソフトウェア R の GDINA package により、容易に実行可能であるという魅力がある。この点に鑑みても、プロセスデータ解析の手法として提案してきた従来の手法と一線を画し、実践への応用可能性の高い方法であると考えられる。