

報告番号	※甲	第	号
------	----	---	---

主論文の要旨

論文題目 心理学実験における推定精度の観点からの実験刺激選択法

氏名 藤田 和也

論文内容の要旨

心理学実験において、参加者の特性などのパラメータを推定精度高く測定することは、検定などの統計分析においても重要となる。サンプルサイズや実験刺激などの実験デザインはパラメータの推定精度などを規定する要因となり、かつ実験者側がコントロールできるという意味で着目すべきものである。しかし、とりわけ実験刺激に関しては、推定精度などの観点から統計的に最適な刺激選択を行ったという記述のある研究は少なく、どのように実験デザインを決定したのか不明瞭な場合もある。そこで本研究では、学力測定や質問紙調査に主に利用されてきた、推定精度を最大化するような刺激選択法の一つである Computerized Adaptive Testing (CAT) に着目した。

第一章では、実験刺激が推定等に影響することを説明し、CATに関する研究のレビューや説明を行った。そしてCATを認知実験に適用するメリットや、認知実験における刺激選択 (Adaptive design optimization, ADO) を紹介した。

第二章では、刺激選択手法であるCATおよびADOの統計理論的な説明を行った。最尤推定量の漸近的な性質や、推定量の評価および推定量が従う分布について触れ、CATの理論的な根拠を説明した。また、ADOによる刺激選択法を定式化しCATとの比較も行った。ADOは観測データやパラメータについて積分計算を必要とするが、CATで利用するフィッシャー情報量 (F情報量) は積分計算が不要なため、計算負荷が低い。最後に、CATという手法において、どのようにF情報量を刺激ごとに計算して、刺激が選択されるかを説明した。

第三章 (研究1) ではCATで提案されているF情報量を基準として刺激を選択する手法を、時系列依存を含んだ認知モデルであるQ学習モデルに対して適用した。第一ステップで計算負荷の低いF情報量で膨大な刺激の中から推定精度が高くなりそうな刺激を絞り込み、第二ステップでMarkov Chain Monte Carlo (MCMC) 推定により精密に刺激デザインの良さを評価するというtwo-step procedureを提案した。F情報量で予測された刺激間での推定精度の大小関係は、MCMCベースでのシミュレ

ーションにおける刺激間での推定精度の大小関係と一致しており、two-step procedure が有効に機能することを支持する結果が得られた。また、標準的な Q 学習モデルのパラメータは 2 つあるが、それぞれのパラメータで適切な刺激は異なることが分かった。これらの結果から、研究者は認知モデリングを適用する実験を行う際には、自らの興味や研究手法 (e. g., どちらのパラメータに関して検定するか) を事前に明確化し、適切な刺激デザインを利用することが望ましいといえる。さらに、刺激により推定精度を向上させると相関係数のバイアスが抑えられることや、Q 学習モデルにおける CAT の効果も一定程度確認された。

第四章 (研究 2) では、単一のモデルを仮定して刺激選択を行う従来の CAT を、複数のモデルを同時に考慮して刺激選択を行える手法に拡張した。この提案手法ではモデルの不確実性を考慮するために Model averaging (MA) という手法と CAT を組み合わせた。シミュレーション実験の結果、誤ったモデルを仮定した CAT は推定精度の向上が、真のモデルを仮定した CAT より悪かった。これは推定時のモデル誤指定ではなく、刺激選択時のモデル誤指定の問題があることを示している。また提案手法は、真のモデルを仮定した CAT と同等のパフォーマンスを示した。実際の実験場面では真のモデルは基本的に未知であるため、単一のモデルを仮定する従来の CAT は真のモデルを仮定した CAT となっているか、誤ったモデルを仮定した CAT となっているかは不明である。提案手法はモデル仮定に関して頑健に刺激選択の効果が得られる手法であるという点で望ましい。また、データが蓄積するほどモデルの不確実性は小さくなり、提案手法は真のモデルを仮定した CAT に近づいていくと考えられる点でも、提案手法は合理的な手法である。

第五章 (研究 3) では、階層モデルにより集団平均などの集団レベルパラメータを推定する時に、従来の CAT が及ぼす影響を検討した。階層モデルは、参加者が複数人いて、かつ各参加者は複数試行を行うような反復測定状況において用いられるモデルである。(認知) 心理実験の多くは反復測定状況であり、また集団平均などの集団レベルパラメータに興味があることが多いため、階層モデルを適用すべき状況は多いと考えられる。従来の CAT は個人ごとに分析するためのモデル (個人レベルモデル) を仮定しており、反復測定状況で得られたネストデータに適用する階層モデルを仮定してはいないという意味でモデル誤指定の問題が存在する。一方で、個人レベルモデルにおける個人レベルパラメータの推定精度が向上すれば、その集団平均など集団レベルパラメータの推定精度も向上すると考えることもできる。実際、シミュレーション実験の結果、従来の個人レベルモデルを仮定した CAT (個人レベル CAT) を行うと、階層モデルにおける集団レベルパラメータの推定精度も向上することが明らかとなった。

第六章 (研究 4) では、研究 3 を踏まえ、薬学領域で提案されている集団レベルパラメータの推定精度を直接最大化する刺激選択法 (集団レベル CAT) を、認知モデルに適用した。さらに、従来の個人レベルモデルを仮定する CAT が集団レベル CAT と比べて、推定精度の意味で損失がどれくらいあるかを確認した。その結果、集団レベル CAT は、集団レベルパラメータの推定精度を最も良くすることが経験的に分かった。また個人レベルモデルを仮定する従来の CAT は、理論的に最適な集団レベル CAT に比べても損失がほとんどないことが分かった。これは研究 3 の結果を補強する結果であり、反復測定状況においても、より実装のコストが低い個人レベル CAT で代用できる可能性が示された。

第七章では、総合考察を行った。本博士論文の研究をまとめ、刺激選択の必要性やどのような実験に適用できるかを議論した。さらに、発展的な話題として、モデル誤指定が推定精度に及ぼす影響や Model averaging に使用するモデルセットの評価方法、CAT におけるモデル誤指定がモデル選択に及ぼす影響についても議論し、推定とモデル選択や検定の関連についても議論した。

付録 A では、実際に F 情報量を計算し、CAT を適用するために必要な情報をまとめた。前半では、ランダムな刺激選択、選択確率を 0.5 に近づける刺激選択、F 情報量を最大にする刺激選択ごとに推定精度などがどの程度変わるかを確認した。後半では、遅延価値割引モデルと呼ばれるモデルを例に、F 情報量を刺激ごとに計算して刺激選択を行う R コードの例を示した。

本研究では、複雑な認知モデルにおいても CAT が適用可能であり、また推定精度を効率的に向上させることを確認した。また、モデル誤指定に着目し、モデル誤指定が問題にならない状況や、問題になりうる際に刺激選択法の効果を保つ手法を提案した。今後は、階層モデルにおける刺激選択法および検出力やモデル選択の観点からの刺激選択法の精緻化、および推定精度とモデル選択の関連などに関する更なる研究を行うことが必要だろう。