

## 論文の内容の要旨

論文題目 項目反応間の局所依存性が受験者、項目およびテストの諸特性の推定に与える影響の検討

氏名 登 藤 直 弥

ある能力を測るためのテストをある受験者集団に実施するとき、そのテストに含まれるどの項目においても、能力水準が高くなるほどその項目への正答率は高くなっていくと考えられる。項目反応理論では、このような能力と各項目の正答率との関係を利用して、各項目の能力水準毎の正答確率を項目反応モデルと呼ばれる関数で表現し、この項目反応モデルと項目反応データから、項目反応モデルに含まれる項目の特性を表す母数（識別力、困難度）と受験者の特性（今の場合は能力）を表す母数を推定することになる。

項目反応理論を用いてテストの作成や運用を行うと、テスト特性関数と呼ばれる受験者特性の関数を利用して、テストを実施する前に、異なるテストの間でその得点を対応づけることが可能となるなど、テストの作成、評価、運用に資する様々な利点を有するため、近年では、日本においても、項目反応理論に基づき作成され、運用されるテストが増えてきている。また、日本で実施されるテストについて考えてみると、その出題形式の特徴として大問形式の出題が多いことが挙げられ、これらのことから、今後日本において大問形式の問題を含むテストを項目反応理論に基づき作成、評価、運用する機会が増加していくであろうことが予想される。

通常、項目反応データを項目反応理論に基づいて分析する際には、分析対象のデータに対し項目反応間の局所独立性が仮定される。ここで、項目反応が局所独立であるとは、受験者特性の値を固定した場合に、当該項目間で項目反応が統計的に互いに独立になることを意味しており、例えば、同じ受験者特性値を有する受験者ならば、ある項目に正答した受験者と誤答した受験者との間で、別の項目に正答する確率が変わらないことを意味している。一方で、大問形式の問題には、ある項目への回答を利用して他の項目への回答を導き出す項目やある文章や図表に対

する理解を問う項目群などが含まれている。前者の項目群においては、例えば、問 1 を誤答してしまった受験者は高い確率で問 2 にも誤答するといった関係が見られ、後者の項目群においては、ある文章や図表をたまたま正しく読み取ることができた（できなかった）受験者は、その文章や図表の内容理解について問う項目群の全てに正答（誤答）する確率が高くなると考えられる。つまり、大問形式の問題内においては、テストが測定している受験者の特性値が同じ値の受験者であっても、ある項目に正答した受験者は、その項目に誤答した受験者よりも、他の項目に正答する確率が高くなるという状況が生じると考えられ、項目反応間の局所独立性は満たされなくなると考えられる（項目反応間の局所依存性）。

先行研究では、この局所依存性を考慮したモデルとして様々なものが提案されているのだが、これらのモデルに関しては、モデルに関する理解や普及が局所独立性を仮定するモデルほど進んでおらず、現時点では、局所依存性が生じていると考えられるデータを項目反応理論に基づいて分析する際には、局所独立性を仮定したモデルを利用することが一般的となっている。

以上のようなことから、今後日本において項目反応理論に基づくテストの作成、評価、運用が増加していくであろうことを見据えると、どのような条件の下であれば局所独立性を仮定した受験者特性、項目特性、テストの特性の推定が局所依存性に対して頑健であると言えるのか、詳細に検討することは重要であると考えられる。また、局所依存性を考慮したモデルを使用することによりこれらの特性の推定精度がどの程度改善されるのか検討し、それが局所独立性を仮定するモデルに代えてより複雑な局所依存性を考慮したモデルを使用するに値するほどのものなのか検討することも重要であると考えられる。さらに、もし局所依存性を考慮したモデルが使用するに値するほどのものであるならば、それはどのような条件下のことであり、その条件下においては用いるモデルの種類によって推定精度の改善度合いが異なるのか、詳細に調べていくことも必要であると考えられる。

そこで、本論文では、これらの点に関してシミュレーションを通じた検討を行い、大問形式の問題を含むテストの項目反応理論に基づく作成、評価、運用に資する知見の獲得を目指した。

第 1 章で局所独立性の概念についての整理と局所依存性を考慮した項目反応モデルのレビューを行った後、まず第 2 章において、局所依存性が局所独立性を仮定した受験者特性の推定に対してどのような影響を与えるのか、異なる発生メカニズムを表現する複数のモデルからデータを発生させてシミュレーションを実施し、検討を行った。その結果、発生モデル毎に局所依存性に対する頑健性の評価結果が異なった一方、多くの発生メカニズムに共通する一般的な知見として、受験者数が少ない場合に局所依存性の影響が無視できないほど大きくなる、という知見が得られた。特に、特定のテストを超えて各受験者の受験者特性値を共通の尺度上で評価するという文脈からは、受験者数が少ない場合、局所独立性を仮定したモデルは頑健ではないといえる。また、一般的な知見として、推定量のバイアスに対する局所依存性の影響は、項目が局所依存性を有している場合、項目数の増加に伴う改善が見られないことが明らかとなった。

次に第 3 章において、局所依存性を有するデータに対して局所独立性を仮定するモデルと局所依存性を考慮するモデルをあてはめてシミュレーションを実施し、局所依存性が局所独立性

を仮定した項目特性の推定に対してどのような影響を与えるのか、項目母数の比較可能性を確保した上で検討を行った。その結果、特定のテスト内において項目の特性を評価するという文脈においては、局所依存関係にある項目のテスト全体に占める割合が高く、項目間の局所依存度が項目群間で異なる場合にのみ、局所依存性の影響が無視できないほど大きくなっていった。一方、特定のテストを超えて各項目の特性値を共通の尺度上で比較するという文脈においては、局所依存性の影響が無視できないほど大きくなっており、局所独立性を仮定したモデルは頑健であるとはいえないことがわかった。また、受験者数や局所依存関係にある項目数などの関係については、対象とする項目特性毎に異なる関係が得られたものの、共通する一般的な知見として、局所依存関係にある項目の割合が増加すると、局所独立性を仮定した場合の推定誤差が局所依存性を考慮した場合に比べて大きくなる、という関係が見られた。加えて、局所依存関係にある項目のテスト全体に占める割合が高い場合、項目間の局所依存度が項目群により異なると、局所独立性を仮定した場合の真値と推定値との相関係数が局所依存性を考慮した場合に比べて低下する、という関係が見られた。さらに、項目母数の比較可能性が頑健性の評価にどのような影響を与えるのか検討したところ、比較可能性を確保した場合としない場合とで識別力の推定量にかかるバイアスの符号が逆転するなど、比較可能性を確保するか否かにより頑健性の評価が異なる結果となった。したがって、項目母数の比較可能性を確保していなかった先行研究における結果は、本研究における結果によって見直される必要がある。

さらに第 4 章において、局所依存性が局所独立性を仮定したテスト情報量の推定に対してどのような影響を与えるのか、シミュレーションを通して検討を行った。その結果、誤って局所独立性を仮定したモデルをあてはめた場合、テスト情報量が過大推定されることがわかった。しかし、局所依存性の程度とこの過大推定の程度との関係は単純ではなく、局所依存度の強い項目群が多い場合には、過大推定の程度はむしろ小さくなっていった。

最後に第 5 章において、局所依存性を有するデータに対して局所依存性を考慮する複数のモデルと局所独立性を仮定するモデルをあてはめてシミュレーションを実施し、局所依存性を考慮することにより受験者特性の推定精度がどれほど改善されるのか、検討を行った。その結果、受験者特性の推定精度に関しては分析モデル間でほとんど差が見られず、特定のテストの成績のみを用いて個人差の評価を行うという文脈においても、特定のテストを超えて各受験者の受験者特性値を共通の尺度上で比較するという文脈においても、局所独立性を仮定するモデルに代えてより複雑な局所依存性を考慮するモデルを使用することを積極的に推奨するような結果は得られなかった。また、受験者数や局所依存関係にある項目の数などとの関係については、分析モデル毎に異なる関係が得られたものの、共通する一般的な知見として、受験者数を増加させることにより、局所独立性を仮定した場合のバイアスが局所依存性を考慮した場合のバイアスと同等かそれよりも値が小さくなる、という関係が得られた。

今後は、局所依存性があると想定される大問形式の実際のテストについて、実データに基づいて局所依存度を推定したうえで、本研究のシミュレーションで得られた結果と同様の結果が得られるかの確認を行う必要があると考えられる。また、局所依存性を考慮した現存のモデル

が、受験者特性の推定において必ずしも有効に機能していないことがわかったことから、現実の局所依存のあり方をよりの確に反映した新たなモデルの開発に取り組むことも重要な課題であると考えられる。

(Microsoft Word 2010 にて作成)