

複数代替案の選択データに適用可能な離散選択モデルの開発とその基本特性

愛媛大学 学生会員 ○江田裕貴

愛媛大学 正会員 倉内慎也

1. はじめに

交通需要分析において SP (Stated Preference) データ¹⁾が実務でも用いられるようになって久しい。SP データは、一種の実験データであり、同一個人に対して様々な状況下での選好を繰り返し尋ねることができるため、効用関数の特定に際して多くの情報量を得ることができる。また、その回答形式についても、複数代替案の中から最も望ましい代替案を 1 つだけ選ぶ選択形式のほか、望ましい順に代替案を並べる順位付け形式など、様々なものがあり、その各々に対応した分析モデルも提案されている。本研究では、SP データの回答形式として、「あなたが望ましいと思う選択肢を 2 つ選んでください」のような形式（以下、上位選択形式）に着目する。上位選択形式は、選ばれた複数代替案間の選好順位が不明であるという点で、得られる情報量は順位付け形式よりは少ないが、逆に回答負荷は軽減されるため、信頼性の高い回答が得られる可能性がある。また、選択形式と比較した場合、最も望ましい代替案が不明であるという問題は生ずるものの、それ以外の代替案に対する選好情報が得られるという利点がある。上位選択形式は、これまでもアンケート調査等で頻繁に用いられているが、収集したデータは単に集計されるだけであり、効用関数の特定に活用された事例は皆無である。これは、離散選択モデルが本質的に複数の代替案の中から最も望ましい 1 つの代替案を選ぶことを前提としたモデルであるのに対し、上位選択形式では複数の代替案の選択を許容するため、そのまま適用することができないことによる。そこで本研究では、上位選択形式のデータに対応した分析モデルを構築すると共に、それにより効用関数がどのように推定されるのかについて、選択形式及び順位付け形式との比較分析を行うことを目的とする。

2. 上位選択モデルの定式化

選択形式のデータに対応した離散選択モデルとしては様々なモデルが提案されているが、その代表例は多項ロジットである。また、順位付け形式のデータについては、例えば 2 番目に選ばれた代替案は、1 番目に選ばれた代替案以外の中で当該代替案が最も望ましいと考え、ロジットモデル等の離散選択モデルを繰り返し適用することで順位付け回答の出現確率を表すことができる。代表例は、効用関数の誤差項に I.I.D. ガンベル分布を仮定したランクロジットモデルであり、個人 n が J 個の代替案から望ましさに順に T 番目までの代替案を回答した場合、その回答ベクトル $\mathbf{I}_n (= \{i_{1n}, \dots, i_{tn}, \dots, i_{Tn}\})$ の出現確率は次式のように表すことができる。

$$P(\mathbf{I}_n) = \prod_{t=1}^T \prod_{i \in G_{tn}} \left[\frac{\exp(V_{in})}{\sum_{m \in G_{tn}} \exp(V_m)} \right]^{d_{in}} \quad (1)$$

ここに、 i_{tn} は個人 n が t 番目に望ましいと回答した代替案、 G_{tn} は個人 n が t 番目に望ましい代替案を選ぶときに直面している代替案の集合、 d_{in} は個人 n にとって t 番目に望ましい代替案が i であった場合は 1、そうでなければ 0 となるようなダミー変数、 V_m は代替案 m の効用の確定項である。

次に、上位選択形式のデータに対応したモデル（以下、上位選択モデル）を構築する。簡単な例として、4 つの代替案 (A,B,C,D) の中から 2 つの代替案 (A,B) が望ましいと回答したケースを考えよう。このとき、回答データと合致するような全ての順位付けパターンは、{A,B,C,D}, {A,B,D,C}, {B,A,C,D}, {B,A,D,C} の 4 パターンである。ゆえに、各順位付けパターンをランクロジットモデル等で表現した上で、その周辺和をとることで上位選択形式の回答の出現確率を厳密に定式化することができる。より一般的に、個人 n が J 個の代替案から望ましさに照らし合わせて T 個の代替案を選んだ場合の回答の出現確率は、次式のように与えられる。

$$P(C_{Tn}) = \sum_{G_{Tn} \in R_{Tn}} P(\mathbf{I}_n) \quad (2)$$

ここに、 C_{Tn} ：個人 n によって選ばれた上位 T 個の代替案の組み合わせ、 R_{Tn} ：個人 n が選んだ上位 T 個の代替案の組み合わせと合致するような順位付けパターン G_{Tn} の集合（ただし $G_{Tn} \in R_{Tn}$ ）である。なお、上位選択モデルにおいて、 R_{Tn} の要素数、すなわち順位付けのパターンは $J P_T$ 通りあり、回答者が直面している代替案の数 J や、その中から選ぶべき代替案の数 T が増加すると膨大な数にのぼることが問題点として挙げられる。

3. 上位選択モデルの推定特性

(1) 解の一意性

多項ロジットモデルは、対数尤度関数のヘッセ行列が正定値行列となる、すなわち狭義凹関数になるため²⁾ 効用パラメータが唯一解を持つことが保証されている。ゆえに、多項ロジットモデルの積で表されるランクロジットモデルについても同様のことが言える。しかしながら、ランクロジットモデルの周辺和で表される上位選択モデルでは、対数尤度関数のヘッセ行列が正定値行列とならないため、解の一意性は保証されない（紙面の都合上、証明については文献3）を参照されたい）。よって、モデル推定においては、初期値を色々と変化させた上で、対数尤度が最大となるような解を導き出す必要がある。

(2) シミュレーションデータによる推定特性の検証

離散選択モデルに含まれる効用パラメータの推定には、BFGS 法などの数値解法⁴⁾ が用いられる。ゆえに、上記のように理論的に解の一意性が保証されるようなモデルでも、効用関数の特定化によっては、最適解が得られない場合がある。そこで、シミュレーションにより、人工的な選択データを作成した上で、多項ロジットモデル、ランクロジットモデル、上位選択モデルを適用し、それにより効用関数がどの程度再現できるのかを検証した。シミュレーションデータの作成手順としては、まず、効用パラメータの真値を設定した上で、説明変数ならびに誤差項に乱数を発生させ、各々の代替案の効用値を計算する。次に、算出した効用値が高い順に各代替案を並べて順位付けデータを作成すると共に、そこから最も効用値が高い代替案だけを取り出したものを選択データ、効用値の高い順に T 個取り出したデータを上位選択データとし、3 種類の回答形式のデータを作成する。その上で、効用パラメータを未知として扱い、各回答形式のデータに対応するモデルを適用し、推定されたパラメータと真値との差やモデルの適合度について比較を行った。なお、ここでは、サンプル数を 100、代替案数を 7 として分析した結果を報告する。

図1は、各モデルを適用して推定された効用パラメータを用いて、真の選好順位の出現確率を算出した場合の適合度として AIC 値を示したものである。なお、分析においては、乱数の影響を考慮するために、誤差項に異なる乱数を発生させた 10 個のデータセットを用いてモデル推定を行っている。図より、解の一意性が保証される多項ロジットモデルでも、サンプル数がやや少ない場合には、データセット 5 のように最適解が得られないケースが生ずることがわかる。一方で、上位選択モデルは、理論的には解の一意性が保証されないものの、比較的安定的に推定されており、適合度についても、最も多くの情報量を活用しているランクロジットモデルと比較して遜色ない。他方、上位 6 個の代替案までを尋ねた場合には、逆に適合度が悪くなっている。これは、7 個中 6 個の代替案を選択することは、最も望ましくない代替案を 1 つ選択することと等価であり、その意味でモデルを過度に複雑化しているものと言えよう。今後は実用化に向けて、計算コストの軽減や回答形式による選好順位の信頼性の検討等を行う必要がある。

参考文献

- 1) 森川高行：ステイティッド・プリファレンス・データの交通需要予測モデルへの適用に関する整理と展望，土木学会論文集，No.413/IV-12，pp.9-18，1990。
- 2) McFadden, D.: Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics*, P. Zarembka ed., Academic Press, New York, pp.105-142, 1973.
- 3) 江田裕貴：複数代替案の選択を考慮した離散選択モデルの開発とその推定特性に関する研究，愛媛大学卒業論文，2016。
- 4) Train, K. E.: *Discrete Choice Methods with Simulation*, Cambridge University Press, 2003.

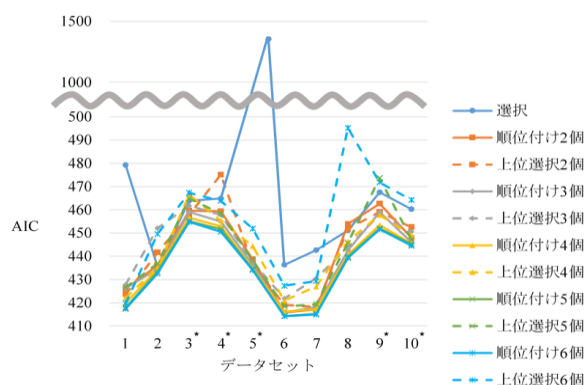


図1 AIC の値