

複数代替案による合成選択肢を含むデータにも適用可能な離散選択モデル

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

交通需要予測に際して、要素選択肢を複数まとめた合成選択肢として分析を行うことが多々ある。ここで、要素選択肢とは、それ以上分解することが不可能な代替案を指し、合成選択肢は、複数の要素選択肢によって構成される代替案を指す。合成選択肢を扱うのは、基本的には分析コストを削減するためであるが、アンケート調査等において要素選択肢レベルでの選択結果が得られている場合と、そうでない場合がある。前者としては、PT 調査データを用いた交通手段選択分析において、「鉄道」や「バス」を「公共交通」という合成選択肢にまとめて集計やモデル化を行うことが代表例として挙げられる。また、後者の例としては、自動車の購買行動調査が挙げられ、被験者が必ずしも保有車両のグレードまで把握しているとは限らないため、車種という合成選択肢での回答を要請したりする。いずれのケースにおいても、集計分析においては特に問題はないが、モデル化のように交通サービスレベルと行動との関係を分析する場合、合成選択肢のサービスレベルが一意に規定できないという問題が生ずる。これに対し実務においては、所要時間等の特定の属性値が最も望ましい要素選択肢を合成選択肢として特定する（以下、「決め打ち」と呼称）ことが頻繁になされる。一般に、IIA が成り立つ多項ロジットモデルを用いてモデル化を行った場合、選択肢をランダムにサンプリングする限りにおいてはパラメータ推定値の一致性は保証される¹⁾。しかし、例えば所要時間に基づいて決め打ちを行った場合は、その時点で所要時間の重要度が極めて大きいとの仮定を暗にしていることになるため、パラメータ推定値にバイアスが生ずるものと考えられる。この問題に対する改善策としては、著者らが提案した複数の代替案の選択を扱うことのできる上位選択モデル²⁾の適用が有効であると思われる。そこで本研究では、合成選択肢を含む状況を想定したシミュレーションデータを用いて、決め打ちを行った上で多項ロジットモデルを適用した場合と、上位選択モデルを適用した場合について、パラメータの推定精度を比較することを目的とする。

2. 上位選択モデル

上位選択モデルは、詳細な選好順位が不明なデータに対し、想定される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により回答の出現確率を定式化したものである。例として、個人 n が J 個の代替案から望ましさを順に T 番目までの回答した場合を考えると、その回答ベクトル $\mathbf{I}_n (= \{i_{1n}, \dots, i_{Tn}\})$ の出現確率は、効用関数の誤差項にガンベル分布を仮定した場合、次式で表すことができる。

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

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_{mn} は個人 n の代替案 m に対する効用の確定項、を表している。

したがって、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、式(1)の周辺和をとることで、次式のように与えられる。

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

ここに、 C_{tn} ：個人 n によって選ばれた上位 T 個の選択肢の組み合わせ、 R_{tn} ：個人 n が選んだ上位 T 個の選択肢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。

このように、上位選択モデルでは、選択代替案が要素選択肢レベルでは不明である場合に対しても、全ての

要素選択肢の効用関数を明示的に考慮した上で、確率論的に厳密にモデル化がなされている点に特長がある。

3. モデル推定特性

シミュレーションにより、合成選択肢を含む人工的な選択データを作成した上で、決め打ちを行った上で多項ロジットモデルを適用した場合（以下、モデル1）と、上位選択モデルを適用した場合（以下、モデル2）の双方について、効用関数の推定精度を検証する。

（1）シミュレーションデータの作成

効用パラメータの真値を設定した上で、2つの説明変数および誤差項に乱数を発生させ、各代替案の効用値を計算する。次に、複数の代替案を1つの代替案としてまとめた合成選択肢を構成した上で、効用値が最も高い選択肢を抽出し選択データを作成する。その後、効用パラメータを未知として扱い、選択データと説明変数を用いて各モデルを推定する。なお、ここでは、サンプル数が300、要素選択肢が4つのケースを報告する。

（2）シミュレーションデータによるモデル推定特性の検証

表1にモデル1・2によるパラメータ推定値を、図1にモデルの適合度を表す指標として、推定パラメータを用いて計算される真の選好順位の出現確率に対応するAIC値を示す。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た。また、データ1とは代替案1・2を合成選択肢、データ2は代替案1・2、代替案3・4をそれぞれ合成選択肢とした上で選択データを作成したものである。ここで、モデル1について、決め打ちを行った説明変数のパラメータは β_1 である。また、データ2に対してモデル2を適用する場合、合成選択肢に含まれる要素選択肢の定数項は分離して推定することはできないため、水色の枠で囲ったように合成選択肢の定数項として推定を行った。

まず、図1より、データ1・2共にモデル2の方が適合度が高くなっていることがわかる。特に合成選択肢を2つ含むデータ2では、その差は顕著である。実際、表1を見ると、モデル1ではほとんどのパラメータが真値から有意に異なった値に過小推定されていることがわかる。これは、ロジットモデルのパラメータ推定値には誤差項の分散が逆数的に影響することを踏まえると、決め打ちにより誤った要素選択肢を合成選択肢として特定してしまった結果、誤差項の分散が大きくなったためであると考えられる。加えて、説明変数間のパラメータの比も真値である1から大きく乖離していることから、決め打ちは、時にモデル適合度を大きく低下させるのみならず、時間価値のような重要な指標を誤って推計する危険性が高いと言える。一方、モデル2では、一部のデータセットではバイアスが見受けられるが、パラメータ推定精度は比較的良好であると言える。これ

表1 パラメータ推定値

		真値	モデル1										モデル2									
			1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
データ1	α_2	1	0.385	-0.323	0.913**	0.552*	-0.123	0.600**	0.373	-0.225	0.594*	0.003	1.293**	1.488**	1.903**	1.282**	1.231**	1.328**	1.660**	1.728**	1.460**	1.636**
	α_3	2	1.212**	0.854**	2.113**	1.343**	0.970**	1.417**	1.252**	1.262**	1.610**	1.172**	1.790**	3.051**	2.302**	2.213**	2.662**	2.093**	2.198**	2.477**	2.491**	2.663**
	α_4	3	1.621**	1.923**	2.508**	2.122**	2.166**	2.030**	1.599**	1.773**	2.514**	1.982**	9.642**	9.834**	9.974**	9.544**	8.831**	8.253**	11.194**	9.291**	8.927**	9.311**
	β_1	10	7.946**	7.353**	9.001**	8.671**	7.797**	7.315**	7.959**	7.493**	8.427**	7.548**	9.361**	10.800**	10.398**	8.141**	9.145**	8.283**	12.255**	9.679**	8.863**	9.743**
	β_2	10	5.95**	6.757**	7.972**	5.772**	6.415**	5.58**	6.589**	6.222**	6.972**	6.006**										
データ2	α_2	1	0.273	-0.120	0.460	0.422	0.002	0.634**	0.425	-0.097	0.586**	-0.091	1.533**	1.800**	2.160**	1.630**	1.818**	1.674**	1.968**	2.123**	1.992**	2.133**
	α_3	2	1.429**	0.947**	1.719**	1.311**	1.493**	1.716**	1.300**	1.459**	2.037**	1.213**	9.35**	7.399**	10.575**	8.492**	7.904**	7.848**	10.810**	9.437**	9.323**	9.303**
	α_4	3	1.093**	1.182**	1.774**	1.606**	1.398**	1.756**	1.126**	1.699**	1.871**	1.551**	9.144**	8.166**	10.128**	7.689**	8.094**	7.853**	12.342**	9.685**	8.553**	9.484**
	β_1	10	7.469**	6.578**	9.064**	8.366**	7.510**	8.405**	7.561**	8.306**	8.808**	8.414**										
	β_2	10	3.057**	2.922**	3.905**	3.271**	3.032**	2.741**	3.541**	3.204**	3.165**	3.116**										

**：推定値 5%有意，*：推定値 10%有意

青字：真値の信頼水準 95%以下，赤字：真値の信頼水準 90%以下

は、上位選択モデルでは、全ての要素選択肢の効用関数を明示的に考慮した上でモデル化がなされているためであると考えられる。

参考文献

- Ben-Akiva, M. and Lerman, S.: Discrete Choice Analysis: Theory and Application to Travel Demand, The MIT Press, 1985.
- 江田裕貴，倉内慎也：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基本特性に関する研究，土木計画学研究・講演集，Vol.54（CD-ROM），2016.

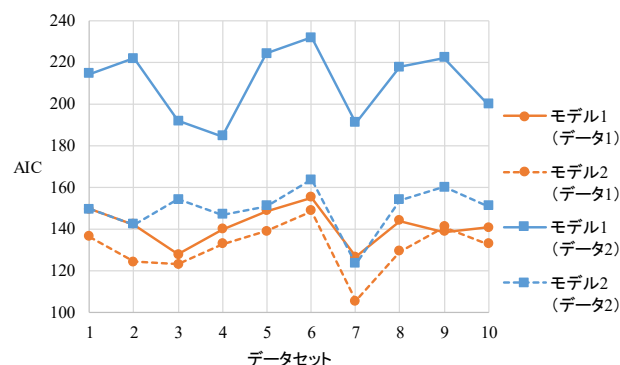


図1 AICの値