

# 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

江田 裕貴<sup>1</sup>・倉内 慎也<sup>2</sup>

<sup>1</sup>学生会員 愛媛大学大学院 理工学研究科 (〒790-8577 愛媛県松山市文京町3番)  
E-mail: eda.yuki.12@cee.ehime-u.ac.jp

<sup>2</sup>正会員 愛媛大学准教授 理工学研究科 (〒790-8577 愛媛県松山市文京町3番)  
E-mail: kurauchi@cee.ehime-u.ac.jp

本研究では、SPデータの回答形式として、望ましい、または望ましくない代替案を複数選択する形式（上位選択形式・下位選択形式）に着目し、それぞれのデータに適用可能な分析モデルを構築すると共に、それによる効用関数の推定精度を検証した。結果、提案モデルは、理論的には未知パラメータ推定における解の一意性は保証されないものの、シミュレーションデータを用いた分析により、実務的には効用関数を安定的かつ比較的精度良く推定することができることを確認した。また、選択数が増加すると尤度関数が複雑化するという問題点はあるものの、上位選択形式と下位選択形式を組み合わせることで、効用関数を効率的かつ精度良く推定できる可能性があることを示した。

**Key Words** : discrete choice model, stated preference data, best selected model, worst selected model

## 1. はじめに

交通需要を計測したデータには、実際の状況における選択行動を観測した RP データと仮想状況下での選好意思表示を観測した SP データ<sup>1)</sup>の 2 つがある。交通需要分析においては、将来の行動予測が目的であることから、信頼性の高い実際の行動と一致する RP データが長らく使用されてきたが、現存しない新規交通サービスの需要分析や、新規政策の効果などを直接計測することが不可能であることから、SP データが利用され始めている<sup>1)</sup>。また、SP データの回答形式についても、複数代替案の中から望ましい代替案を 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（多項ロジットモデル・ランクロジットモデル等）も提案されている<sup>2)</sup>。なお、近年では望ましくない代替案を尋ねる方法も使用されており、同様に選択形式・順位付け形式が存在する。

本研究では、回答者に提示した代替案の中から望ましいと思う代替案を複数選択してもらう形式（以下、上位選択形式）、反対に望ましくないと思う代替案を複数選択してもらう形式（以下、下位選択形式）の 2 つに着目する。これらの形式は、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、第 1 位の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。本形式は、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 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、本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。

本稿では、SP データの回答形式とその分析モデルの特徴を整理した後、上位選択・下位選択形式に対応したモデルを導入する。最後に、シミュレーションデータによるモデルの推定精度及び適合度の検証結果を報告する。

## 2. SPデータの回答形式と分析モデル

### (1) SPデータの回答形式

#### a) 選択形式

選択形式は、代替案集合の中から最も望ましい代替案のみを回答してもらう形式で、主に実務において使用されている。場合によっては、最も望ましくない代替案を答えてもらうこともある。この特別な形式として、回答者に仮想の代替案を 1

つ与え、その選択意向を問うものも存在し、現存しない新規交通手段の将来需要の予測などで用いられる。さらに、選好順位を回答する順位付けに対して、課題が複数の代替案の中から、ただ1つだけ選択するという比較的単純な形式であることから回答の信頼性は高いと言われている。しかし、得られる情報は最も望ましい選択結果のみのため、情報量は少ない。

### b) 順位付け形式

順位付け形式は、回答者に提示した複数の代替案を、望ましい順、あるいは望ましくない順に並べ替えてもらう形式である。ちなみに、このときすべての代替案を並べ替える必要はなく、10個のうち上位3個の順位をつけてもらうというものでもよい。その一例が、先に述べた最も望ましい代替案のみを回答してもらう選択形式である。順位付け形式のデータは、選択形式のデータと異なり、第1位の代替案以外の関係性を把握可能であることから、モデル化に用いる情報は多いものの、回答の信頼性の観点から疑問視されている。例えば、上位の順位付けの信頼性は高いが、下位の順位付けの信頼性は低いとされており、上位の順位付けのみのデータを用いた推定結果と下位を含めたすべての順位付けデータを使用した推定結果は異なることがあると報告されている<sup>3)</sup>。

## (2) 分析モデルとその推定特性

ここでは、選択・順位付け形式のデータ分析に対応した離散選択モデルを整理する。具体的には、各モデルによる選択確率式と推定データに最も適合するパラメータ値を求める最尤推定法 (MLE) を用いた場合のモデル推定特性を確認する。

### a) 多項ロジットモデル

多項ロジットモデルは、IIA特性が成り立っている場合の選択データの分析に用いられるロジットモデルであり、それぞれの代替案の効用の確率項に、独立で同一なガンベル分布を仮定している。

個人 $n$ の代替案 $z$ に対する効用は式(2a)で表される。

$$U_{zn} = V_{zn} + \varepsilon_{zn} \quad (2a)$$

ここに、 $U_{zn}$ : 効用、 $V_{zn}$ : 効用の確定項、 $\varepsilon_{zn}$ : 効用の誤差項。

個人 $n$ が代替案集合 $X$ の中から代替案 $i$ を選択する場合、「 $X$ に含まれるすべての代替案の中から代替案 $i$ が最大の効用をもたらすときに代替案 $i$ が選択される」と仮定すると、個人 $n$ の代替案 $i$ の選択確率は、

$$P_n(i) = \Pr \left[ U_{in} = \max_{z \in X} U_{zn} \right] \quad (2b)$$

ゆえに、

$$P_n(i) = \frac{\exp(V_{in})}{\sum_{z \in X} \exp(V_{zn})} \quad (2c)$$

離散選択モデルの推定ではデータの得られるもつともらしさ (尤度) が最大になるようにパラメ

ータを推定する最尤推定法が用いられる。そこで、最尤推定法によるパラメータの推定特性を確認する。データの尤度は選択確率式を用いて以下のように表される。

$$L = \prod_{n=1}^N \prod_{i \in X} P_n(i)^{d_{in}}$$

ここに、 $d_{in}$ : 個人 $n$ が代替案 $i$ を選択したとき1、そうでないとき0。

なお、最尤推定法で最大化を行う尤度関数は、上式の両辺に対数をとった式(2d)で与えられる。

$$\begin{aligned} \ln L &= \sum_{n=1}^N \sum_{i \in X} d_{in} \ln P_n(i) \\ &= \sum_{n=1}^N \sum_{i \in X} \left[ d_{in} V_{in} - d_{in} \ln \sum_{z \in X} \exp(V_{zn}) \right] \end{aligned} \quad (2d)$$

次に、パラメータの推定特性を整理する。対数尤度関数を未知パラメータに関して1階微分し、その値が0となるとき、関数が未知パラメータに関して極値を持つ。さらに、関数のヘッセ行列が正定値行列となるとき、大局的な凹関数となり、極値は最大値となる。ここで、効用関数の確定項を以下のように仮定する。

$$V_{in} = \beta_1 x_{i1n} + \beta_2 x_{i2n} + \dots + \beta_k x_{kin} + \dots + \beta_K x_{Kin}$$

ここに、 $x_{kin}$ : 個人 $n$ の代替案 $i$ に対する $k$ 番目の説明変数、 $\beta_k$ :  $k$ 番目の未知パラメータ。

式(2d)を $\beta_k$ で1階微分すると

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{n=1}^N \sum_{i \in X} x_{kin} [d_{in} - P_n(i)] \quad (2e)$$

式(2e)を $\beta_l$ で1階微分すると

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \beta_k \partial \beta_l} &= - \sum_{n=1}^N \sum_{i \in X} P_n(i) \left( x_{kin} - \sum_{z \in X} x_{kzn} P_n(z) \right) \\ &\quad \times \left( x_{lin} - \sum_{z \in X} x_{lzn} P_n(z) \right) < 0 \end{aligned} \quad (2f)$$

特に、 $k=l$ のとき、

$$\frac{\partial^2 \ln L}{\partial \beta_k \partial \beta_l} = - \sum_{n=1}^N \sum_{i \in X} P_n(i) \left( x_{kin} - \sum_{z \in X} x_{kzn} P_n(z) \right)^2 < 0$$

式(2e)より、関数が極値を持つとき、

$$\sum_{n=1}^N \sum_{i \in X} x_{kin} d_{in} = \sum_{n=1}^N \sum_{i \in X} x_{kin} P_n(i)$$

両辺を $N$ で割ると

$$\frac{1}{N} \sum_{n=1}^N \sum_{i \in X} x_{kin} d_{in} = \frac{1}{N} \sum_{n=1}^N \sum_{i \in X} x_{kin} P_n(i)$$

これより、回答者 $N$ 人のうち代替案 $i$ を実際に選択した人の各属性平均と、モデルから計算される代替案 $i$ の選択確率の属性平均が等しくなることがわかる。特に、各代替案の定数項を最大限取り入れた場合 ( $x_{kin}=1$ のとき)、

$$\frac{1}{N} \sum_{n=1}^N \sum_{i \in X} d_{in} = \frac{1}{N} \sum_{n=1}^N \sum_{i \in X} P_n(i)$$

代替案 $i$ を実際に選択した人の総和と、モデルから計算される代替案 $i$ の選択確率の総和が等しくなる。

また、式(2f)より、関数のヘッセ行列が正定値行列となるため、大局的な凹関数となり、極値は

最大値となる<sup>4)</sup>。

以上より、尤度を最大化する未知パラメータが唯一存在することが保証されることが示された。

**b) ランクロジットモデル**

ランクロジットモデルは順位付け形式のデータに対応した分析モデルである。代替案集合  $X$  に含まれる代替案に順番を付けるときを考える。第1位の代替案は、選択データと同じく式(2c)を用いて定式化できる。第2位の代替案は、 $X$  から第1位の代替案を除いた代替案集合の中から最も望ましいと選択されたとみなし式(2c)を適用する。以下、同様に第  $T$  位まで順位付けを行ったとすると、順位付けデータは選好順位ベクトル  $\mathbf{I}_n$  で表すことができ、その確率を順位付け確率とする。第  $T$  位までの順位付け確率は、以下のように表される。

$$P_n(\mathbf{I}_n) = \prod_{t=1}^T \frac{\exp(V_m)}{\sum_{m \in G_{Tn}} \exp(V_m)}$$

ここに、 $V_m$  : 個人  $n$  が  $t$  番目に選好する代替案に対する効用、 $G_{Tn}$  : 個人  $n$  が  $t$  番目に望ましい代替案を選ぶ際に直面している代替案集合 ( $i \in G_{Tn} \subseteq X$ )

続いて、多項ロジットモデル同様に最尤推定法によるパラメータの推定特性を確認する。ランクロジットモデルの尤度関数と対数尤度関数は以下の通りである。

$$L = \prod_{n=1}^N P_n(\mathbf{I}_n)$$

$$\ln L = \sum_{n=1}^N \sum_{t=1}^T \sum_{i \in G_{Tn}} d_{im} \ln \left[ \frac{\exp(V_m)}{\sum_{m \in G_{Tn}} \exp(V_m)} \right] \quad (2g)$$

ここに、 $d_{im}$  : は個人  $n$  が代替案  $i$  を  $t$  番目に選択したとき1, そうでないとき0。

また、最尤推定法によるパラメータの推定特性として、対数尤度関数が極値を持つ条件、及び関数のヘッセ行列が正定値行列となるかどうかを確認する。

対数尤度関数を  $\beta_k$  で1階微分すると

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{n=1}^N \sum_{t=1}^T \sum_{i \in G_{Tn}} x_{kim} [d_{im} - P_n(i_t)] \quad (2h)$$

ただし、 $P_n(i_t)$  は個人  $n$  が代替案  $i$  を  $t$  番目に選択する確率である。

式(2h)を  $\beta_t$  で1階微分すると、

$$\frac{\partial^2 \ln L}{\partial \beta_k \partial \beta_t} = - \sum_{n=1}^N \sum_{t=1}^T \sum_{i \in G_{Tn}} P_n(i_t) \left( x_{kim} - \sum_{t'=1}^T \sum_{i' \in G_{Tn}} x_{kit'n} P_n(i_{t'}) \right) \times \left( x_{lit'n} - \sum_{t'=1}^T \sum_{i' \in G_{Tn}} x_{lit'r'n} P_n(i_{t'}) \right) < 0 \quad (2i)$$

式(2h)より、関数が極値を持つとき、

$$\sum_{n=1}^N \sum_{t=1}^T \sum_{i \in G_{Tn}} x_{kim} d_{im} = \sum_{n=1}^N \sum_{t=1}^T \sum_{i \in G_{Tn}} x_{kim} P_n(i_t)$$

両辺を  $N$  で割ると

$$\frac{1}{N} \sum_{n=1}^N \sum_{t=1}^T \sum_{i \in G_{Tn}} x_{kim} d_{im} = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^T \sum_{i \in G_{Tn}} x_{kim} P_n(i_t)$$

これより、回答者  $N$  人のうち実際に代替案  $i$  を  $T$  番目までに選択した人の各属性平均と、モデルから計算される代替案  $i$  の  $T$  番目までの選択確率の属性平均が等しくなることがわかる。特に、各代替案の定数項を最大限取り入れた場合 ( $x_{kim}=1$  のとき) ,

$$\frac{1}{N} \sum_{n=1}^N \sum_{t=1}^T \sum_{i \in G_{Tn}} d_{im} = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^T \sum_{i \in G_{Tn}} P_n(i_t)$$

実際に代替案  $i$  を  $T$  番目までに選択した人の総和と、モデルから計算される  $T$  番目までの代替案  $i$  の選択確率の総和が等しくなる。

また、式(2i)より、関数のヘッセ行列が正定値行列となるため、大局的な凹関数となり、極値は最大値となる。

以上より、ランクロジットモデルは多項ロジットモデルの積で表されることから、尤度を最大化する未知パラメータが唯一存在することが保証されることが示された。

**3. 上位選択モデル**

**(1) 上位選択形式**

上位選択形式は、代替案集合の中から望ましいと思う代替案を複数選択してもらう形式で、例えば、「鉄道・バス・自動車・自転車・タクシーという代替案の中からよく利用する交通手段を2つ選んでください」という形式がこれらにあたる。形式としては、順位付けに似ており、単純に望ましい代替案を回答してもらうだけであるため、データの信頼性は比較的高いと思われる。しかし、選ばれた代替案同士の順位付けは不明であることから、情報量の観点では、順位付けに劣る。

**(2) 上位選択モデル**

**a) 上位選択モデルの一般式**

本モデルは、上位選択形式のデータの分析に適用可能なモデルである。多項ロジットモデル・ランクロジットモデルの場合と同様にIIA特性が成り立つと仮定する。ここで、回答者が代替案集合  $X$  から上位  $T$  個 ( $T > 2$ ) の代替案を回答した場合を考える。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を考慮することで、上位選択モデルを定式化することができる。よって、回答者が上位  $T$  個を選択したときの上位選択モデルの一般式は次式のようになる。

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

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 ( $G_{Tn} \subseteq R_{Tn}$ ) .

b) 上位選択モデルの例

前記の上位選択モデルの適用例を述べる。「代替案集合Xから上位3つの代替案を回答する」ケースを想定する。今、回答者が、代替案1, 代替案2, 代替案3の3つを望ましいと回答したものとすると、回答者の3つの代替案の選好順位として考えるパターンは表-1の6種である。ここで、各々の順位付けパターンが生起する確率は、前章で示したランクロジットモデルを用いて表-1のように表せる。ここに、 $P(1,2,3)$ は代替案1が第1位に、代替案2が第2位に、代替案3が第3位に順位付けされる同時確率、 $P(1)$ は代替案集合Xから代替案1が第1位に選択される確率、 $P(2|1)$ は代替案1が第1位に選択された後に、代替案集合の中から代替案2が選択される確率、 $P(3|1,2)$ は代替案1, 代替案2の順で選択された後に、代替案集合の中から代替案3が選択される確率である。ゆえに、代替案集合Xから、代替案1, 代替案2, 代替案3の3つが望ましいと回答する確率は、表-1に示した全パターンの周辺和をとることで、次式のように表すことができる。

$$P(1,2,3) = P_{(1,2,3)} + P_{(1,3,2)} + P_{(2,1,3)} + P_{(2,3,1)} + P_{(3,1,2)} + P_{(3,2,1)}$$

ただし、 $P(1,2,3)$  は代替案集合Xから、代替案1, 代替案2, 代替案3の3つを選択する確率である。

なお、上記の上位選択モデルにおいて、XにJ個の代替案が含まれるとすると、すなわち順位付けのパターン数 $R_{Tn}$ は $J^3 P_T$ 通りあり、回答者が直面している代替案数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。例えば、10個の代替案から上位3つを選ぶときでは、組み合わせは ${}_{10}P_3 = 720$ 通りとなる。

(3) パラメータの推定特性

ここでは最尤推定法を用いた場合の、上位選択モデルの推定特性を確認する。

a) 尤度関数

上位選択モデルの尤度関数及び対数尤度関数は、ランクロジットモデルによる順位付け確率の和で表すことができるため、以下のように表される。

$$L = \prod_{n=1}^N P_n(C_{Tn})$$

$$\ln L = \sum_{n=1}^N \ln P_n(C_{Tn})$$

$$= \sum_{n=1}^N \ln \sum_{G_{Tn} \in R_{Tn}} \prod_{t=1}^T \prod_{i \in G_{Tn}} \left( \frac{\exp(V_{in})}{\sum_{m \in G_{Tn}} \exp(V_m)} \right)^{d_{in}} \quad (3a)$$

b) パラメータの推定特性

前章と同じく対数尤度関数が極値を持つ条件、及び関数のヘッセ行列が正定値行列となるかどうかを確認する。なお、効用関数は、前章と同様に線形効用関数を用いる。

式(3a)を $\beta_k$ で1階微分する。表記の簡略化のために、ランクロジットモデルの選択確率に相当する部分を、

$$L_r = \prod_{t=1}^T \prod_{i \in G_{Tn}} \left( \frac{\exp(V_{in})}{\sum_{m \in G_{Tn}} \exp(V_m)} \right)^{d_{in}}$$

とおいた上で微分を行うと、

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{n=1}^N \left\{ \frac{\sum_{G_{Tn} \in R_{Tn}} L_r}{\sum_{G_{Tn} \in R_{Tn}} L_r} \right\}$$

$$= \sum_{n=1}^N \frac{\sum_{G_{Tn} \in R_{Tn}} \left\{ L_r \sum_{t=1}^T \sum_{i \in G_{Tn}} x_{kim} [d_{in} - P_n(i_r)] \right\}}{\sum_{G_{Tn} \in R_{Tn}} L_r} \quad (3b)$$

式(3b)を $\beta_l$ で1階微分すると、

$$\frac{\partial^2 \ln L}{\partial \beta_k \partial \beta_l} = \sum_{n=1}^N \left[ \frac{\sum_{G_{Tn} \in R_{Tn}} L_r \{ (\ln L_r)^2 + (\ln L_r)' \}}{\sum_{G_{Tn} \in R_{Tn}} L_r} \right]$$

$$- \left[ \frac{\sum_{G_{Tn} \in R_{Tn}} [L_r (\ln L_r)]^2}{\left( \sum_{G_{Tn} \in R_{Tn}} L_r \right)^2} \right] \quad (3c)$$

式(3b)より、関数が極値を持つとき、

$$\sum_{n=1}^N \sum_{G_{Tn} \in R_{Tn}} \left\{ L_r \sum_{t=1}^T \sum_{i \in G_{Tn}} x_{kim} d_{in} \right\} = \sum_{n=1}^N \sum_{G_{Tn} \in R_{Tn}} \left\{ L_r \sum_{t=1}^T \sum_{i \in G_{Tn}} x_{kim} P_n(i_r) \right\}$$

このとき、条件を満たす未知パラメータは複数存在する可能性があることが分かる。

表-1 上位選択確率式の導出例

パターン	1位	2位	3位	順位付け確率式
1	1	2	3	$P_{(1,2,3)} = P(1) \times P(2 1) \times P(3 1,2)$
2	1	3	2	$P_{(1,3,2)} = P(1) \times P(2 1) \times P(2 1,3)$
3	2	1	3	$P_{(2,1,3)} = P(2) \times P(1 2) \times P(3 2,1)$
4	2	3	1	$P_{(2,3,1)} = P(2) \times P(3 2) \times P(1 2,3)$
5	3	1	2	$P_{(3,1,2)} = P(3) \times P(1 3) \times P(2 3,1)$
6	3	2	1	$P_{(3,2,1)} = P(3) \times P(2 3) \times P(1 3,2)$

一方、式(3c)より、関数のヘッセ行列が確実に正定値行列となるとは断定できないため、大局的な凹関数となるとは限らない。

以上より、上位選択モデルに関しては、解の一意性は保証されないことから、モデル推定において、初期値を変化させた上で、対数尤度が最大となるような解を導き出す必要がある。

#### 4. Best-Worst選択

ここまででは、望ましい代替案を選択する場合を考えてきたが、望ましくない代替案を尋ねた場合も効用関数の推定精度の向上に貢献しうると考えられる。望ましい代替案を選択する場合は望ましさが最大になる代替案を選択すると仮定したが、望ましくない代替案を選択する場合は、望ましくなさが最大になる代替案を選択すると仮定する。ここで、 $B_n(i)$ ：個人 $n$ が代替案集合 $X$ から代替案 $i$ を望ましいと選択する確率、 $W_n(j)$ ：個人 $n$ が代替案集合 $X$ から代替案 $j$ を望ましくないと選択する確率（ただし、 $i \neq j$ ）、 $B_{zn}$ ：個人 $n$ の代替案 $z$ に対する望ましさ（Best効用）、 $W_{zn}$ ：個人 $n$ の代替案 $z$ に対する望ましくなさ（Worst効用）とすると、以下の関係が成り立つ。

$$B_n(i) = \Pr(B_{in} = \max_{z \in X} B_{zn})$$

$$W_n(j) = \Pr(W_{jn} = \max_{z \in X} W_{zn})$$

Best効用が最大となるとき、Worst効用は最小となるはずである。 $B_{zn} = U_{zn}$ とすると、

$$B_n(i) = \Pr(U_{in} = \max_{z \in X} U_{zn}) \quad (4a)$$

$$W_n(j) = \Pr(U_{jn} = \min_{z \in X} U_{zn}) \quad (4b)$$

$$B_{zn} = -W_{zn} = U_{zn} \quad (4c)$$

ゆえに、

$$B_{zn} = V_{zn} + \varepsilon_{zn}$$

$$W_{zn} = -V_{zn} - \varepsilon_{zn}$$

となる<sup>5)</sup>。

このことから、上記の効用をモデルに代入すると、順位付け、上位選択形式についても同様のことがいえる。なお、望ましくない代替案を複数選択する形式を今後は「下位選択形式」と述べる。ただし、このとき尤度関数の極値は最小値をとることに留意すべきである。

#### 5. シミュレーション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<sup>6)</sup>が用いられることから、理論的に解の一意性が保証されるようなモデルであっても、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、Best側の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、及びWorst側の多項ロジットモデル、ランクロジットモデル、下位選択モデルの6つをそれぞれ適用し、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Best選択データ、最も低い代替案だけを取り出したWorst選択データ、高い順に $T$ 個取り出した上位選択データ、低い順に $T$ 個取り出した下位選択データと、5種類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、各データに対応したモデルを適用し、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。今回は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値を用いた。なお、AICはその値が小さくなるほど適合度が高いことを表す。また、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。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い、効用関数は以下の通りとする。

$$U_{in} = \alpha_i + \beta_1 x_{in} + \varepsilon_{in}$$

ただし、 $\alpha_i$ ：代替案 $i$ の定数項（代替案1では0に固定）、 $\beta_1$ ：説明変数の限界効用、 $x_{in}$ ：個人 $n$ の代替案 $i$ に対する説明変数、 $\varepsilon_{in}$ ：誤差項

##### (1) 上位選択モデルの適合度

図-1、図-2ではBest順位付けデータ、Worst順位付けデータについてランクロジットモデルを適用し、図-3では、すべての代替案の選好順位が判明しているデータに対し、Best・Worstランクロジットモデルを適用した際の適合度を比較している。また、図-4は上位選択モデル、多項ロジットモデル、順位がすべて判明している際のランクロジットモデルの適合度、図-5は下位選択モデル、多項ロジットモデル、順位がすべて判明している際のランクロジットモデルの適合度を載せる。図-1、図-2より全体的に、ランクロジットモデルに関して、選好順位が判明していくごとにAICは小さくなっていくが、ある程度順位が判明すると値はほぼ変わらなくなる。つまり、順位付け調査では、ある程度順位付けが行われていればすべての代替案に順位付けがされていなくとも、選好順位は正しく予測できることを示している。また、データセット5では、多項ロジットモデルを用いた選択データのAIC値が大きく剥離しているところがある。この理由として、サンプル数が100と少ないことが影響していると考えられ、選択調査はサンプル数200以

上で行うべきであり，サンプル数100程度では，順位付けデータが有用であると考えられる。

図-3から，両モデルの適合度はいずれのセットにおいても大差ないことがわかる．これより，Best・Worstランクロジットモデルともに同程度のデータの再現性を有していると判断できる。

図-4，図-5から，全体的に上位選択モデル・下位選択モデルの適合度は，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個の代替案までを求めた場合には，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，選択していない最も望ましくない代替案を1つ選択することと実務的な意味においては同義であるためである．このことは下位選択においても同

様である．ゆえに，選択数が増加すると尤度関数が複雑化するという問題点があるといえる．

そこで，同義となる条件を比較する．図-6～図-11はそれぞれ上位選択形式と下位選択形式の同義となる場合における両モデルの適合度である．図-6，図-11より，選択時と選択数6個の際のモデルの適合度は低下している．一方，両形式とも選択数が2個以上であるとみなせる場合には，上位選択・下位選択モデルとも同程度の適合度となっていることから，その条件下では上位選択形式と下位選択形式を組み合わせることで，効用関数を効率的かつ精度良く推定できる可能性があると考えられる．

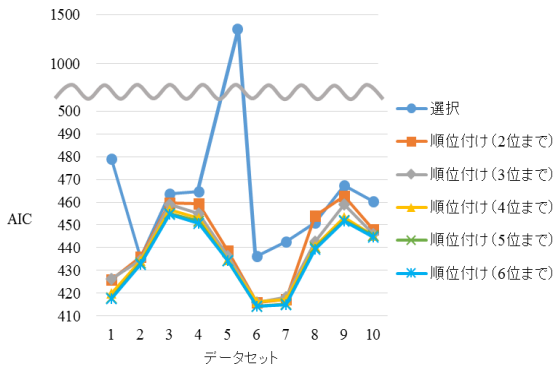


図-1 Bestランクロジットモデルの適合度

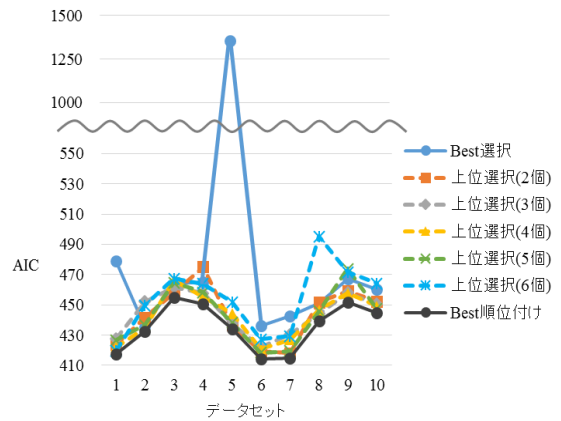


図-4 上位選択データの適合度

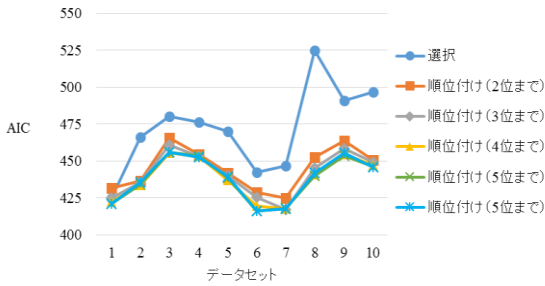


図-2 Worstランクロジットモデルの適合度

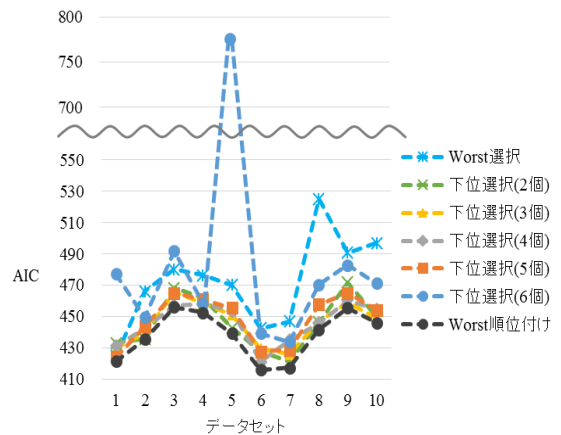


図-5 下位選択データの適合度

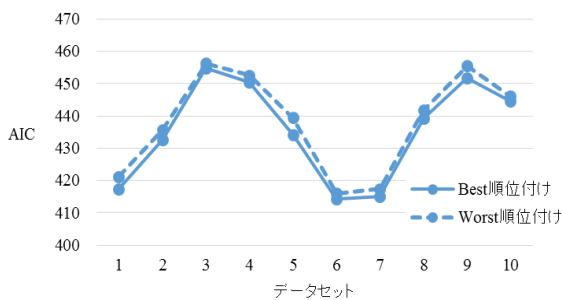


図-3 ランクロジットモデルの適合度

(2) Best-Worst選択によるパラメータ推定精度

表-2にBest選択における多項ロジットモデル，ランクロジットモデル，上位選択モデルのパラメータ推定値，表-3にWorst選択における多項ロジットモデル，ランクロジットモデル，下位選択モデルのパラメータ推定値を示す．これより，ランクロジットモデルでは，順位付けの位が増加するにつれて推定値の有効性は高くなる傾向にあるものの，一部では有効性が低下するものも確認された．これは，モデル推定に際して，変数が不足している状態で順位付けデータを用いると，その影響が確率項に含まれ，バイアスがかかり推定精度が低下してしまうことを示唆している．一方，上位選択モデル・下位選択モデルの推定値は局所最適解であるが，サンプル数が少ない中でもある程度安定した結果が得られたが，ランクロジットモデルに比べて，推定精度は低い．

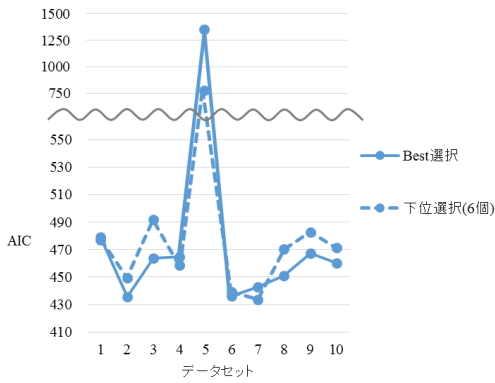


図-6 Best選択と下位6個選択の適合度

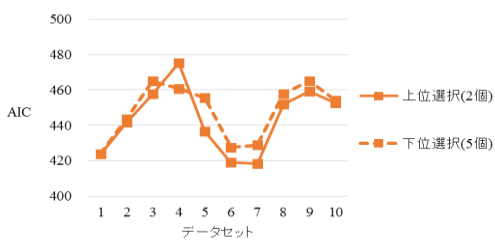


図-7 上位2個選択と下位5個選択の適合度

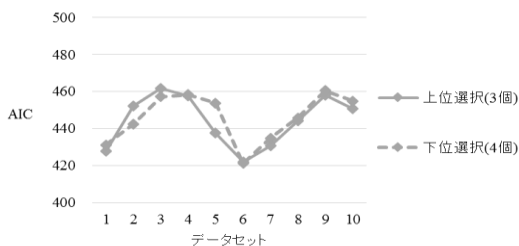


図-8 上位3個選択と下位4個選択の適合度

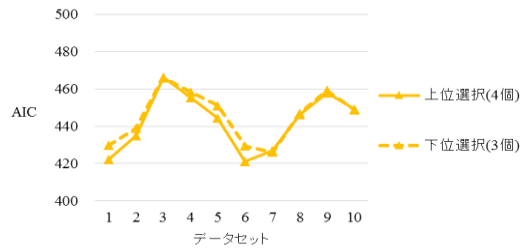


図-9 上位4個選択と下位3個選択の適合度

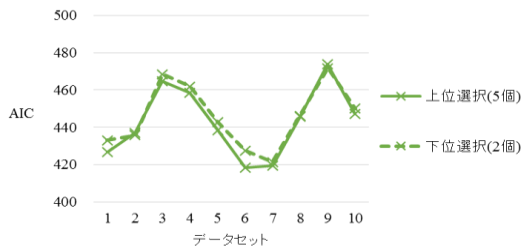


図-10 上位5個選択と下位2個選択の適合度

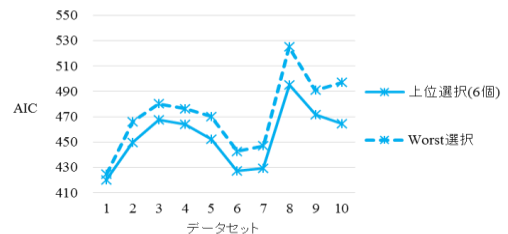


図-11 上位6個選択とWorst選択の適合度





## 6. まとめ

本研究では、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提案に関して、その基本特性を整理した。具体的には、提案モデルは既存の多項ロジットモデル・ランクロジットモデルとは異なり、対数尤度関数のヘッセ行列が正定値行列となることが保証されず、数値解法によるパラメータ推定値は局所的最適解となることがあるものの、シミュレーションデータによる推定結果より、提案モデルはある程度選好順位は表現できること、及び選択数が増加すると順位付けの組み合わせが増大し推定精度が低下する問題はあるが、実務的には効用関数を安定的かつ比較的精度良く推定することができることを確認した。また、選択数が増加すると尤度関数が複雑化するという問題点はあるものの、上位選択形式と下位選択形式を組み合わせることで、効用関数を効率的かつ精度良く推定できる可能性と考えられる。今後の課題として、選好意思表示と実際の選好の一致性に疑問が持たれていることから、実験により、回答形式による回答負荷や回答の信頼性を検証する必要がある。

### 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) 北村隆一，森川高行，佐々木邦明，藤井聡，山本俊行：交通行動の分析とモデリング—理論/モデル/調査/応用—，技報堂出版，2002
- 3) Bradley, M. and Daly, A.: Use of the logit scaling approach to the test for rank-order and fatigue effects in stated preference data, *Transportation* 21, pp.167-184, 1994.
- 4) McFadden, D.: Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics*, P. Zarembka ed., Academic Press, New York, pp.105-142, 1973.
- 5) A.A.J. Marley, J.J. Louviere: Some probabilistic models of best, worst, and best-worst choices, *Journal of Mathematical Psychology* 49, pp.464-480, 2005.
- 6) Train, K. E.: *Discrete Choice Methods with Simulation*, Cambridge University Press, 2003.