

選択肢間の類似性を考慮した時刻選択モデル* Time choice models considering similarity among alternatives

藤原章正**・杉恵頼寧***・岡村敏之****・神田佑亮*****

By Akimasa FUJIWARA, Yoriyasu SUGIE, Toshiyuki OKAMURA, Yusuke KANDA

1. はじめに

ランダム効用最大化理論に基づく非集計多項ロジット(MNL)モデルは、効用の誤差分布は選択肢間で独立で同一分布に従う(IID: Independent and Identically Distribution)と仮定している¹⁾。この仮定により、モデル式が極めてシンプルでかつ推定が容易であるという特徴を持っており、交通行動の離散選択問題の分析に頻繁に適用されている。しかし、選択肢間に類似性が存在したり誤差分布が不均一である場合には、このような前提条件(IIA 特性)が成立しないため、推定結果にバイアスが生じるという重大な問題点がある。

選択肢間の類似性や誤差分布の類似性が無視できない事例の1つとして、「時刻」選択問題が考えられる。時刻選択問題では、通常連続量を離散データにカテゴリライズしてMNLモデルに適用することが多く見られ、わずかな数分の違いで別々の選択肢として扱われる場合が生じ、類似性が全くないとは言えないであろう。また、カテゴリライズの違いによって、誤差分散が変動することも予想される。

本研究では上記2つの問題のうち、選択肢間の類似性の問題に着目して分析する。選択肢間の誤差分布の類似性を考慮できる離散選択モデルを用い、時刻のカテゴリライズの違いが選択肢間の類似性に及ぼ

す影響について分析を行うことを目的とする。時刻選択モデルとして選択肢間の類似性をパラメータとして計量化できるPCLモデル(Paired Combinatorial Logit model)²⁾を採用する。

2. PCLモデル

PCLモデルは選択肢のペアごとの類似性を考慮することにより、MNLモデルで仮定するIIA特性を緩和するモデルである。IIA特性の緩和を目指したモデルはいくつか存在するが、その中でPCLモデルは一般的に用いられているネスティッドロジット(NL)モデルと同様に、誤差分布を非独立かつ同一分布に従うと仮定し、一般化極値分布に仮定した

$$P_i = \frac{\sum_{j=1}^{n-1} (1 - \sigma_{ij}) \left\{ \exp\left(\frac{V_i}{1 - \sigma_{ij}}\right) + \exp\left(\frac{V_j}{1 - \sigma_{ij}}\right) \right\}^{-\sigma_{ij}} \exp\left(\frac{V_i}{1 - \sigma_{ij}}\right)}{\sum_{q=1}^{n-1} \sum_{r=q+1}^n (1 - \sigma_{qr}) \left\{ \exp\left(\frac{V_q}{1 - \sigma_{qr}}\right) + \exp\left(\frac{V_r}{1 - \sigma_{qr}}\right) \right\}^{-\sigma_{qr}}} \quad (1)$$

GEVモデルから導出される。PCLモデルの一般式は以下ようになる。

ここで P_i は選択肢 i を選択する確率、 V_i は選択肢 i の効用を表す。PCLモデルは全ての選択肢間の類似性を表すパラメータ σ を含んでいる。類似度パラメータは $0 \leq \sigma < 1$ の値をとり、 σ の値が 0 に近いほど選択肢の誤差は独立であり、 1 に近くなるほど相関が大きいことを示す。もし全ての $\sigma = 0$ ならば、MNLモデルとなる。

PCLモデルがNLモデルと異なる点は、PCLモデルが全ての選択肢ペア間の類似性パラメータを含んでいるのに対して、NLモデルでは同一のネスト内の選択肢間における類似性は認めるものの、ネスト外の選択肢とは完全に独立であると仮定している。

* キーワード：PCLモデル、交通行動分析、発生交通

** 正会員 工博 広島大学大学院国際協力研究科
(東広島市鏡山 1-5-1 Phone&Fax 0824-24-6921)

*** 正会員 工博 広島大学大学院国際協力研究科
(東広島市鏡山 1-5-1 Phone&Fax 0824-24-6919)

**** 正会員 工博 広島大学大学院国際協力研究科
(東広島市鏡山 1-5-1 Phone&Fax 0824-24-6922)

***** 学生員 広島大学大学院国際協力研究科
(東広島市鏡山 1-5-1 Phone&Fax 0824-24-5970)

場合によってはネストを越えて選択肢間に類似性が存在する場合もあり、このような場合には非現実的となる。

また、NL モデルではケースに最も適したネスト構造を分析者が事前に想定し、その検証を必要とするが、PCL モデルではこのような全てのペア間の類似性パラメータの推定を行うことにより、NL モデルで必要であった構造の想定や検証を省略することが可能である。

3. 会社到着時刻選択行動の分析

(1) 分析データの概要

1997 年にフレックスタイム制度を導入した広島市内の企業を対象にして出社行動の実態調査を行った³⁾。本研究ではこの調査データを用いて出社時刻選択モデルを推定する。有効回答は 301 人であり、初期サンプルの 96.2%を占める。

出社時刻の質問ではあらかじめカテゴリ化した時間帯を選ぶのではなく、回答用紙に直接時刻を記入してもらった。そのため連続データが得られており、離散選択モデルの目的変量として使用するには次の作業が必要であった。

1 つ目は時刻のカテゴリ化であり、図 1 で示す 4 つのケースを想定する。図 1 のケース 1～ケース 3 に示す、標準始業時刻の 9:00 を基準とし、時間幅を 20 分と 40 分、60 分とした場合であり、ケース 4 に示す 60 分の時間幅で 9:30 を基準とした区分とした。いずれも 3 項選択モデルとなる。カテゴリ化の違いによって選択肢間の類似性の違いが生じるかどうか確認するためである。

もう 1 つは代替選択肢に関する説明要因の水準値の補填である。実際に出社した時間帯以外の時間帯について交通サービス水準に関する情報が得られていないため、これらの情報をアイテム欠損データと考え、EM アルゴリズム法⁴⁾によってデータを補填した。EM アルゴリズムの詳細な説明は参考論文 4) に委ねるが、基本的な考え方は欠損データの期待値を有効データ相互の多重相関関係から推定する統計手法である。

(2) 集計分析

図 1 に 10 分間隔の時間帯別の回答者数を示す。

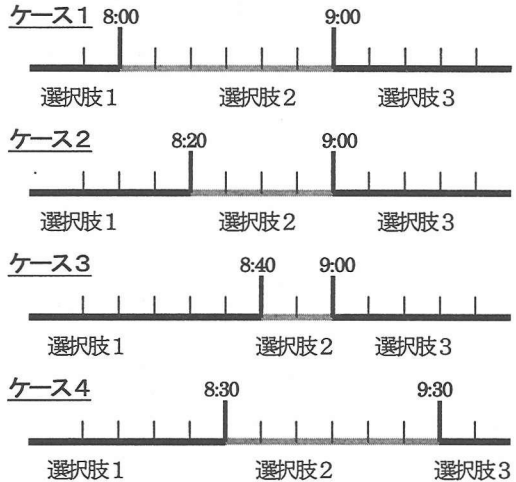


図 1. 選択肢のカテゴリ化

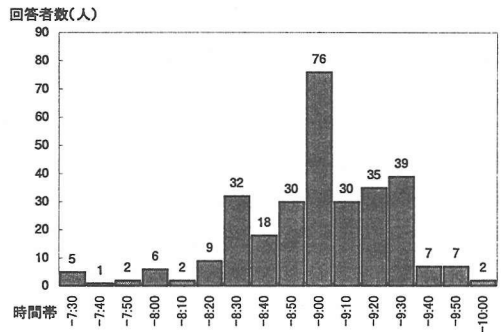


図 2. 会社到着時刻の分布

表 1. ケースによる選択肢別サンプル数の比較

	選択肢1	選択肢2	選択肢3
ケース1	14人	167人	120人
ケース2	25人	156人	120人
ケース3	75人	106人	120人
ケース4	57人	228人	16人

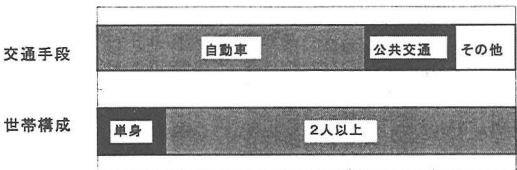


図 3. 回答者の通勤交通手段と世帯構成

平均会社到着時刻は 8 時 56 分となり、その時刻前後に回答が集中している。これを図 2 のようにカテゴリ化したところ、選択肢別のサンプル数は表 1 に示すとおりとなった。

図 3 に回答者の通勤交通手段と、世帯構成について集計した結果を示す。半数以上が自動車通勤を行っており、公共交通利用者は 2 割程度である。また、全回答者の 8 割程度が世帯人数 2 人以上であり、単身者は 2 割にすぎない。

図 4 に回答者の年齢と会社到着時刻のクロス集計結果を示す。出発時刻が遅くなるほど若年層の占める割合が多い。

図 5 は回答者の通勤所要時間と会社到着時刻のクロス集計結果を示している。所要時刻が短い人ほど会社に遅く到着する傾向にある。これは、通勤時間の短い人ほど交通渋滞など遅れに対するリスクが少ないためであると考えられる。よって、会社到着時刻の選択には、これらの要因の影響を考慮する必要がある。

4. 会社到着時刻選択モデルの推定

(1) 選択肢の時間幅が類似性に及ぼす影響

上記の会社到着時刻データを用いて、9:00 を基準としてケース 1 ～ケース 3 に PCL モデルで推定し、選択肢の時間幅を変化させた場合に現れる変化を確認した。表 2 に推定結果を示す。

まずモデルの適合指標を比較すると、尤度比より、すべてのモデルとも説明力は妥当と判断される。

「公共交通」、「世帯構成」などの一部パラメータ推定値の符号がケースにより変化している。これは時間幅の大きさにより、公共交通利用者や単身者が異なった選択肢に属するためである。「所要時間」、「年齢」についてはカテゴリ化の違いによる影響は小さい。

類似性パラメータに着目すると、ケース 1 のモデルでは全ての類似性パラメータが有意となった。また、そのほかのモデルでは 3 つのうち 2 つのパラメータが有意となった。全てのモデルについて共通していえることは、 σ_{23} (選択肢 2 と選択肢 3 の間) パラメータ推定値、 t 値共に大きいことである。その原因として考えられることは、回答者の回答時刻がこの 2 つの選択肢の境界付近に集中しているため

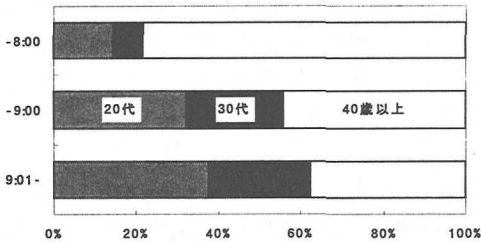


図 4. 年齢と会社到着時刻のクロス集計

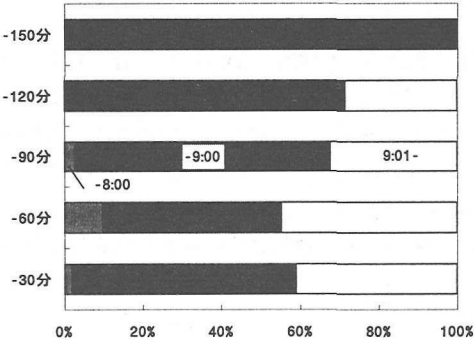


図 5. 通勤所要時間と会社到着時刻のクロス集計

表 2. PCL モデル推定結果 (時間帯幅別)

モデル	ケース 1	ケース 2	ケース 3
説明変数			
所要時間 [C]	-0.430 * (-2.493)	-0.451 ** (-2.997)	-0.563 (-0.298)
自動車通勤 ¹⁾ [3]	0.306 (1.591)	0.528 * (2.383)	0.637 ** (2.718)
公共交通 ²⁾ [2]	0.417 (1.613)	-0.330 (-1.593)	0.130 (0.631)
年齢 [1]	0.033 * (2.308)	0.050 (3.861)	0.058 * (2.037)
世帯構成 ³⁾ [2]	0.061 (0.372)	0.124 (0.442)	-0.076 (-0.321)
定数項1	3.537 ** (5.517)	3.526 ** (5.053)	3.263 ** (2.849)
定数項2	3.543 ** (5.882)	3.816 ** (5.768)	3.200 ** (2.860)
類似性 σ_{12}	0.126 ** (2.578)	0.252 (1.715)	0.000 (0.939)
パラメータ σ_{13}	0.414 ** (3.778)	0.787 ** (4.600)	0.008 ** (2.440)
σ_{23}	0.909 ** (13.427)	0.855 ** (12.888)	0.912 ** (11.648)
初期尤度	-330.680	-330.680	-330.680
最大尤度	-242.355	-258.033	-303.928
尤度比	0.255	0.218	0.065
的中率 (%)	59.47	45.85	54.49
サンプル数	301	301	301

1) 1: 自動車通勤者 2) 1: 公共交通利用者 3) 1: 単身者

(): t 値 * 1% 有意 ** 5% 有意

[#]: 効用関数に説明変数を入れた選択肢番号

[C]: 共通変数

と考えられる。仮にわずかな数分カテゴリー区分がずれたならば、選択肢の境界付近にある多くの回答が他の選択肢に入る可能性がある。このことからこの選択肢間の類似性が非常に強くなったものと考えられる。

σ_{13} (選択肢1と選択肢3の間) は全てのモデルにおいて有意な推定値となっている。これは両選択肢に共通の例えば車内混雑や他の人との勤務時間のずれなどといった、省略変数の影響が作用していることを表している。

σ_{12} (選択肢1と選択肢2の間), σ_{23} のt値について着目すると、選択肢の時間幅が小さくなるにつれてt値も小さくなっている。このことから時間幅が小さくなるにつれて選択肢2の独立性が高まる傾向にあり、類似性が小さくなっていると考えられる。

以上の結果から、選択肢の時間幅を変更した場合にはそれに伴って類似性の大きさが変化することが確認できた。

(2) 選択肢の区切り方の違いが類似性に及ぼす影響

同一のデータを用い、選択肢の時間幅を60分にして回答が集中する9:00を基準にカテゴリ化したモデル(ケース1)と、回答の集中を選択肢の中心付近に回答が集中するような、9:30を基準にカテゴリ化したモデル(ケース4)で推定を行い、結果を比較した。推定結果を表2に表す。

まずモデルの適合度を見ると、尤度比から両モデルとも妥当であるといえる。説明変数についてはモデルにより符号が異なるものがあるが、原因は前節のモデルと同じである。そのほかのパラメータについては妥当である。

類似性パラメータについて着目すると、2つのモデルで傾向が異なっている。9:00が基準のモデルでは、選択肢が集中している付近の類似性パラメータ σ_{23} の推定値、t値共に非常に大きい。9:30のモデルでは σ_{23} の値は比較的小さい。このことから、選択肢の境界付近に回答が集中している場合は類似性が強く現れやすいことを示している。

これらのことから、選択肢の区切り方を変化させた場合、現れる類似性が変化することが確認できた。

5. 結論

以上の結果から、時刻選択問題のような本来連続料であるものを離散的に取り扱う場合、選択肢の時間幅や区切り方によって、出現する選択肢間の類似性が変化することが確認できた。このような類似性を予想した上で時刻の選択肢をカテゴリ化することは、多大な労力を必要とする。PCLモデルを適用すると、カテゴリ化の違いによって生ずる類似性パターンの変化をモデルパラメータとして取り扱うことができるため、このような労力は不要となる。

この点は従来のロジットモデルの時刻選択問題への適用可能性を大いに高める結果として評価される。

表3. PCLモデル推定結果(境界別)

説明変数	PCLモデル-1		PCLモデル-2	
	推定値	t値	推定値	t値
所要時間 [C]	-0.430 *	-2.493	-0.375 *	-3.113
自動車通勤 [3]	0.306	1.591	0.177	1.125
公共交通 [2]	0.417	1.613	0.156	0.694
年齢 [1]	0.033 *	2.308	0.033 *	3.448
世帯構成 [2]	0.061	0.372	-0.403	-3.104
定数項1	3.537 **	5.517	2.520 **	6.046
定数項2	3.543	5.882	1.183	2.596
類似性 σ_{12}	0.126 *	2.578	0.105 *	0.673
パラメータ σ_{13}	0.414 **	3.778	0.630 **	3.058
σ_{23}	0.909 **	13.427	0.690 **	2.518
初期尤度	-330.680		-330.680	
最大尤度	-242.355		-195.779	
尤度比	0.255		0.398	
的中率 (%)	59.47		75.75	
サンプル数	301		301	

説明変数の定義・単位は表1と同じ

*1%有意

**5%有意

[#]: 効用関数に説明変数を入れた選択肢番号

[C]: 共通変数

参考文献

- 1) 社団法人交通工学研究会：やさしい非集計分析，1993.
- 2) Chu, C.: A Paired Combinatorial Logit Model for travel demand analysis, Proc. of 5th WCTR, Vol. 4, Yokohama, pp.295-309, 1989
- 3) 周藤，杉恵，藤原：フレックスタイム制度下における通勤行動の時間的変化，土木計画学研究論文集，No.12, pp.655-662, 1998
- 4) 藤原，杉恵，原田：交通調査における無回答バイアスの取り扱い方法，土木計画学研究講演集，No.21(2), pp.61-64, 1998