【土木計画学研究・論文集 No.16 1999年9月】

# Mixed Logit Model とプロビットモデルの推定特性に関する比較分析*
## -鉄道経路選択モデルを例に-

# A Comparative Study on the Estimation Characteristics of Mixed Logit Model and Probit Model
## -In the Case of Railway Route Choice Model-

清水　哲夫**　屋井　鉄雄***
*By Tetsuo SHIMIZU and Tetsuo YAI*

## 1．はじめに

従来から，IIA 特性を持つロジットモデルを改良し，選択肢間の相関を表現する試みは数多く行われてきた．この流れの中で，筆者らはプロビットモデルの誤差項を構造化し（以下では構造化プロビットモデルと呼ぶ），分散共分散行列内の推定パラメータ数を減らすことにより，主に鉄道経路選択に対するプロビットモデルの適用可能性を拡大してきた[1),2),3)]．

一方海外では，ロジットモデルをベースとし，選択肢間の相関を表現する Mixed Logit Model が交通計画分野に適用されつつある．このモデルでは，効用関数の誤差項を選択肢間の相関関係によって記述する項と選択肢固有の項に分割し，サンプルごとの選択肢の類似性を表現するものである．例えば，Brownstone *et al* (1997)[4)]は省エネカー購入の選好モデルに，Bhat (1998)[5)]は都市間交通を対象とした出発時刻と交通機関の同時選択モデルへ Mixed Logit Model を適用し，共に SP データを用いて推定を試みている．ただし，これらの分析対象は選択肢間の相関が大きくないと考えられ，選択肢間の相関が大きいケースを対象とした推定特性については引き続き検討する必要がある．

これらモデルの巧拙については，現在様々な議論が展開されているが，筆者らは以下の 2 点に関して構造化プロビットの方が優れていると考える．第一に，誤差項を完全に正規分布に仮定している構造化プロビットモデルの方が誤差項の展開に関しては一般性が高いことである．第二に，筆者らは鉄道経路選択モデルへ Mixed Logit Model を適用した研究[6)]は既に行っているが，誤差の分散共分散の定式化がまだ不十分であり，予測の際に新しい選択肢を考慮できないモデルとなっていた．即ち，Mixed Logit Model においても構造化プロビットモデルと同様の分散共分散の定義が可能であれば，新たな選択肢への対応も可能となる．

以上の論点を受けて，本研究では比較的選択肢間の相関が大きい東京圏の鉄道経路選択モデルを対象として，Mixed Logit Model の新たな定式化を試み，その推定特性を構造化プロビットモデルのそれと比較した．

*Keywords: 交通行動分析，経路選択
** 正会員　工修　東京工業大学助手　工学部土木工学科
***正会員　工博　東京工業大学教授　工学部土木工学科
(〒152-8552　東京都目黒区大岡山 2-12-1)
(FAX: 03-5734-3578　E-mail: sim@plan.cv.titech.ac.jp)

## 2．Mixed Logit Model の概要

始めに Mixed Logit Model の定式化を簡単に説明しておく．モデルの効用関数は以下の形で表す．

$$U_{in} = V_{in} + \left[\eta_{in} + \varepsilon_{in}\right] \tag{1}$$

ここで，$V_{in}$ はサンプル $n$，選択肢 $i$ の効用関数の確定項，$\eta_{in}$ は平均 0 の確率分布に従う誤差，$\varepsilon_{in}$ はガンベル分布に従う誤差（通常のロジットモデルの誤差項）である．$\eta_{in}$ に何らか値を与えれば，選択肢 $i$ の選択確率は以下のようになる．

$$L_{in}(\eta) = \frac{\exp(V_{in} + \eta_{in})}{\sum_i \exp(V_{in} + \eta_{in})} \tag{2}$$

今，$\eta$ の確率密度分布を $f(\eta|\Omega)$ とすると，選択肢 $i$ の選択確率は以下の積分の形式で与えられる．

$$P_{in} = \int L_{in}(\eta) f(\eta|\Omega) d\eta \tag{3}$$

式(3)では積分を計算できないため，モンテカルロシミュレーション法などを利用して近似解を算出することになる．

なお，$\eta_{in}$ については相互に独立な誤差項に分解が可能である．この場合，式(1)を以下の形に書き換える．

$$U_{in} = V_{in} + \left[\mu' \mathbf{z}_{in} + \varepsilon_{in}\right] \tag{4}$$

ここで，$\mu$ は平均 0 の分布に従う確率変数ベクトル，$\mathbf{z}_{in}$ は選択肢 $i$ に関する特性変数ベクトルである．これらの与え方を工夫すれば，選択肢間の類似性を任意に表現することができる．選択肢 $i$ と $j$ の共分散は以下のように表せる．

$$E\left(\left[\mu' \mathbf{z}_{in} + \varepsilon_{in}\right]\left[\mu' \mathbf{z}_{jn} + \varepsilon_{jn}\right]\right) = \mathbf{z}_{in}' \mathbf{W}(\mu') \mathbf{z}_{jn} \tag{5}$$

ここで，$\mathbf{W}(\mu')$ は $\mu$ の分散を対角要素に持つ行列である．即ち，特性ベクトルを構成する要因を複数考慮することが可能となる．

## 3．鉄道経路選択モデルの定式化

次に，3 肢選択を例に Mixed Logit Model を利用した鉄道経路選択モデルの定式化を考える．以下では，経路 $i$ の経路長を $L_{in}$，経路 $i,j$ の重複区間長を $L_{ijn}$ とする．清水ら(1998)[6]による定式化では，経路 1 の誤差分散に経路 2,3 との重複距離が再度カウントされることになり，理論的整合性に欠いていた．そこで，本研究では定式化を以下のように改良する．

図 1 のように各サンプルに対して選択肢集合内の全路線を多数の単位区間に分割することを考える．各区間 $k$ からは等しい誤差が生じ，区間相互で独立であると仮定する．即ち，式(4)の $\mu$ は $N(0,\omega^2)$ に従うと仮定し，各選択肢の特性ベクトル $\mathbf{z}_{in}$ は以下のように設定する．

$$\mathbf{z}_{1n}=\begin{bmatrix}1\\1\\1\\0\\0\\0\end{bmatrix}\quad \mathbf{z}_{2n}=\begin{bmatrix}1\\1\\0\\1\\0\\0\end{bmatrix}\quad \mathbf{z}_{3n}=\begin{bmatrix}1\\0\\0\\0\\1\\1\end{bmatrix} \tag{6}$$

即ち，$\mathbf{z}_{in}$ の行数は総区間数と等しく，各要素に関しては，区間 $k$ が経路 $i$ 上に存在すれば 1，そうでなければ 0 が入る．式(6)を式(4)に代入すれば，選択肢 $i$ の誤差分散，選択肢 $i,j$ の誤差共分散は以下の形に導くことができる．

$$\mathrm{var}\left(\mu'\mathbf{z}_{in}+\varepsilon_{in}\right)=L_{in}\omega^2+\frac{\pi^2}{6} \tag{7}$$

$$\mathrm{cov}\left[\left(\mu'\mathbf{z}_{in}+\varepsilon_{in}\right),\left(\mu'\mathbf{z}_{jn}+\varepsilon_{jn}\right)\right]=L_{ijn}\omega^2 \tag{8}$$

ここで，式(7)の右辺第 2 項はガンベル分布の分散である．これは後述の構造化プロビットモデルによる定式化と酷似した形となっている．即ち，単位区間の考え方を導入し，従来のモデルの有する理論整合性についての課題を解決することができた．

一方，構造化プロビットモデルについては，効用関数は以下のように表す．

$$U_{in}=V_{in}+\left[\varepsilon_{in}^1+\varepsilon_{in}^0\right] \tag{9}$$

ここで，$\varepsilon_{in}^1$ は経路 $i$ の単位長ごとに独立に発生する誤差（単位長当たり分散 $\sigma_1^2$），$\varepsilon_{in}^0$ は経路ごとに独立に発生する誤差（分散 $\sigma_0^2$）であり，共に平均 0 の正規分布である．詳しいモデルの展開は参考文献[1),2)]に譲るが，選択肢 1 の誤差項の分散，選択肢 1 と 2 の共分散は，経路長，重複区間長を用いて以下のようになる．

$$\mathrm{var}\left(\varepsilon_{1n}^1+\varepsilon_{1n}^0\right)=L_1\sigma_1^2+\sigma_0^2 \tag{10}$$

$$\mathrm{cov}\left[\left(\varepsilon_{1n}^1+\varepsilon_{1n}^0\right),\left(\varepsilon_{2n}^1+\varepsilon_{2n}^0\right)\right]=L_{12}\sigma_1^2 \tag{11}$$

以上より，Mixed Logit Model においても構造化プロビットモデルと同様の誤差項の構造化が行えた．

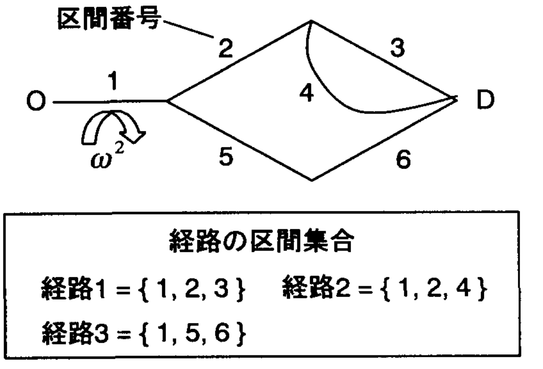


図1　3肢選択の場合の例

## 4．モデルの基本的な推定特性

3 で定式化を行った両モデルについて，平成 7 年「大都市交通センサス」データの一部を用いて，3 肢モデルを対象に推定を行った．サンプル数は選択可能経路が 3 つしかない 637 を抽出した．確定効用項の説明変数として，鉄道運賃，各所要時間，乗換回数，混雑指標を選択した．モデル推定結果を表 1 に示す．推定に際して，構造化プロビットモデルはシミュレーション法の一種である GHK 法[2)]，Mixed Logit Model はモンテカルロシミュレーション法を用いている．GHK 法では各に対して発生させる乱数は(選択肢数-2）となる．一方，モンテカルロシミュレーション法では選択肢数が必要であり，計算速度の観点からは GHK 法の方が有利であることが予想される．各モデルの推定パラメータの変動を比較するため，モデル推定を異なる乱数セットを用いて 10 回ずつ行っている．また，推定において尤度関数に与える選択確率は 10 パターンの異なる乱数セットを与えたときの平均確率としている．なお，表 1 は 10 回の推定結果の中から平均的なものを示している．ここで，分散比とは構造化プロビットモデルが $\sigma_1^2/\sigma_0^2$，Mixed Logit Model が $6\omega^2/\pi^2$ であり，これが大きいほど，選択間の相関をより表現したモデルとなる．従来の研究[7)]から分散比は選択肢数の増加，サンプルの経路重複度合いの増加するほど小さくなることが分かっている．

表 1 から分かる通り，8 つの確定項のパラメータ値は両モデルでほとんど差がない結果となっている．分散比パラメータについては，表 1 では Mixed Logit Model の方が大きいが，10 回の平均ではほぼ同じ値であった．これらパラメータの変動は図 2 に示す通りである．図の縦軸である変動係数は，10 回の推定パラメータの平均値 $\mu$ と標準偏差 $\sigma$ の比 $\sigma/\mu$ であり，これが大きければパラメータの安定性が低いということができる．また，確定項パラメータに関しては，推定ごとの変動は大きくても 10%程度であり，両モデルでほとんど差がない．従来

表1 パラメータの推定結果の比較

| | Probit | | Mixed Logit | |
|---|---|---|---|---|
| | パラメータ | $t$ 値 | パラメータ | $t$ 値 |
| 運賃(円) | -0.00184 | -3.47 | -0.00255 | -3.81 |
| アクセス時間(分) | -0.107 | -7.97 | -0.136 | -9.69 |
| イグレス時間(分) | -0.0755 | -6.16 | -0.0962 | -7.03 |
| 乗車時間(分) | -0.0112 | -1.01 | -0.0114 | -0.866 |
| 乗換時間(分) | -0.0284 | -1.85 | -0.0327 | -1.75 |
| 待ち時間(分) | -0.123 | -4.30 | -0.171 | -4.94 |
| 乗換回数 | -0.194 | -1.35 | -0.274 | -1.51 |
| 混雑指標(*) | -0.00726 | -2.24 | -0.00892 | -2.36 |
| 分散比 | 0.0123 | 0.739 | 0.0264 | 1.38 |
| 尤度比 | 0.262 | | 0.272 | |

表2 推定に要した時間の比較

| | 平均(分) | 標準偏差(分) |
|---|---|---|
| Probit | 2.12 | 0.234 |
| Mixed Logit | 8.55 | 2.77 |

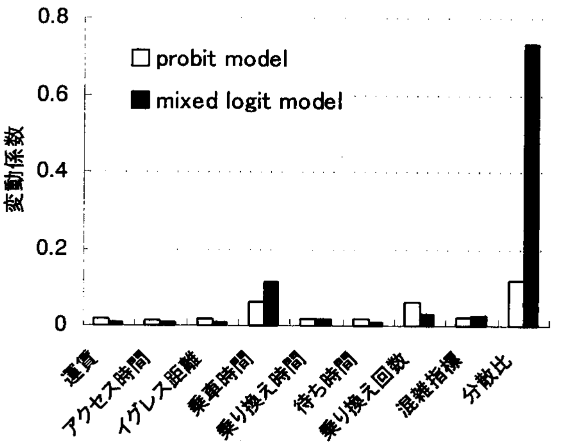


図2 パラメータの変動係数の比較

の研究[6]では，Mixed Logit Model の変動係数の方がプロビットモデルのそれと比べて 3 倍程度であったが，定式化の改良により，より安定的な推定が行えていることが分かる．しかし，分散比の変動係数については Mixed Logit Model の方が非常に大きい．この性質については次章で詳細に分析する．

表 2 は両モデルの推定時間を比較したものである．使用したプログラムは GAUSS(Ver.3)であり，同一のコンピュータ(CPU は PentiumⅡ-300MHz)で計算を行った．先にも述べた通り，推定が不安定であったこと，発生させる乱数の数が多いことなどの理由で，Mixed Logit Model の方が時間がかかり，収束時間が一定でないことが示されている．

## 5．分散比の推定特性

分散比は，選択肢独自の誤差分散と他の選択肢から影響を受ける誤差分散の比であり，これが大きいほど選択肢間の相関を強く表現できることになる．図 3 では，3 肢選択を例にして，選択経路と代替経路の重複関係により推定される分散比の大きさが異なることを示している．もし，選択経路が代替経路との重複が小さければ，推定アルゴリズムは選択経路の選択確率を大きくするために分散比を大きくし代替経路の効用を低減させるように働く．逆に言えば，選択経路が代替経路と重複していないサンプルを用いて推定を行えば，大きい分散比を得ることができる．4 では Mixed Logit Model の分散比の安定性が非常に悪かったが，以下では重複の状況により推定サンプルを取捨選択し，推定を試みることで，Mixed Logit Model が安定的に推定できる条件を簡単に検討する．

先の全 637 サンプルは，各経路間の重複関係を全体的にカバーしており，この中から経路 1（選択経路）と経路 2,3（代替経路）の重複率が 20%以下のサンプルを抽出したもの（ケース A），50%以下のサンプルを抽出したものの（ケース B），80%以下のサンプルを抽出したもの（ケース C）の 3 種類のサブサンプルを設定した．重複率とは，経路 $i,j$ の重複区間長を経路 $i$ 及び $j$ の区間長の平方根で割った値（区間[0,1]で定義）で，0 であれば両経路は重複しておらず，1 であれば完全に重複している．この時，全サブサンプルで，推定される分散比は全サンプルのそれよりも大きくなると考えられる．

表 3 に各ケースにおける分散比の変動係数を示す．なお，分散比の比較にのみ着目したため，説明変数は運賃と所要時間の 2 つとし，各ケースで 10 回の推定を行い，4 と同じ手順で変動係数を算出した．先の仮定に従えば，ケース A,B,C の順で分散比パラメータが大きくなり，かつ安定性も向上すると考えられるが，ケース A では変動係数が大きくなっている．この原因としてサンプル数の減少が考えられるが，その確認のために 637 サンプルから各ケースと同じサンプル数のサブサンプルを抽出して同様に推定を行った（表 3 で各ケースの括弧内の数値）．サンプル数の減少とともに推定が不安定となっており，特に 195 サンプルの場合には，重複度合いの減少により変動係数が約半分となっていることが見て取れる．

以上の分析から，重複度合いの高いサンプルが少ないような条件設定においては Mixed Logit Model でも安定的な推定が可能であることが示せた．

## 6．おわりに

本研究では，鉄道経路選択を例に，誤差項を構造化し選択肢間の類似性を定義できる 2 つのモデルの定式化を行ったが，Mixed Logit Model については，1 で述べた第二の課題を解決することができた．さらにトリップデータを用いてモデルの推定特性を把握した．Mixed Logit Model は分散比の変動係数が大きく推定が不安定であるが，サンプルの構成によってはこれを小さくすることが可能であることが示された．ただし，今後はモンテカル

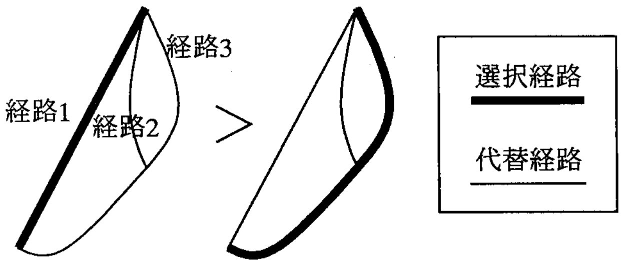


図3 経路の重複関係による分散比の大小関係

表3 重複度合いによる分散比の変動係数

| ケース | 分散比の変動係数 | サンプル数 |
|---|---|---|
| A | 0.373 (0.801) | 195 |
| B | 0.173 (0.291) | 311 |
| C | 0.235 (0.403) | 505 |
| 全サンプル | 0.286 | 637 |

ロシミュレーション法に替わる新たな推定アルゴリズムの開発により，推定の安定性が向上する可能性がある．

今後の課題としては，選択肢集合の設定方法や選択肢数の増加によってどのような推定特性が得られるか，様々な検討が必要となる．プロビットモデルの場合についても言えることであるが，IIA 特性のないモデルであるため，選択肢集合の設定については注意が必要である．鉄道経路選択行動において，経路間の類似性が非常に高い，即ちほとんど重複しているケースでは，利用者が両者を別の経路と考慮している可能性は非常に小さいと考えられ，どちらかの経路を選択肢集合から外すべきである．その論拠は５の分析結果からも想定することが可能であり，より望ましい選択肢集合の設定により推定の安定性をも向上させることができると考えられる．

謝辞：本研究の遂行に当たり，小林亜紀子氏（東京工業大学大学院修士課程）の協力を得た．記して謝意を表する次第である．

（参考文献）

1) Yai, T., Iwakura, S. and Morichi, S.: Multinomial probit model with structured covariance for route choice behavior, Transportation Research B, Vol.31, No.3, pp.195-207, 1997.

2) 屋井鉄雄，中川隆広，石塚順一：シミュレーション法による構造化プロビットモデルの推定特性, 土木学会論文集, No.694, pp.11-21, 1998.

3) Yai, T. and Shimizu, T.: Multinomial probit with structured covariance for several choice situations with similar alternatives, Transportation Research Record, 1645, pp.69-75, 1998.

4) Brownstone, D., Bunch, D. and Train, K.: Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-fuel Vehicles, IATBR, 1997.

5) Bhat, C.: Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice, Transportation Research B, Vol.32, No.7, pp.455-466, 1998.

6) 清水哲夫，屋井鉄雄，坂井康一：鉄道経路選択モデルにおける選択肢間の類似性の表現方法，土木計画学研究・講演集, No.21(1), pp.459-460, 1998.

7) Yai, T. and Iwakura, S.: An application of multinomial probit model with structured covariance, Journal of the Eastern Asia Society for Transportation Studies, Vol.2, No.4, pp.1163-1176, 1997.

## Mixed Logit Model とプロビットモデルの推定特性に関する比較分析

清水　哲夫，屋井　鉄雄

本研究は，ロジットモデルの IIA 特性を緩和する２つのモデル，Mixed Logit Model と構造化プロビットモデルについて，鉄道経路選択モデルを対象に推定特性を比較分析した．Mixed Logit Model は従来から定式化の制約があったものの，これ解決する新たな定式化が可能となり，理論的整合性のより高いモデルとなった．3 肢選択を対象とした両モデルによる推定では，構造化プロビットモデルの方が推定の安定性，所要時間とも優れている結果となったが，選択肢集合の設定方法など新たな検討課題も明らかとなった．

## A Comparative Study on the Estimation Characteristics of Mixed Logit Model and Probit Model

Tetsuo Shimizu and Tetsuo YAI

The aim of this study is to compare the estimation characteristics of mixed logit model and probit model with structured covariance for railway route choice. We improve the formulation for mixed logit model in order to be more theoretical formula. We analyze the case of 3 alternatives and it is suggested that probit model is prior to get more stable parameters rather than mixed logit model. In the further study, method for setting choice set may be obviously important issue.