多項型サンプルセレクションモデルによる居住地-交通行動選択モデリングとベイズ推定アルゴリズムの提案

渡邉　萌[1]・円山　琢也[2]

[1]学生会員　熊本大学　大学院自然科学教育部（〒860-8555 熊本市中央区黒髪2-39-1）
E-mail: 197d9225@st.kumamoto-u.ac.jp

[2]正会員　熊本大学准教授　くまもと水循環・減災研究教育センター（〒860-8555 熊本市中央区黒髪2-39-1）
E-mail: takumaru@kumamoto-u.ac.jp

人々の居住地選択において，あらかじめ利用する交通手段を想定した上でその交通手段を利用しやすい居住地を選択する場合が考えられる．サンプルセレクションモデルはそのような居住地選択－交通行動間の依存関係を明示的に記述することができる．しかし，既存のサンプルセレクションモデルが扱える居住地選択は都市部・郊外部のような2項選択のみであり，現実の居住地選択を記述するには不十分であった．本研究では，既存のサンプルセレクションモデルを拡張し，複数の選択肢から一つの居住地を選択する多項選択を記述可能なモデルを提案する．提案モデルは多項プロビットモデルによる居住地選択問題と2値または連続モデルによる交通行動選択問題により構成され，大規模なオープン・フォーム構造を有している．そのため，尤度関数を直接評価してパラメータを推定することが困難である．そこで本研究ではベイズ推定を採用し，効率的なMCMCアルゴリズムをあわせて提案する．

Key Words : *residential location choice, sample selection models, Markov chain Monte Carlo, discrete-continuous models*

1.　はじめに

望ましい都市開発に向けて，土地利用と交通の連携を強調するニューアーバニズムやコンパクトシティといった都市の計画・デザインのコンセプトが長年にわたり提唱されている．交通路や施設等の都市における建造環境の密度，多様性，デザインは人々の交通行動に影響を与えることが考えられるが，それらは同時に居住地を選択する際の決定要因にもなりうる．そのため，土地利用計画と交通計画の望ましい連携のための政策提言に向けて，都市開発に伴う建築環境の変化が居住地選択と交通行動に与える影響をそれぞれ適切に把握する必要がある[1]．

居住地選択と交通行動には依存関係が存在しうることが指摘されている[2]．これは，人々が居住地を選択する際に，あらかじめ利用する交通手段や移動そのものに対する選好を考慮する場合が考えられるためである．具体例として，自動車を保有していない人々が公共交通インフラの整備水準が高い地域に居住する傾向や，長距離の運転が趣味の人々が郊外部に居住する傾向がある場合などが挙げられる．このとき，居住地選択と交通行動は独立ではないため，両者の間には依存関係があるとされる．しかし，交通調査等によりこの依存関係の存在が正しく把握されることは稀である．この依存関係を無視してしまうと分析結果にバイアスが生じてしまい，建築環境の変化が居住地選択と交通行動に与える影響を正しく捉えることができない．この場合に生じるバイアスは，居住地における自己選択バイアスとも呼ばれる (residential self-selection bias)[3]．誤った分析結果に基づく政策提言を避けるためにも，バイアスの補正は重要な課題となる．

上述の依存関係により生じるバイアスの対処法の一つとして，サンプルセレクションモデルが挙げられる[4]．Heckman (1974, 1979)[5),6)] や Heckman et al. (2001)[7)] により提案されたサンプルセレクションモデルは，離散-連続モデルとも呼ばれ，離散的な選択問題と連続量の選択問題との相関関係を誤差相関により記述するモデルである．Zhou and Kockelman (2008)[8)] や Cao (2009)[9)] らの研究では，このサンプルセレクションモデルの誤差相関により，居住地選択と交通行動との間の依存関係を記述することで，

上述したバイアスを補正している．しかし，それらの既存研究で使用されているサンプルセレクションモデルのフレームワークでは，都市部と郊外部のどちらかを選択するような，単純な2項選択の居住地選択問題しか取り扱うことができないという課題がある．現実の人々の居住地選択をより詳細に記述するためには，複数 (3選択肢以上) の選択肢からの選択を記述できる，より一般的なフレームワークが求められる．

本稿では，多項選択による居住地選択と交通行動との依存関係を記述できる新しいサンプルセレクションモデルを提案する．具体例には，選択肢相関を考慮できる多項プロビットモデルにより居住地選択問題を記述し，交通行動は2項プロビットモデルまたは線形回帰モデルにより記述する．そして，この居住地選択モデルと交通行動モデルの依存関係を，誤差相関により記述する．

提案モデルと既存のモデルとの違いは，Heckman et al. (2001) や渡邉・円山 (2021) [10]のモデルが扱うことができる居住地選択は2項選択のみであるのに対し，提案モデルは多項選択の居住地選択を扱うことができる点にある．

提案モデルはオープン・フォームでかつ大規模な尤度関数を有する．しかし，全体構造は多変量正規分布によって記述されるため，居住地選択モデルと，居住地選択を条件付けた交通行動モデルの2つに確率分布を分割することが可能である．この性質は，ベイズ推定の利点である，パラメータの逐次的な推定と非常に相性が良い．そこで本研究では，この点を踏まえたMarkov chain Monte Carlo (MCMC) アルゴリズムもあわせて提案する．

2. 居住地選択-交通行動モデル

居住地域の建造環境が交通行動に与える影響を明らかにした先行研究は多く[11]，そのための方法論的検討も行われている[3]．しかし，一般にそれらの研究における居住地はあくまで交通行動を説明する変数として扱われており，居住地選択そのものをモデリングの対象として含めている先行研究は非常に限られている．人々の居住地移転が交通需要に直接的に関係することを考慮すると，交通行動とあわせて居住地選択自体を記述する意義は大きいと考えられる．

Heckman et al. (2001) のサンプルセレクションモデルは，交通行動と居住地選択両方のモデリングを行う．例えば，Zhou and Kockleman (2008) らの研究では，都市部から郊外部に移住した場合，世帯における一日の自動車走行距離に約27km以上増加すること示した．しかし，既存のサンプルセレクションモデルの限界として，非常に単純化された居住地選択しか扱えないという点が挙げられる．したがって，提案モデルのように，より一般的な居住地選択を扱うことができるモデルが必要である．

3. 多項型サンプルセレクションモデル

本章ではまず， Heckman et al. (2001) のサンプルセレクションモデルを用いた，居住地選択と交通行動の同時選択を記述する既存のモデリングに関する説明を行う．その後，本稿で提案する多項型のサンプルセレクションモデルが，既存のモデリングをより一般的なモデルに拡張したモデルであるという点を説明する．

(1) 既存モデルにおける内生的スイッチング構造

Heckman et al. (2001) のサンプルセレクションモデルの最も特徴的な点として，内生的スイッチング構造が挙げられる．居住地選択と交通行動のモデリングの文脈における内生的スイッチング構造とは，観測される交通行動が居住地選択によって内生的に決定，すなわちモデル内部で交通行動の観測と非観測が引き起こされる構造のことを指す．以下では，サンプルセレクションモデルとその内生的スイッチング構造について説明する．

サンプルセレクションモデルを用いて居住地選択と交通行動の同時選択を記述する場合，一般的に居住地選択は2項プロビットモデルにより記述される．例えば，ある個人 i が居住地の選択肢 J (ここでは $J=2$) から一つを選択する際の潜在効用を y_i^*，選択結果を y_i とすると，個人 $i=1,\ldots,n$ の居住地選択は以下のように記述できる．

$$y_i^* = x_i'\beta + \varepsilon_i, y_i = \mathrm{I}\{y_i^* > 0\}, \tag{1}$$

x_i は説明変数，β は説明変数に対応するパラメータ，ε_i は誤差項であり，プロビットモデルの場合は正規分布が仮定される．$\mathrm{I}\{\cdot\}$ は指示関数であり，$\{\cdot\}$ が真の場合に1，それ以外の場合に0を取る．

次に交通行動を記述するモデルの説明を行う．ここでは，自動車走行距離や歩行距離といった連続的な交通行動を想定する．そのため交通行動を記述するモデルとして線形回帰モデルを想定する．上述したように，内生的スイッチング構造により居住地選択と交通行動の関係性を記述するためには，候補となる居住地別の交通行動を想定する必要がある．式 (1) よりここで想定している居住地選択は2項選択であるため，2つの居住地にあわせて区別された個人 i の交通行動 z_{i1}，z_{i2} に対しそれぞれ線形回帰モデルを以下のように仮定する．

$$z_{i1} = \begin{cases} w_{i1}'\alpha_1 + \xi_{i1} & \text{if } y_i = 1 \\ \text{missing} & \text{if } y_i = 0, \end{cases} \tag{2}$$

$$z_{i2} = \begin{cases} \text{missing} & \text{if } y_i = 1 \\ w_{i2}'\alpha_2 + \xi_{i2} & \text{if } y_i = 0, \end{cases} \tag{3}$$

w_{i1}', w_{i2}' は説明変数，α_1, α_2 はそれぞれ説明変数に対応するパラメータ，ξ_{i1}, ξ_{i2} は誤差項であり，ここでは正規分布を仮定する．式 (2), (3) より，z_{i1} は居住地が $y_i = 1$ の場合に観測される交通行動であり，z_{i2} は $y_i = 0$ の

場合にのみ観測される．したがって，居住地の選択肢を都市部($y_i = 1$)と郊外部($y_i = 0$)と仮定した場合，z_{i1} は都市部における交通行動，z_{i2} は郊外部における交通行動と解釈される．都市部 (郊外部) における交通行動は，当然ながら対象者が都市部 (郊外部) に居住していなければ観測できない．

このモデルにおける居住地選択と交通行動の依存関係は，誤差の相関により記述される．したがって居住地選択モデルと交通行動モデルの誤差項は，以下に示す分散共分散行列を持つ三変量正規分布に従うと仮定する．

$$\begin{pmatrix}\varepsilon_i\\ \xi_{i1}\\ \xi_{i2}\end{pmatrix}\sim N\left[\begin{pmatrix}0\\0\\0\end{pmatrix},\begin{pmatrix}1 & \sigma_{12} & \sigma_{13}\\ \sigma_{12} & \sigma_2 & 0\\ \sigma_{13} & 0 & \sigma_3\end{pmatrix}\right]. \tag{4}$$

σ_2 と σ_3 はそれぞれ ξ_{i1} と ξ_{i2} の分散，また σ_{12}, σ_{13} は共分散である．2項プロビットの誤差項 ε_i の分散は通常1に固定される．式 (2), (3) に示すように，同じ個人においてて z_{i1} と z_{i2} の両方の交通行動が同時に観測されることはないため，ξ_{i1} と ξ_{i2} の誤差相関は通常0に固定されることが多い．したがって，全体のモデル構造は以下となる．

$$\begin{pmatrix}y_i^*\\ z_{i1}\\ z_{i2}\end{pmatrix}\sim N\left[\begin{pmatrix}x_i'\beta\\ w_{i1}'\alpha_1\\ w_{i2}'\alpha_2\end{pmatrix},\begin{pmatrix}1 & \sigma_{12} & \sigma_{13}\\ \sigma_{12} & \sigma_2 & 0\\ \sigma_{13} & 0 & \sigma_3\end{pmatrix}\right]. \tag{5}$$

これより，Heckman et al. (2001) のサンプルセレクションモデルにおいて最も特徴的な点である，内生的スイッチング構造について説明する．内生的スイッチング構造は，将来の交通行動を所与とした場合の居住地選択確率 $P(y_i|z_{i1}, z_{i2})$ を想定すると説明が容易となる．これは，公共交通機関が発達している都市部とそうでない郊外部から一つを選択し移住するとき，仮に移住した場合に将来どのような交通行動を取るかを想定し，居住地を選択する場合などが例として挙げられる．より具体的には，車を保有していない人は公共交通機関へのアクセスを考慮する傾向にあるため，都市部を選択する確率が高い，といったように交通行動を条件付けた場合の居住地選択確率を想定する．ここで，$z_{i1} \neq z_{i1}'$ の場合の都市部 ($y_i = 1$) を選択する確率を考えると，

$$P(y_i = 1|z_{i1}, z_{i2}) \neq P(y_i = 1|z_{i1}', z_{i2}), \tag{6}$$

となる．これは，式 (5) に示すように居住地選択モデルの潜在効用 y_i^* が σ_{12} を通して交通行動 z_{i1} と依存関係を持っているためである．同様に，$z_{i2} \neq z_{i2}'$ の場合の都市部 ($y_i = 1$) を選択する確率は

$$P(y_i = 1|z_{i1}, z_{i2}) \neq P(y_i = 1|z_{i1}, z_{i2}'), \tag{7}$$

となる．よって式 (6), (7) は，「居住地選択，すなわち将来的に交通行動が観測されるかどうかに，将来の交通行動そのものが影響を与える」ということを示している．端的に言えば，内生的スイッチング構造とは一般に上記の点を明示的に記述する構造のことを指す．

(2) 内生的スイッチング構造の多項型への拡張

前節では，居住地選択問題が2項選択の場合におけるサンプルセレクションモデルについて説明した．本節では，以上のサンプルセレクションモデルを，内生的スイッチング構造を保持したまま，多項型の居住地選択を扱うことができるモデルへ拡張する．

ある個人 i の居住地選択 $y_i \in \{1, 2, \dots, J\}$ が $j \in J$ の場合の潜在効用 $y_i^* = (y_{i1}^*, y_{i2}^*, \dots, y_{iJ}^*)'$ は以下の通り定義する．

$$y_{ij}^* = x_{ij}'\beta_j + \varepsilon_{ij}, \tag{8}$$

x_{ij} は説明変数，β_j は説明変数に対応するパラメータ，ε_{ij} は誤差項であり，以下の通り相関行列 Σ_y の多変量正規分布に従うと仮定する．

$$\begin{pmatrix}\varepsilon_{i1}\\ \varepsilon_{i2}\\ \vdots\\ \varepsilon_{iJ-1}\\ \varepsilon_{iJ}\end{pmatrix}\sim N\left[\begin{pmatrix}0\\0\\ \vdots\\0\\0\end{pmatrix},\begin{pmatrix}1 & \gamma_{1,2} & \cdots & \gamma_{1,J-1} & \gamma_{1,J}\\ \gamma_{1,2} & 1 & \cdots & \gamma_{2,J-1} & \gamma_{2,J}\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \gamma_{1,J-1} & \gamma_{2,J-1} & \cdots & 1 & \gamma_{J-1,J}\\ \gamma_{1,J} & \gamma_{2,J} & \cdots & \gamma_{J-1,J} & 1\end{pmatrix}\right]. \tag{9}$$

式 (9) の γ は相関係数であり，居住地選択における選択肢相関を記述するパラメータである．また以下に示すように，潜在効用 y_{ij}^* が最大の時，居住地 j を選択すると仮定する．

$$y_i = j \text{ if } \max(y_i^*) = y_{ij}^*, \tag{10}$$

すなわち $\forall q \neq j, \forall q \in J$ の場合，居住地 j を選択する確率 $P(y_i = j)$ は,

$$P(y_i = j) = P\left[x_{ij}'\beta_j + \varepsilon_{ij} > x_{iq}'\beta_q + \varepsilon_{iq}\right], \tag{11}$$

となり，提案モデルにおける居住地選択問題は，誤差相関を考慮した多項プロビットモデルにより記述される．

次に，提案モデルにおける交通行動選択問題について説明する．提案モデルは2値または連続的な交通行動を想定しており，それぞれ以下の通り2項プロビットモデルと線形回帰モデルにより記述される．

$$z_{ij}^* = w_{ij}'\alpha_j + \xi_{ij}, z_{ij} = \mathrm{I}\{z_{ij}^* > 0\}, \tag{12}$$

$$z_{ij} = w_{ij}'\alpha_j + \xi_{ij}, \tag{13}$$

z_{ij} は居住地 j における交通行動であり，2値の場合は潜在効用 z_{ij}^* を仮定する．w_{ij} は説明変数，α_j は説明変数に対応するパラメータ，ξ_{ij} は誤差項である．また，前節で説明したサンプルセレクションモデルと同様に，居住地 j における交通行動 z_{ij} は，対象者の居住地選択 y_i によって観測されるかどうかが決定されると仮定する．すなわち，

$$z_{ij} = \begin{cases}\text{observed} & \text{if } y_i = j\\ \text{missing} & \text{if } y_i \neq j\end{cases}, j \in J. \tag{14}$$

提案モデルは，前節で説明したサンプルセレクションモデルと同様に，誤差相関により居住地選択と交通行動の依存関係を記述する．$\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iJ})'$．$\xi_i = (\xi_{i1}, \xi_{i2}, \dots, \xi_{iJ})'$ とするとき，居住地選択と交通行動における誤差項である ε_i と ξ_i は以下の多変量正規分布に従

う．

$$\begin{pmatrix}\varepsilon_i\\ \xi_i\end{pmatrix} = N_{2J}\left[\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}, \begin{pmatrix}\Sigma_y & \Sigma_{y,z}\\ \Sigma_{y,z}^{\mathrm{T}} & \Sigma_z\end{pmatrix}\right],$$
$$\text{where } \Sigma_z = \mathrm{diag}\left(\Sigma_{z_1}, \Sigma_{z_2}, \dots, \Sigma_{z_J}\right), \text{and} \tag{15}$$
$$\Sigma_{y,z} = \left(\Sigma_{y,z_1}, \Sigma_{y,z_2}, \dots, \Sigma_{y,z_J}\right)' = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_J),$$

Σ_y は式 (9) 内に示す次元 J の相関行列である．また計算効率のため，誤差項 ξ_{ij} の分散 Σ_{z_j} は交通行動が2項または連続の場合でそれぞれ以下のように仮定する．

$$\Sigma_{z_j} = \begin{cases} 1 + \Sigma_{y,z_j}^{\mathrm{T}} \Sigma_y^{-1} \Sigma_{y,z_j} & \text{if } z_j \text{ is binary} \\ \nu_j^2 + \Sigma_{y,z_j}^{\mathrm{T}} \Sigma_y^{-1} \Sigma_{y,z_j} & \text{if } z_j \text{ is continuous}\cdot \end{cases} \tag{16}$$

これにより，誤差項 ε_{ij} を条件付けた場合の誤差項 ξ_{ij} の分散がそれぞれ1と ν_j^2 となり，次章で説明する推定アルゴリズムが簡便なものになる．

以上より，交通行動が2項または連続の場合の提案モデルの全体的な構造は以下の通りとなる．

$$\begin{pmatrix}y_i^*\\ z_i^*\end{pmatrix} \text{ or } \begin{pmatrix}y_i^*\\ z_i\end{pmatrix} \sim N_{2J}\left[\begin{pmatrix}X_i\beta\\ W_i\alpha\end{pmatrix}, \begin{pmatrix}\Sigma_y & \Sigma_{y,z}\\ \Sigma_{y,z}^{\mathrm{T}} & \Sigma_z\end{pmatrix}\right], \tag{17}$$

このとき，$z_i^* = (z_{i1}^*, z_{i2}^*, \dots, z_{iJ}^*)', z_i = (z_{i1}, z_{i2}, \dots, z_{iJ})', X_i = \mathrm{diag}(x_{i1}', x_{i2}', \dots, x_{iJ}')$ ， $W_i = \mathrm{diag}(w_{i1}', w_{i2}', \dots, w_{iJ}')$ ， $\beta' = (\beta_1', \beta_2', \dots, \beta_J'), \alpha' = (\alpha_1', \alpha_2', \dots, \alpha_J')$．したがって，式 (16-17) より，提案モデルは居住地選択における居住地 j の潜在効用 y_{ij}^* と居住地 j における交通行動 z_{ij} との依存関係を，共分散パラメータ σ_j により記述する．

上記の誤差相関により，提案モデルにも内生的スイッチング構造が生じる．前節と同様に，将来の交通行動を所与とした場合の居住地選択確率を考えると，$z_{ij} \neq z_{ij}'$ の場合は

$$P(y_i = j | z_{i1}, \dots, z_{ij}, \dots, z_{iJ}) \neq P(y_i = j | z_{i1}, \dots, z_{ij}', \dots, z_{iJ}), \tag{18}$$

が成り立つ．さらに，例えば $z_{i1} \neq z_{i1}', j \neq 1$ の場合，

$$P(y_i = j | z_{i1}, \dots, z_{ij}, \dots, z_{iJ}) \neq P(y_i = j | z_{i1}', \dots, z_{ij}, \dots, z_{iJ}), \tag{19}$$

となる．式 (6), (7) と式 (18), (19) を比較すると明らかなように，提案モデルは前節で説明したサンプルセレクションモデルをより一般的な内生的スイッチング構造を持つモデルへと拡張したものである．

最後に，提案モデルの実用的な面について述べる．最も考えられうる適用例は，因果推論への適用である．提案モデルにより，渡邉・円山 (2021)，Zhou and Kockleman (2008) やCao (2009) らの分析と同様に，居住地の違いが交通行動にもたらす平均因果効果を計算ことができる．提案モデルは多項選択による居住地選択を扱うことができ，既存のモデリングでは不可能であった分析が可能となる．また，内生的スイッチング構造は様々な場面で生じていると解釈できるため，居住地選択と交通行動のモデリング以外にも適用が期待される．例えば，活動場所選択と活動時間 (または2値の活動有無) に生じる内生的スイッチング構造のモデリングなど，アクティビティモデルとして適用することも可能である．

4. ベイズ推定

これまで説明したように，サンプルセレクションモデルやその拡張である提案モデルは，居住地選択モデルにおける潜在効用と交通行動 (2値の場合は潜在効用) が多変量正規分布に従うと仮定されている．多変量正規分布が持つ高い操作性は，ベイズ推定の効率化において重要な役割を果たす．そこで本章では，はじめに多変量正規分布の基本的な性質について説明する．その後，その性質が提案モデルの推定に果たす役割と，MCMCによる推定アルゴリズムを説明する．

(1) 多変量正規分布における周辺分布と条件付き分布

ベクトル x が以下の2変量正規分布に従うと仮定する．

$$x = \begin{pmatrix}x_1\\ x_2\end{pmatrix} \sim N\left[\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right]. \tag{20}$$

ここで $\Sigma_{12} = \Sigma_{21} \neq 0$ を仮定すると，x_1 と x_2 は独立ではなくなる．同時分布 $f(x_1, x_2)$ を x_2 について積分して得られる x_1 の周辺分布 $f(x_1)$ は,

$$f(x_1) = \int f(x_1, x_2) dx_2 = N(\mu_1, \Sigma_{11}), \tag{21}$$

となり，1次元の正規分布で表現される．また，x_1 を固定した場合の x_2 の分布，すなわち x_1 を条件付けた場合の x_2 の条件付き分布は

$$f(x_2|x_1) = N[\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}], \tag{22}$$

となり，同じく1次元の正規分布となる．

このとき，確率の乗法定理より

$$f(x_1, x_2) = f(x_2|x_1) f(x_1), \tag{23}$$

が成り立つ．したがって，2次元の同時分布は一次元の条件付き分布と周辺分布を掛け合わせたものに等しい．

重要なのは，多次元の確率分布を少ない次元の確率分布の積に分解できる点である．これは多変量正規分布における一般的な性質であり，x_1 と x_2 がそれぞれ複数の次元を持つベクトルの場合でも成り立つ．例えば，同時分布 $f(x_1, x_2)$ が5次元と仮定すると，$f(x_1)$ を1次元とすれば条件付き分布 $f(x_2|x_1)$ は4次元となり，この場合にも式 (23) が成り立つ．したがって，左辺の次元数と右辺の合計した次元数が一致するように，複数の正規分布に分割できる．これは多変量正規分布の基本的な性質であり，ベイズ推定の効率化に大きな役割を果たす．

(2) 提案する推定アルゴリズム

サンプルの居住地選択のベクトルを $\boldsymbol{y}$，交通行動のベクトルを $\boldsymbol{z}$ とし，θ をモデルのパラメータベクトルとする．パラメータの事後分布 $f(\theta|\boldsymbol{y}, \boldsymbol{z})$ は，ベイズの定理より以下のように求められる．

$$f(\theta|\boldsymbol{y}, \boldsymbol{z}) \propto f(\boldsymbol{y}, \boldsymbol{z}|\theta) f(\theta), \tag{24}$$

$f(\theta)$ はパラメータの事前分布，$f(\boldsymbol{y}, \boldsymbol{z}|\theta)$ は尤度である．

推定の際に尤度を必要とする (本稿では尤度を評価すると表現する) という点では，ベイズ推定も最尤推定も同じである．

一般的に，推定手法に関係なく，複雑なモデルのパラメータ推定の際にモデル全体の大規模な尤度が必ずしも必要なわけではない．提案モデルのパラメータ推定を例に挙げると，居住地選択モデルにおけるパラメータ β,γ を推定する際には，交通行動モデルの尤度は必要ない．これは，パラメータ β,γ は交通行動モデルとは完全に独立しているためである．同様に，交通行動モデルにおけるパラメータ α,ν^2 の推定の際には居住地選択モデルの尤度は必要ない．ただし，問題なのは σ のように居住地選択モデルと交通行動選択モデル両方に影響を及ぼすパラメータの推定である．このようなパラメータを推定する必要があるとき，前節で説明した多変量正規分布の性質とベイズ推定を組み合わせることで，効率的なベイズ推定アルゴリズムの構築が可能となる．

ここで，式 (22) に示した多変量正規分布の性質を利用して，式 (16), (17) より居住地選択を条件付けた交通行動の選択確率は，2値の交通行動の場合

$$P\left(z_{ij}=1\middle|y_i^*\right)=\Phi\left\{w_{ij}'\alpha_j+\Sigma_{y,z_j}^{\mathrm{T}}\Sigma_y^{-1}(y_i^*-X_i\beta)\right\},\qquad(25)$$

交通行動が連続の場合は

$$P\left(z_{ij}\middle|y_i^*\right)=\frac{1}{\nu_j}\phi\left\{\frac{z_{ij}-w_{ij}'\alpha_j-\Sigma_{y,z_j}^{\mathrm{T}}\Sigma_y^{-1}(y_i^*-X_i\beta)}{\nu_j}\right\},\qquad(26)$$

というように，それぞれ標準的な2項プロビットモデルと，分散 ν_j^2 の線形回帰モデルとなる．また，$\Sigma_{y,z_j}^{\mathrm{T}}=\sigma_j$ である．重要なのは，σ_j は説明変数 $\Sigma_y^{-1}(y_i^*-X_i\beta)$ に対応するパラメータとみなすことができる点である．したがって，居住地選択を条件付けた交通行動選択モデルは，パラメータ α,σ (交通行動が連続の場合は ν^2 も) を持つ関数である．

前節で説明した内容を応用すると，提案モデルにおける選択確率の同時分布は，多項プロビットモデルによる居住地選択確率の分布と，2項プロビットモデルまたは線形回帰モデルによる，居住地選択を条件付けた交通行動の選択確率分布に分割することができる．すなわち

$$f(\boldsymbol{y},\boldsymbol{z}|\theta)=f(\boldsymbol{z}|\boldsymbol{y},\alpha,\nu^2,\sigma)f(\boldsymbol{y}|\beta,\gamma).\qquad(27)$$

加えて，式 (16) で示した誤差の仮定により，式 (25), (26) に示す居住地選択を条件付けた交通行動の選択確率分布は非常にシンプルな構造となる．これにより，σ_j (for all $j\in J$) の推定が非常に効率的になる．

上述したように，パラメータ β,γ の推定の際には，

$$f(\beta,\gamma|\boldsymbol{y})\propto f(\boldsymbol{y}|\beta,\gamma)f(\beta,\gamma),\qquad(28)$$

に示すように居住地選択モデルの尤度 $f(\boldsymbol{y}|\beta,\gamma)$ だけを評価すればよい．$f(\beta,\gamma), f(\beta,\gamma|\boldsymbol{y})$ はそれぞれパラメータ β,γ の事前分布，事後分布である．パラメータ α,ν^2,σ の事後分布 $f(\alpha,\nu^2,\sigma|\boldsymbol{y},\boldsymbol{z})$ は，居住地選択を条件付けた交通行動モデルの尤度 $f(\boldsymbol{z}|\boldsymbol{y},\alpha,\nu^2,\sigma)$ を評価することにより，以下のように得られる．

$$f(\alpha,\nu^2,\sigma|\boldsymbol{y},\boldsymbol{z})\propto f(\boldsymbol{z}|\boldsymbol{y},\alpha,\nu^2,\sigma)f(\alpha,\nu^2,\sigma),\qquad(29)$$

$f(\alpha,\nu^2,\sigma)$ は事前分布である．MCMCによるパラメータ推定では，複数のパラメータのサンプリングを異なる事後分布から別々に行うことが可能である．したがって，パラメータ β,γ とパラメータ α,ν^2,σ のサンプリングを，式 (28), (29) に示す2つの事後分布から別々に行うことができる．

提案モデルはサンプルセレクション，すなわち交通行動が一部のサンプルにしか観測されない状況を想定しているため，厳密には，式 (28), (29) に示した単純な構造では推定できない．しかし，基本的な考えは同じであり，本節で説明した点は提案モデルのパラメータ推定の際に最も重要な部分である．提案するアルゴリズムの詳細はここでは割愛するが，大まかな手順は以下の通りとなる．

交通行動が 2 値の場合

Step 1. Set the initial values $\beta^{(0)},\gamma^{(0)},\alpha^{(0)},\sigma^{(0)}$ and go to Step 2 with $k=1$.

Step 2. Sample $\boldsymbol{y}^{*(k)}$ and $\boldsymbol{z}^{*(k)}$ using a data augmentation step.

Step 3. Sample $\beta^{(k)}$ using Gibbs sampling.

Step 4. Sample $\gamma^{(k)}$ using MH sampling.

Step 5. Sample $\alpha^{(k)}$ and $\sigma^{(k)}$ for all $j\in J$ using Gibbs sampling.

Step 6. $k=k+1$, and return to Step 2 until $k=K$.

交通行動が連続の場合

Step 1. Set the initial values $\beta^{(0)},\gamma^{(0)},\alpha^{(0)},\sigma^{(0)},\nu^{2\,(0)}$ and go to Step 2 with $k=1$.

Step 2. Sample $\boldsymbol{y}^{*(k)}$ using a data augmentation step.

Step 3. Sample $\beta^{(k)}$ using Gibbs sampling.

Step 4. Sample $\gamma^{(k)}$ using MH sampling.

Step 5. Sample $\alpha^{(k)}$ and $\sigma^{(k)}$ for all $j\in J$ using Gibbs sampling.

Step 6. Sample $\nu^{2\,(k)}$ for all $j\in J$ using Gibbs sampling.

Step 7. $k=k+1$, and return to Step 2 until $k=K$.

---

k: MCMCの繰り返し数, K: 繰り返し総数, Data augmentation: データ拡大, Gibbs sampling: ギブスサンプリング, MH: メトロポリス・ヘイスティングス法

5. シミュレーション

(1) 設定した条件

本章では，人工的に生成されたデータを用いて，これまで説明した提案モデルとその推定アルゴリズムの検証を行う．ここでは，居住地選択における選択肢の数 J は3とし，連続的な交通行動を想定する．

設定する効用関数を以下に示す．

$$\begin{aligned} y_{i1}^* &= 1.0 - 0.5x_1 + \varepsilon_{i1}, \\ y_{i2}^* &= 0.0 + 0.5x_2 + \varepsilon_{i2}, \\ y_{i3}^* &= -1.0 + 1.5x_3 + \varepsilon_{i3}, \\ z_{i1} &= 1.0 + 1.0x_4 + \xi_{i1}, \\ z_{i2} &= 1.0 + 1.0x_5 + \xi_{i2}, \\ z_{i3} &= 1.0 + 1.0x_6 + \xi_{i3}. \end{aligned} \tag{30}$$

説明変数は，$x_1 \sim N(1,1)$, $x_2 \sim N(1,1)$, $x_3 \sim N(1,1)$, $x_4 \sim U(-1,1), x_5 \sim U(-1,1)$, $x_6 \sim U(-1,1)$ とした．その他のパラメータの真値は，$\gamma_{1,2} = \gamma_{1,3} = 0, \gamma_{2,3} = 0.30$, $\nu_1^2 = \nu_2^2 = \nu_3^2 = 1.0, \sigma_1 = \sigma_2 = \sigma_3 = 0.30$に設定した．以上の条件に基づいて3,000サンプルを生成した．

ここでは，サンプルセレクション，すなわち居住地が j の場合に z_{ij} が観測される現象を再現するため，$\max(y_i^*) = y_{ij}^*$ のサンプル i の z_{ij} 以外の交通行動，つまりz_{i-j}はデータから除外した．また実際には居住地選択の際の潜在効用 y_i^* は観察されないため，式 (10) により y_i に変換した後，潜在効用 y_i^* をデータから除外した．3章にて説明したように，提案モデルの内生的スイッチング構造を規定するのは誤差項の共分散パラメータ σ である．そこで本章では，σ を推定した場合と，σ を0に固定した場合の2つのパターンで推定を行い，結果を比較する．

(2) 推定結果

2つのパターンでの推定結果を**表-1**と**表-2**に示す．パラメータの識別を確保するため，居住地選択モデルにおける定数項と誤差相関パラメータをそれぞれ一つずつ0に固定している．**表-1**より，σ を推定した場合はおおよそ真値を再現することができていることがわかる．一方で，σ を0に固定した場合では真値から外れた推定値が得られた．これは，サンプルセレクションにより生じた自己選択バイアスによる影響が原因であると考えられる．

また，最終対数尤度と渡辺-赤池情報量基準 (WAIC)[12] を比較すると，σ を推定した場合の方がデータへの適合度とサンプル外予測精度の両方の点からモデルが σ を固定した場合より改善していることが示された．

表-1　推定結果 (σ を推定した場合)

変数[真値]	パラメータ	t値	
居住地選択モデル			
β_{10} [1.0]	1.09	13.75	**
β_{11} [-0.5]	-0.50	-12.50	**
β_{20} [0.0]	Fixed to 0		
β_{21} [0.5]	0.51	13.02	**
β_{30} [-1.0]	-0.93	-5.17	**
β_{31} [1.5]	1.50	10.56	**
$\gamma_{1,2}$ [0.0]	Fixed to 0		
$\gamma_{1,3}$ [0.0]	0.13	0.98	
$\gamma_{2,3}$ [0.3]	0.31	2.61	**
交通行動モデル			
α_{10} [1.0]	0.97	11.52	**
α_{11} [1.0]	0.99	17.58	**
α_{20} [1.0]	1.08	13.05	**
α_{21} [1.0]	1.03	16.86	**
α_{30} [1.0]	1.07	22.43	**
α_{31} [1.0]	0.99	18.09	**
ν_1^2 [1.0]	0.99	14.09	**
ν_2^2 [1.0]	1.06	17.57	**
ν_3^2 [1.0]	1.03	20.53	**
共分散パラメータ			
σ_1 [0.3]	0.34	2.98	**
σ_2 [0.3]	0.22	2.24	*
σ_3 [0.3]	0.21	2.95	**
サンプルサイズ		3,000	
最終対数尤度		-6681.28	
WAIC		13530.82	

*: p < 0.05, **: p < 0.01.

表-2　推定結果 (σ を0に固定した場合)

変数[真値]	パラメータ	t値	
居住地選択モデル			
β_{10} [1.0]	1.14	14.06	**
β_{11} [-0.5]	-0.53	-12.94	**
β_{20} [0.0]	Fixed to 0		
β_{21} [0.5]	0.54	13.80	**
β_{30} [-1.0]	-1.24	-5.27	**
β_{31} [1.5]	1.73	9.84	**
$\gamma_{1,2}$ [0.0]	Fixed to 0		
$\gamma_{1,3}$ [0.0]	-0.13	-0.67	
$\gamma_{2,3}$ [0.3]	0.12	0.74	
交通行動モデル			
α_{10} [1.0]	1.20	36.60	**
α_{11} [1.0]	0.99	17.58	**
α_{20} [1.0]	1.25	35.10	**
α_{21} [1.0]	1.03	17.05	**
α_{30} [1.0]	1.18	38.13	**
α_{31} [1.0]	0.99	18.28	**
ν_1^2 [1.0]	1.08	22.49	**
ν_2^2 [1.0]	1.11	20.86	**
ν_3^2 [1.0]	1.07	23.86	**
共分散パラメータ			
σ_1 [0.3]	Fixed to 0		
σ_2 [0.3]	Fixed to 0		
σ_3 [0.3]	Fixed to 0		
サンプルサイズ		3,000	
最終対数尤度		-6693.60	
WAIC		13548.82	

*: p < 0.05, **: p < 0.01.

6. 結論

本研究では，Heckman et al. (2001) のサンプルセレクションモデルをより一般的な構造へ拡張したモデルと，その推定アルゴリズムを提案した．Heckman et al. (2001) をはじめとした既存のサンプルセレクションモデルとの違いとして，既存のモデルが取り扱う居住地選択は2項選択なのに対し，提案モデルは多項選択により居住地選択を記述することができる．

既存のサンプルセレクションモデルと同様に，提案モデルは内生的スイッチング構造を有しており，因果推論をはじめとした様々な場面で適用が期待される．また，提案したMCMCアルゴリズムにおいて採用した，多変量正規分布とベイズ推定の性質を組み合わせて利用するというアイディアは，交通分野での適用は非常に限られているため，今後の積極的な活用が望まれる．

生成したデータを用いたシミュレーションでは，提案したMCMCアルゴリズムが適切にパラメータを推定できたことが示された．また，居住地選択と交通行動の依存関係を無視した場合，パラメータ推定値にバイアスが生じることが示唆された．

謝辞：本研究は日本学術振興会特別研究員奨励費 (DC2: 20J15157) とJSPS科研費 (18H01561) の助成を受けている．ここに感謝の意を表します．

参考文献

1) 織田澤利守, 大平悠季：交通インフラ整備効果の因果推論：論点整理と展望, 土木学会論文集 D3（土木計画学）, Vol.75, No.5, pp.I_1-I_15, 2019.
2) Kitamura, R., Mokhtarian, P. L. and Laidet, L.: A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area, *Transportation*, Vol.24, No.2, pp.125-158, 1997.
3) Mokhtarian, P. L. and Cao, X.: Examining the impacts of residential self-selection on travel behavior: A focus on methodologies, *Transportation Research Part B: Methodological*, Vol.42, No.3, pp.204-228, 2008.
4) Mokhtarian, P. L. and van Herick, D.: Quantifying residential self-selection effects: A review of methods and findings from applications of propensity score and sample selection approaches, *Journal of Transport and Land Use*, Vol.9, No.1, pp.9-28, 2016.
5) Heckman, J. J.: The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, *Annals of economic and social measurement*, Vol.5, No.4, pp.475-492, 1976.
6) Heckman, J. J.: Sample selection as a specification error, *Econometrica*, Vol.47, pp.153-161, 1979.
7) Heckman, J., Tobias, J. L., Vytlacil, E.: Four parameters of interest in the evaluation of social programs, *Southern Economic Journal*, Vol.68, No.2, pp.210-223, 2001.
8) Cao, X.: Disentangling the influence of neighborhood type and self-selection on driving behavior: An application of sample selection model, *Transportation*, Vol.36, pp.207-222, 2009.
9) Zhou, B. and Kockelman, K. M.: Self-selection in home choice: Use of treatment effects in evaluating relationship between built environment and travel behavior, *Transportation research record*, Vol.2077, No.1, pp.54-61, 2008.
10) Watanabe, H. and Maruyama, T.: A Bayesian sample selection model with a binary outcome: Examining the impact of residential self-selection on car ownership, 2021 (投稿中).
11) Cao, X., Mokhtarian, P. L. and Susan L. Handy.: Examining the impacts of residential self-selection on travel behavior: a focus on empirical findings, *Transport Reviews*, Vol.29, No.3, pp.359-395, 2009.
12) Watanabe, S.: Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory, *Journal of Machine Learning Research*, Vol.11, pp.3571-3594, 2010.

A BAYESIAN SAMPLE SELECTION MODEL WITH MULTINOMIAL ENDOGENOUS SWITCHING: APPLICATION FOR RESIDENTIAL LOCATION CHOICE AND TRAVEL BEHAVIOR

Hajime WATANABE and Takuya MARUYAMA