離散選択とその選択肢集合形成に生じる内生性バイアスの補正モデル

渡邉　萌 [1]

[1] 正会員　東京大学　大学院工学系研究科（〒113-8656 東京都文京区本郷 7-3-1）
E-mail: hwatanabe@bin.t.u-tokyo.ac.jp

交通行動分析に代表される離散選択モデリングにおいて，選択肢集合の形成とその選択肢集合を所与とした離散選択という二段階の意思決定をそれぞれモデリングする必要性が指摘されている．本研究では，既存モデルでは対処できなかった (a) 選択肢の利用可否の非独立性，(b) 離散選択・選択肢集合形成モデル間の非観測異質性，(c) 離散選択・選択肢集合モデルの両方に生じる内生性を対処する新しいモデルを提案する．これにより，政策変数が離散選択行動に与える影響の推定に影響を及ぼすバイアスを補正できる．提案モデルは居住地選択や目的地選択など，離散選択分析におけるその他の文脈においても適用可能である．

Key Words: *Endogeneity, Instrumental variable, Travel mode choice, Consideration set formation, Markov chain Monte Carlo*

1. はじめに

人々の離散的な交通行動は，選択肢集合の形成とその選択肢集合を所与とした離散選択という二段階の意思決定から構成されると想定し分析を行うのが望ましい．交通手段選択の例を用いて説明すると，人々が選択した移動手段は (1) 利用可能な交通手段の認知，(2) 利用可能な交通手段の中から得られる効用が最も高い交通手段を選択，という二段階の意思決定の結果であると解釈できる．この場合，公共交通が利用可能な交通手段として人々に認知されていない，すなわち公共交通が人々の選択肢集合にそもそも含まれない場合には，公共交通のサービスレベルを改善したとしても公共交通利用への転換にはつながらない可能性が高い．また，公共交通が人々の選択肢集合に含まれるために必要な施策の一つとして公共交通へのアクセスの改善が挙げられるが，アクセス改善による総所要時間の減少により公共交通を選択する場合の効用も同時に高まると考えられる．したがって，より効率的に人々の交通行動を望ましい方向へ転換させるためにも，人々の選択肢集合形成と選択肢集合を所与とした離散選択にそれぞれ影響を及ぼす要因とその影響度を正確に明らかにすることが必要である．

離散選択分析において選択肢集合の形成過程を考慮する必要性は長年指摘されており，そのための方法論も数多く開発されてきた．なかでも，Manski (1977)[1] は上述した二段階の意思決定を明示的にモデリングするフレームワークを提示しており，このフレームワークに従う形で Swait and Ben-Akiva (1987)[2] は Independent Availability Logit (IAL) モデルを提案している．IAL モデルは，選択肢が利用可能か否かがそれぞれ独立に決定されると仮定した上で，確率的な選択肢集合形成過程を離散選択モデルに組み込んだモデルである．Habib (2019)[3] はこの IAL モデルを拡張する形で Semi-Compensatory Independent Availability Logit (SCIAL) モデルを提案しており，SCIAL モデルはある選択肢が利用可能であるということ自体が，その選択肢を選択する確率に影響を与えるという現象，すなわち選択肢集合形成と離散選択の関連性を (部分的に) 説明することができる．これにより，離散選択における非保証型意思決定を表現している．しかし，これらのモデリングは選択肢集合が観測されないことを前提としており，そのためモデルの記述力が限定的である．具体的には，

(a)　選択肢の利用可否の非独立性
(b)　離散選択・選択肢集合形成モデル間の非観測異質性
(c)　離散選択・選択肢集合モデルの両方に生じる内生性

を記述できないという課題を抱えている．以下ではそれぞれ具体的に説明する．

それぞれの選択肢が利用可能かどうかは必ずしも独立に決定されるとは限らず，時には同時に決定されうる．

交通手段選択の例を挙げると，自動車への依存度が高く公共交通利用志向の低い人々にとって，バスと電車はどちらも選択肢集合に入らない可能性が高い．この場合，自家用車の利用可否と公共交通の利用可否は負の相関関係，また同じ公共交通であるバスと電車の間には正の相関関係があると考えられる．選択肢の利用可否が独立に決定されると仮定しているIAL モデルやSCIAL モデルでは，選択肢集合形成におけるこのような相関関係を十分に捉えることができない．

離散選択とその選択肢集合形成を確率的にモデリングするためには，それぞれのモデルに誤差項を仮定する必要がある．しかし，この 2 つのモデルの誤差項は互いに独立ではないと考えられる．ここでも交通手段選択の例を挙げるが，公共交通利用志向のある個人は，公共交通が選択肢集合に含まれる可能性が高く，また実際に公共交通を選択する可能性が高い．このように両方のモデルに同時に正の影響を及ぼす主観的な選好が未観測の場合，モデルの誤差項間に正の共分散，すなわち観測されない異質性が生じる．交通利用志向等の主観的な選好は一般に観測されないため，このような 2 つのモデル間の非観測異質性を記述可能なモデルが必要である．

また，誤差項と説明変数を設定し，説明変数に対応するパラメータを推定するということは，説明変数と誤差項が独立しているという仮定を暗に採用することになる．このようなモデルにおいて，説明変数と誤差項が相関している，すなわち説明変数が内生変数である場合，対応するパラメータの推定値にバイアスが発生する．一般にこの現象を内生性，生じるバイアスは内生性バイアスと呼ばれている．離散選択モデルにおける内生性への対処法は数多く提案されているが，筆者の知る限り，選択肢集合形成モデルにおける内生性に対処した既存研究は存在しない．

本研究では，人々の選択肢集合が観測されたという条件下で，上述の 3 つの課題に同時に対処するモデルを提案する．近年では選択肢集合を実際に観測する試みが行われており，例えば，米国で数年毎に実施されている世帯旅行調査 (National Household Travel Survey; NHTS) の 2017 年度に実施された最新の調査において，交通手段選択における選択肢集合を直接的に尋ねる設問が新たに追加されている．このような調査により実際の選択肢集合が観測された場合，選択肢集合形成モデルはそれぞれの選択肢の利用可否の 2 値情報 (選択肢集合に含まれる/含まれない) を利用して複数の 2 項選択モデルの同時確率で表現することができる．本提案モデルはこの選択肢集合形成モデルと選択肢集合を所与とした離散選択モデル，内生変数を従属変数とする線形モデルの間の誤差構造をプロビットモデルのフレームワークで柔軟に表現することにより，上述した 3 つの課題へ対処する．

2. 離散選択とその選択肢集合形成のモデリング

ここでは，Manski (1977) による離散選択とその選択肢集合形成の同時モデリングについて説明する．その後，提案モデルとManskiのモデルの違いについて説明する．

まず，全ての個人に共通な取りうる最大の選択肢集合をCとし，そのサブセットである個人$i \in n$の選択肢集合を$\tilde{C}_i$と仮定する．Manski (1977) は離散選択モデルとその選択肢集合形成モデルをかけあわせた以下のモデルを提案している．

$$\Pr(Y_i = t) = \sum_{\tilde{C}_i \subseteq G_i} \Pr\left(Y_i = t \middle| \tilde{C}_i\right) \Pr\left(\tilde{C}_i\right), t \in \tilde{C}_i, \quad (1)$$

Y_iは個人iの離散選択結果，G_iは個人iの取りうる全ての選択肢集合の組み合わせである．これより，このモデルでは全ての選択肢集合の組み合わせ毎の同時確率$\Pr\left(Y_i = t \middle| \tilde{C}_i\right)\Pr\left(\tilde{C}_i\right)$を足し合わせることにより，選択肢集合形成の確率を消去している．そのため，モデルの結果変数はY_iのみである．これにより，実際の選択肢集合$\tilde{C}_i$の観測を必要とせずに，離散選択モデルに選択肢集合形成過程の影響を組み込むことが可能である．

個人の選択肢集合$\tilde{C}_i$は通常観測されないため，このManski によるフレームワークは非常に実用的であり，多くの研究により踏襲・拡張されている．しかし，個人の選択肢集合が観測されたと仮定すると，選択肢集合形成そのものをモデル化でき，選択肢集合形成の確率を消去する必要はない．本研究では個人の選択肢集合が観測されたと仮定し，以下のように個人の選択肢集合と選択結果の同時確率のモデリングを行う．

$$\Pr\left(Y_i = t, \tilde{C}_i\right) = \Pr\left(Y_i = t \middle| \tilde{C}_i\right) \Pr\left(\tilde{C}_i\right), t \in \tilde{C}_i. \quad (2)$$

この点において，本稿にて提案するモデルはManskiとその拡張モデルと異なる．

3. 提案モデル

提案モデルは (1) 選択肢集合形成モデル，(2) 選択肢集合を所与とした離散選択モデル，(3) 内生変数を従属変数とする線形回帰モデル，の 3 つのモデルから構成されており，誤差構造を通じて互いに依存関係があることを許容する．以下ではそれぞれのモデルと誤差構造について説明する．

(1) 選択肢集合形成モデル

提案モデルではIAL モデルのフレームワークに従い，ある選択肢jの利用可否を，選択肢jが個人iの選択肢集合に含まれる，すなわち$j \in \tilde{C}_i$の場合は$Z_{ij} = 1$，それ以外の場合は$Z_{ij} = 0$として，2 値の確率変数Z_{ij}として定

義する．例えば，$C=(1,2,3,4)$かつ$\tilde{C}_i=(1,3)$の場合，$(Z_{i1},Z_{i2},Z_{i3},Z_{i4})=(1,0,1,0)$となる．IAL モデルの拡張モデルである SCIAL モデルでは，選択肢jが個人の選択肢集合$\tilde{C}_i$に含まれる確率$\Pr(Z_{ij}=1)$を2項ロジットモデルによりモデル化している．拡張性の高い柔軟な誤差

構造を持たせるため，提案モデルでは以下に示すように2項プロビットモデルを用いる．

$$\Pr(Z_{ij}=1)=\Pr(z_{ij}^*>0),\quad \text{where } z_{ij}^*=w_{ij}'\alpha_j+D_i\omega_{Z_j}+\eta_{ij}, \tag{3}$$

z_{ij}^*は潜在効用，w_{ij}とα_jはそれぞれ説明変数と対応するパラメータベクトルである．D_iは連続量の内生変数であり，ω_{Z_j}は内生変数に対応するパラメータである．η_{ij}は平均 0，分散$\Sigma_{Z_{jj}}$の正規分布に従う誤差項であり，全ての誤差項$\eta_i=(\eta_{i1},\eta_{i2},\dots,\eta_{iJ})'$は以下のような分散共分散を持つ多変量正規分布から発生すると仮定する．

$$\eta_i\sim N_J[\mathbf{0},\Sigma_Z],\ \text{where } \Sigma_Z=\begin{pmatrix}\Sigma_{Z_{11}} & \Sigma_{Z_{12}} & \cdots & \Sigma_{Z_{1J}}\\ \Sigma_{Z_{21}} & \Sigma_{Z_{22}} & \cdots & \Sigma_{Z_{2J}}\\ \vdots & \vdots & \ddots & \vdots\\ \Sigma_{Z_{J1}} & \Sigma_{Z_{J2}} & \cdots & \Sigma_{Z_{JJ}}\end{pmatrix}, \tag{4}$$

$\Sigma_{Z_{lm}}$は選択肢lと選択肢mとの間の誤差共分散であり，Jは個人の選択肢集合に含まれうる選択肢の総数である．$\Sigma_{Z_{lm}}$が正の場合，選択肢lと選択肢mが同時に選択肢集合に含まれる確率に影響を及ぼす未観測な要因が存在していることを示唆する．

ここで，選択集合形成モデルに生じる内生性について説明する．本稿では説明の簡略化のため，提案モデル全体で1種類の内生変数D_iを想定している．考えられる内生変数D_iとして代表的なものとして，居住地周辺の建築環境 (built environment) が挙げられる．人々は自らの居住地を選択する際，将来とりうる交通行動を想定し，自分が好む交通行動 (例: 電車通勤) に適した居住地を選択する可能性がある．このような交通行動に関連する主観的な選好が居住地選択に影響を及ぼす現象は居住地の自己選択 (residential self-selection) と呼ばれ，交通行動モデルにおいて内生性の問題を引き起こす代表的な例である [4]．そのような主観的な選好が観測され説明変数w_{ij}の中に含まれた場合，(3) のモデル上では内生性の問題は解消される，つまりD_iは外生変数となる．しかし，主観的な選好は一般に観測が困難であり，その影響は誤差項η_{ij}に含まれることがほとんどである．D_iに影響を与える主観的な選好が誤差項η_{ij}に含まれることにより，主観的な選好を通してD_iと誤差項η_{ij}との間に相関が生じる．この場合D_iは内生変数となり，対応するパラメータω_{Z_j}の推定値にバイアスが生じる．したがって多くの場合，交通インフラや公共空間の質といった近隣の建築環境に関する変数は交通行動分析においては内生変数である可能性が高いと考えられ，離散選択モデルだけでなく選択肢集合形成モデルにおいても内生性への対処は非常に重要な課題である．

提案モデルは選択肢集合形成確率$\Pr(\tilde{C}_i)$を全選択肢の利用可否$Z_i=(Z_{i1},Z_{i2},\dots,Z_{iJ})'$の同時確率として以下のように多変量プロビットモデルにより記述する．

$$\Pr(\tilde{C}_i)=\Pr(Z_i|\alpha,\omega_Z,\Sigma_Z)=\int_{A_{iJ}}\cdots\int_{A_{i1}}\phi_J(z_i^*|w_i\alpha+D_i\omega_Z,\Sigma_Z)\,dz_i^*, \tag{5}$$

$z_i^*=(z_{i1}^*,z_{i2}^*,\dots,z_{iJ}^*)'$，$w_i=\mathrm{diag}(w_{i1}',w_{i2}',\dots,w_{iJ}')$，$\alpha=(\alpha_1,\alpha_2,\dots,\alpha_J)'$，$\omega_Z=(\omega_{Z_1},\omega_{Z_2},\dots,\omega_{Z_J})'$，$\phi_J(\cdot|w_i\alpha+D_i\omega_Z,\Sigma_Z)$は平均$w_i\alpha+D_i\omega_Z$，分散共分散行列$\Sigma_Z$の J 次元正規分布の確率密度関数，A_{ij}は$Z_{ij}=1$のとき $(0,\infty)$，$Z_{ij}=0$のとき$(-\infty,0]$をとる区間である．

(2) 離散選択モデル

個人の選択肢集合$\tilde{C}_i$を所与とした離散選択Y_iの確率は

$$\Pr(Y_i=t|\tilde{C}_i)=\Pr(\max(\tilde{y}_i^*)=\tilde{y}_{it}^*),\forall t\in\tilde{C}_i, \tag{6}$$

$\tilde{y}_i^*=(\tilde{y}_{i1}^*,\tilde{y}_{i2}^*,\dots,\tilde{y}_{i\tilde{J}_i}^*)'$は選択肢集合$\tilde{C}_i$に含まれる選択肢を選択した場合の潜在効用ベクトルであり，$y_i^*=(y_{i1}^*,y_{i2}^*,\dots,y_{iJ}^*)'$の部分集合となる．また，$\tilde{J}_i$は個人の選択肢集合$\tilde{C}_i$に含まれる選択肢の総数である．選択肢$j$を選択した場合の潜在効用は以下のように定義される．

$$y_{ij}^*=x_{ij}'\beta_j+D_i\omega_{Y_j}+\varepsilon_{ij}, \tag{7}$$

x_{ij}とβ_jは説明変数と対応するパラメータベクトル，D_iは内生変数，ω_{Y_j}は対応するパラメータである．誤差項ベクトル$\varepsilon_i=(\varepsilon_{i1},\varepsilon_{i2},\dots,\varepsilon_{iJ})'$は以下の分散共分散を持つ平均0の$J$次元多変量正規分布に従う．

$$\Sigma_Y=\begin{pmatrix}\Sigma_{Y_{11}} & \Sigma_{Y_{12}} & \cdots & \Sigma_{Y_{1J}}\\ \Sigma_{Y_{21}} & \Sigma_{Y_{22}} & \cdots & \Sigma_{Y_{2J}}\\ \vdots & \vdots & \ddots & \vdots\\ \Sigma_{Y_{J1}} & \Sigma_{Y_{J2}} & \cdots & \Sigma_{Y_{JJ}}\end{pmatrix}, \tag{8}$$

$\Sigma_{Y_{lm}}$は選択肢lと選択肢mとの間の選択肢相関を捉える共分散である．したがって選択肢集合を所与とした離散選択確率$\Pr(Y_i=t|\tilde{C}_i)$は，多項プロビットモデルの枠組みで以下のように記述される．

$$\Pr(Y_i=t|\tilde{C}_i)=\int_{y_i^*\in\mathcal{Y}(i)}\phi_J(y_i^*|x_i\beta+D_i\omega_Y,\Sigma_Y)dy_i^*, \tag{9}$$

$x_i=\mathrm{diag}(x_{i1}',x_{i2}',\dots,x_{iJ}')$，$\beta=(\beta_1,\beta_2,\dots,\beta_J)'$，$\omega_Y=(\omega_{Y_1},\omega_{Y_2},\dots,\omega_{Y_J})'$，$\mathcal{Y}(i)$は$\max(\tilde{y}_i^*)=\tilde{y}_{it}^*$となるため

に潜在効用$y_i^* = \left(y_{i1}^*, y_{i2}^*, \dots, y_{iJ}^*\right)'$が取りうる範囲である．$\tilde{J}_i = 1$の場合，潜在効用$y_i^*$はどのような値でもとりうるため，選択確率$\Pr\left(Y_i = t \middle| \tilde{C}_i\right)$は1に帰着する．

(3) 内生変数を従属変数とする線形モデル

提案モデルは操作変数法により内生性バイアスの補正を行う．離散選択モデルにおいて操作変数法を援用した内生性への対処法は数多く提案されており [5]，なかでもコントロール関数法が有名である．本研究では，コントロール関数法の一種である操作変数モデル [6]のフレームワークを応用し，選択肢集合形成モデルと離散選択モデルに生じる内生性への対処を試みる．そのためには，以下のように内生変数を従属変数とする線形モデルを仮定する必要がある．

$$D_i = \nu_i'\delta + \mu_i, \tag{10}$$

ν_iとδは操作変数を含む説明変数と対応するパラメータベクトルである．μ_iは誤差項であり，平均 0，分散ξ^2の正規分布に従う．操作変数は内生変数D_iと相関している必要があるが，選択肢集合形成モデルと離散選択モデルの誤差項η_i, ε_iとは相関してはならない (操作変数の除外制約条件)．除外制約条件を満たす適切な操作変数を用いた上で，内生変数D_iを従属変数とする線形モデルの誤差項と選択肢集合形成モデルと離散選択モデルの誤差項η_i, ε_iとの共分散を推定することにより，内生変数D_iに対応するパラメータω_Z, ω_Yの推定値に生じる内生性バイアスを補正できる．提案モデル全体の具体的な誤差共分散構造については次節にて説明する．

(4) 提案モデルの誤差構造

選択肢集合形成モデルと離散選択モデルの潜在効用z_i^*, y_i^*と内生変数D_iは以下の多変量正規分布に従う．

$$\begin{pmatrix} z_i^* \\ y_i^* \\ D_i \end{pmatrix} \sim N_{2J+1}\left[\begin{pmatrix} w_i\alpha + D_i\omega_Z \\ x_i\beta + D_i\omega_Y \\ \nu_i'\delta \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y} & \sigma_Z \\ \Sigma_{Z,Y}^{\mathrm{T}} & \Sigma_Y & \sigma_Y \\ \sigma_Z^{\mathrm{T}} & \sigma_Y^{\mathrm{T}} & \xi^2 \end{pmatrix}\right], \tag{11}$$

$\sigma_Z = \left(\sigma_{Z_1}, \sigma_{Z_2}, \dots, \sigma_{Z_J}\right)'$，$\sigma_Y = \left(\sigma_{Y_1}, \sigma_{Y_2}, \dots, \sigma_{Y_J}\right)'$，$\sigma_{Z_j}$は誤差項$\eta_{ij}, \mu_i$の間の共分散，$\sigma_{Y_j}$は$\varepsilon_{ij}, \mu_i$の間の共分散である．$\Sigma_{Z,Y} = \mathrm{diag}\left(\rho_1, \rho_2, \dots, \rho_J\right)$，$\rho_j$は誤差項$\eta_{ij}, \varepsilon_{ij}$の間の共分散である．(1)で説明したように，分散共分散行列Σ_Zを式 (4) のようにパラメータ化することにより，提案モデルは選択肢の利用可否の非独立性を表現している．また，$\Sigma_{Z,Y}$により選択肢集合形成モデルと離散選択モデルの非観測異質性を捉えることが可能である．また，両方のモデルに生じる内生性バイアスは，内生変数と誤差項の相関をσ_Z, σ_Yにより捉えることで補正を行う．

4. おわりに

本研究では，離散選択とその選択肢集合形成を統合したモデリングにおいて，(a) 選択肢の利用可否の非独立性，(b) 離散選択・選択肢集合形成モデル間の非観測異質性，(c) 離散選択・選択肢集合モデルの両方に生じる内生性，に対処するモデルを提案した．提案モデルは交通行動選択だけでなく，居住地選択や目的地選択等，様々な文脈の離散選択分析にも適用可能である．

提案モデルはある政策変数が離散選択に与える影響を精査する必要がある場合に有効である．交通手段選択の例を挙げると，政策変数として公共交通のサービスレベルを想定した場合，公共交通のサービスレベルの向上が真に公共交通への転換につながるのかどうかを調べることは需要予測の点からも重要である．このとき，既存手法による分析では生じうる上述の (a-c) の問題に対応することができず，それにより政策変数の感度を過大 (または過少) に推定している可能性がある．また，政策変数に対応するパラメータの推定値に生じるバイアスの発生要因を考察することで，離散選択とその選択肢集合形成についてより理解を深めることにつながると考えられる．

謝辞：本研究は日本学術振興会特別研究員奨励費 (PD: 22J01164) の助成を受けている．ここに感謝の意を表します．

REFERENCES

1) Manski, C.F., 1977. The structure of random utility models, *Theory and Decision*, Vol.8, 229–254.
2) Swait, J., Ben-Akiva, M., 1987. Incorporating random constraints in discrete models of choice set generation, *Transportation Research Part B: Methodological*, Vol.21, pp.91–102.
3) Habib, K.N., 2019. Mode choice modelling for hailable rides: An investigation of the competition of Uber with other modes by using an integrated non-compensatory choice model with probabilistic choice set formation, *Transportation Research Part A Policy Practice*, Vol.129, pp.205–216.
4) Mokhtarian, P.L., Cao, X., 2008. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies, *Transportation Research Part B: Methodological*, Vol.42, pp.204–228.
5) Guevara, C.A., 2015. Critical assessment of five methods to correct for endogeneity in discrete-choice models, *Transportation Research Part A Policy Practice*, Vol.82, pp.240–254.
6) Watanabe, H., Maruyama, T. A Bayesian instrumental variable model for multinomial choice with correlated alternatives, *Journal of Choice Modelling*, 100400.

A MODEL ADDRESSING ENDOGENEITY IN DISCRETE CHOICE AND ITS CHOICE SET FORMATION

Hajime WATANABE