第6章 離散選択モデル
交通行動と分析モデリング p.103～p.145

福田研究室　修士2年

阿部紘之

目次

はじめに

- この章では離散選択モデルについて取り扱う。

→福田先生の「選択行動の数理モデル」の講義資料
交通ネットワークの均衡分析第4章も参照のこと。

- 離散選択モデルは様々な場面で用いられる。

→交通, マーケティング等で主に利用

- キーワード: 確率効用最大化モデル(RUM), 2項選択モデル, 多項選択モデル(MNL), 最尤法,
IIA特性, GEVモデル,ネスティッドロジットモデル(NL)

6.1 離散選択モデルの導出

- **離散選択と確率効用**

個人の行動を「選択」として考える

■ **連続型**

例) 1ヶ月の電話通話時間, 1ヶ月の労働日数

$$\boldsymbol{Y_n = \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_k x_{kn} + \varepsilon_n}$$

(Y_n:個人nの連続型被説明変数,
x_{kn}: 個人nのk番目の説明変数,
β_k: k番目の未知パラメータ ε_n: 確率項)

- 最小二乗法を用いて上式を推定

■ **離散型**

例)ブランドの選択, 交通手段の選択,
子供の有無, 本社の設置都市の選択

$$\boldsymbol{V_{in} = \beta_1 x_{1in} + \beta_2 x_{2in} + \cdots + \beta_k x_{kin}}$$

(V_{in}:個人nの選択肢iに対する効用,
x_{kin}: 個人nの選択肢iに対するk番目の説明変数,
β_k: k番目の未知パラメータ)

$$\boldsymbol{U_{in} = \beta_1 x_{1in} + \beta_2 x_{2in} + \cdots + \beta_k x_{kin} + \varepsilon_{in}}$$
$$\boldsymbol{= V_{in} + \varepsilon_{in}}$$

(U_{in}: 確率効用, V_{in}: 効用の確定部分, ε_n: 効用の確率項)

6.1 離散選択モデルの導出

<注意> 確率効用において行動者が確率的に行動することを表すわけではない
→確率項が存在しているために観測される行動が確率的に見えるため

効用の確率項が表す要因は以下の要因がある。

- 確定部分に含まれる変数以外の要因 (抜け落ちた変数)
- 確定部分を線形和とした関数形の誤差
- パラメータ β_k を個人間で均一とした誤差
- 説明変数の測定誤差

6.1 離散選択モデルの導出

ある個人が選択肢の中から選択

⇒その選択が選択肢の中で効用が最も高いということ

個人nが選択肢iを選択する確率

$$P_n(i) = Pr\left[U_{in} \geq U_{jn} \, , for \, \forall \, j \, , i \neq j\right] \quad (6.1.4)$$

個人が確率効用を最大化するモデルを<u>確率効用最大化モデル(RUM)</u>と呼ぶ。

6.1 離散選択モデルの導出

◈ **二項選択モデル (Binary choice model)**

(6.1.4)から導出

$$
\begin{aligned}
P_n(i) &= Pr\left[U_{in} \geq U_{jn}\right] \\
&= Pr\left[V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}\right] \\
&= Pr\left[\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}\right] \\
&= Pr[\varepsilon_n \leq V_{in} - V_{jn}]\ (\because \varepsilon_n \equiv \varepsilon_{jn} - \varepsilon_{in}) \\
&= \boldsymbol{F_\varepsilon(V_{in} - V_{jn})}
\end{aligned}
$$

※F_εはε_nの累積分布関数(CDF)

- プロビットモデル

中心極限定理によって正規分布を仮定
確率項を正規分布に従うと仮定

$$
\begin{aligned}
P_n(i) &= \Phi_\varepsilon\left(V_{in} - V_{jn}\right) \\
&= \int_{-\infty}^{V_{in}-V_{jn}} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2\right] d\varepsilon \\
&= \int_{-\infty}^{\frac{V_{in}-V_{jn}}{\sigma}} \frac{1}{\sqrt{2\pi}} exp\left[-\frac{1}{2}z^2\right] dz \\
&= \Phi\left(\frac{V_{in}-V_{jn}}{\sigma}\right)
\end{aligned}
$$

(Φ:標準正規分布の累積分布関数,
σ: ε_nの標準誤差)

※選択確率式に積分形が残るから計算が複雑!

6.1 離散選択モデルの導出

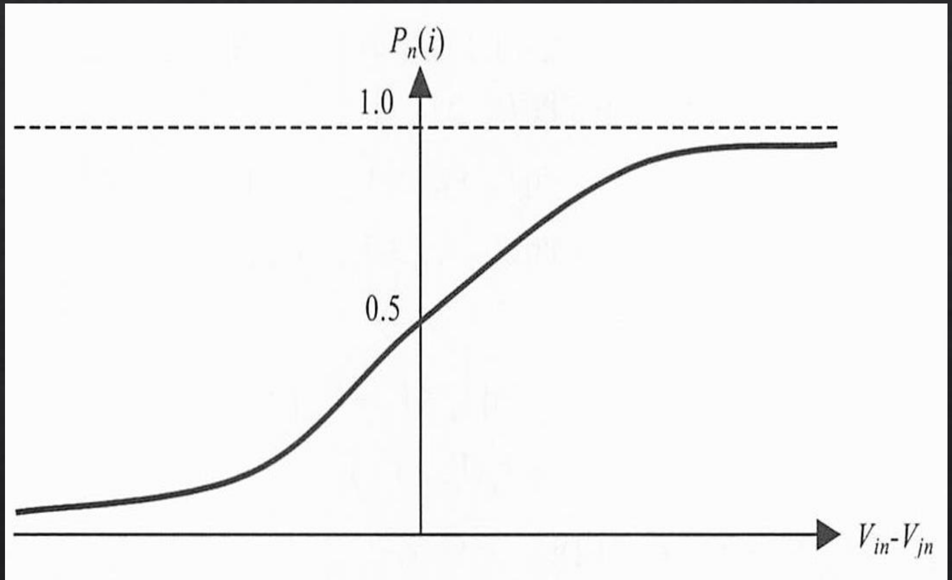
- **ロジットモデル(Logit model)**

2つの確率項の差ε_nにロジスティック分布を仮定

$$P_n(i) = \frac{1}{1+exp\{-\mu(V_{in}-V_{jn})\}}$$

$$= \frac{exp(\mu V_{in})}{exp(\mu V_{in})+exp(\mu V_{jn})}$$

(μ: スケールパラメータ[ε_nのばらつき程度を表現])



※二項選択モデルの選択確率は上図の様にS字型を描いている

6.1 離散選択モデルの導出

◈ 多項選択モデル(Multinomial choice model)

選択肢を3以上にしたモデルのことをいう

- **多項ロジットモデル(MNL)**

(定式化)

$$P_n(i) = Pr\left[U_{in} \geq U_{jn}, for \forall j, j \neq i\right]$$
$$= Pr\left[U_{in} \geq, \max_{\forall j, j \neq i} U_{jn}\right]$$

選択確率式に積分形が残らない(クローズドフォーム)
確率項に独立で同一(IID)のガンベル分布を仮定

- 累積分布関数

$$F(\varepsilon) = exp[-exp\{-\mu(\varepsilon - \eta)\}]$$

- 確率密度関数

$$f(\varepsilon) = \mu exp\{-\mu(\varepsilon - \eta)\}exp[-exp\{-\mu(\varepsilon - \eta)\}]$$

(μ: ガンベル分布のスケールパラメータ[ε_nのばらつき], η: ロケーションパラメータ[分布の位置(最頻値)])

6.1 離散選択モデルの導出

- **ガンベル分布の性質**

(1)最頻値はη, 平均値は$\eta+\gamma/\mu$　(γ(オイラー定数)≅0.577)

(2)分散は${}^{\pi^2}/_{6\mu^2}$

(3)$\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}$がそれぞれ$(\eta_1, \mu)$ (η_2, μ)のパラメータを持つガンベル分布に従う時$\varepsilon = \boldsymbol{\varepsilon_1} - \boldsymbol{\varepsilon_2}$は以下のロジスティック分布に従う。

$$\boldsymbol{F(\varepsilon) = \frac{1}{1+exp\{\mu(\eta_2 - \eta_1 - \varepsilon)\}}} \quad \textbf{(6.1.12)}$$

(4)$\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \dots, \boldsymbol{\varepsilon_j}$がそれぞれ$(\eta_1, \mu)$ (η_2, μ),…, (η_j, μ)のパラメータを持つガンベル分布に従う時$\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \dots, \boldsymbol{\varepsilon_j}$の最大値max$(\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \dots, \boldsymbol{\varepsilon_j})$もガンベル分布に従う。パラメータは、

$$\left(\frac{1}{\mu}\ln\sum_{j=1}^{J}\exp(\mu\eta_j), \mu\right)$$

6.1 離散選択モデルの導出

- **MNLの導出**

$\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \dots, \boldsymbol{\varepsilon_j}$がそれぞれ(0,$\mu$)を持つ互いに独立なガンベル分布に従い、$U_n^* \equiv \max_{\forall j, j \neq i} (V_{jn} + \varepsilon_{jn})$ と定義すると、

$\left(\frac{1}{\mu} \ln \sum_{j \neq i} \exp\left(\mu V_{jn}\right)\ , \mu\right)$ を持つガンベル分布となる

$U_n^* = V_n^* + \varepsilon_n^*$ とすると、

$$V_n^* = \frac{1}{\mu} \ln \sum_{j \neq i} \exp\left(\mu V_{jn}\right)$$

(6.1.4)および、ガンベル分布の性質(3)より、

$$\begin{aligned}
\boldsymbol{P_n(i)} &= \boldsymbol{Pr}[\boldsymbol{V_{in}} + \varepsilon_{\boldsymbol{in}} \geq V_n^* + \varepsilon_n^*] \\
&= \boldsymbol{Pr}[\varepsilon_{\boldsymbol{in}} - \varepsilon_n^* \leq V_n^* - \boldsymbol{V_{in}}] \\
&= \frac{\boldsymbol{1}}{\boldsymbol{1+exp(\mu(}V_n^* \boldsymbol{- V_{in}))}} = \frac{\boldsymbol{exp(\mu V_{in})}}{\boldsymbol{exp(\mu V_{in}) + exp(\mu} V_n^* \boldsymbol{)}} \\
&= \frac{\boldsymbol{exp(\mu V_{in})}}{\boldsymbol{exp(\mu V_{in}) + exp}\left(\ln \sum_{j \neq i} \exp(\mu V_{jn})\right)} \\
&= \boxed{\frac{\boldsymbol{exp(\mu V_{in})}}{\sum_j \exp\left(\mu V_{jn}\right)}}
\end{aligned}$$

MNLの選択確率

※通常のスケールパラメータμの値は1としている

6.2 離散選択モデルの推定

- **モデルの特定化(specification)**

説明変数の関数を決定する、誤差項に用いる確率変数を特定の分布系に仮定したりすること

(例)ある地域に住む人の通勤における3つの交通手段選択

自動車(Auto), 鉄道(Rail), バス(Bus)

交通交通手段:手段を決定する主な要因: 所要時間(t), 費用(c), 性別(f), 免許有無(l)

$$
\begin{aligned}
V_{An} &= \beta_1 + \beta_3 t_{An} + \beta_4 c_{An} + \beta_5 l_n \\
V_{Rn} &= \beta_2 + \beta_3 t_{Rn} + \beta_4 c_{Rn} \\
V_{Bn} &= \qquad\ \ \beta_3 t_{Bn} + \beta_4 c_{Bn} + \beta_6 f_n
\end{aligned}
$$

	β_1	β_2	β_3	β_4	β_5	β_6
Auto	1	0	t_A	c_A	1	0
Rail	0	1	t_R	c_R	0	0
Bus	0	0	t_B	c_B	0	f
	定数項		LOS変数		SE変数	

「定数項」‥説明変数で表せなかった効用内で全個人で共有する値

「サービスレベル(LOS)変数」‥各選択肢独自のLOSを表す変数

「社会経済(SE)変数」‥個人属性等を表し、選択肢間で値が変化しない

6.2 離散選択モデルの推定

- **最尤推定法(maximum likelihood estimation, MLE)**

行動モデル(未知パラメータ含む)が正しいことが前提で観測されたデータがもっともらしさ(尤度)が最大になるようにパラメータを定めている。

モデル推定においてパラメータ値を求める方法として利用されている

$$L = \prod_{n=1}^{N} \prod_{i=1}^{J} P_n(i)^{d_n}$$

(d_n: 個人nが選択肢iを選んだ時に1、選んでない時は0)

実際のモデル推定では推定計算の簡便性のために上式の尤度関数に対数をとる

$$\ln L = \prod_{n=1}^{N} \prod_{i=1}^{J} d_n P_n(i)$$

上式の対数尤度関数を未知パラメータに関して最大化する

6.2 離散選択モデルの推定

◆ **統計的検定**

得られたパラメータ推定値がどれほど意味を持つかを検定

それぞれの係数が0から統計的に優位に離れているか
➡各説明変数が効用値に影響を与えているか ➡ t検定

得られたパラメータ値を標準偏差を除した値をt値という
t値が出力されて、その値が±1.96以上でパラメータ値が有意(有意水準5%)であると言える
→0から十分離れている

6.2 離散選択モデルの推定

- **モデルの適合度**

モデルの適合度指標として、**尤度比(likelihood ratio)**に基づく

パラメータ推定値を定める最大尤度$L(\beta^*)$は無情報モデル(パラメータが0)の初期尤度 $L(0)$と完全情報モデル(選択肢が全て1になる時)の尤度1(対数尤度では0)の間に存在

尤度比: -2(ln $L(0)$ − ln $L(\beta^*)$) ➡ カイ二乗分布に従っている

対数尤度の差を用いた適合度指標(McFaddenの決定係数)

$$\rho^2 = \frac{\ln L(\beta^*) - \ln L(0)}{\ln 1 - \ln L(0)} = \frac{\ln L(0) - \ln L(\beta^*)}{\ln L(0)} = 1 - \frac{\ln L(\beta^*)}{\ln L(0)}$$

本教科書ではサンプル数を数百、選択肢を5つの時ならば0.2程度以上でよい

※この適合度指標は回帰分析の決定係数の様に基準はない

6.2 離散選択モデルの推定

McFaddenの決定係数の欠点: 説明変数を増やせば値が増加(回帰分析の決定係数と同様)

自由度調整済み尤度比の導入 (回帰分析の自由度調整済み決定係数と同様の考え方)

$$\bar{\rho}^2 = \frac{\ln L(0) - (\ln L(\beta^*) - K)}{\ln L(0)} = 1 - \frac{\ln L(\beta^*) - K}{\ln L(0)}$$

(K: モデルに含まれる未知パラメータの数)

$\ln L(\beta^*) - K$ の値が小さいほど適合度が良いと言われている

6.3 離散選択モデルによる予測

- 交通行動に離散選択モデルを導入する目的

(1)個人の行動をシステマティックに理解➡推定された効用のパラメータから知る
(2)交通現象の予測➡非集計レベルで推定したモデルを予測段階で集計化

- 集計方法

(1)数え上げ法(総当たり法)
モデル推定に使われたサンプルがマーケット全体をよく表すと仮定

$$S(i) = \frac{1}{N}\sum_{n=1}^{N} P_n(i) \ , i = 1, \dots, j$$

($S(i)$: 選択肢iのマーケットシェア, N: サンプル数, $P_n(i)$個人nの選択肢iの選択確率)

※ある選択肢のマーケットシェアは個人の選択確率の重み付き総和であって各個人がどの選択肢を選択するかを0〜1で予測してから求めるわけではない

6.3 離散選択モデルによる予測

なお、選択肢別に標本抽出を行った時の各選択肢ごとのマーケットシェア

$$S(i) = \sum_{j=1}^{J} W_j \frac{1}{N_j} \sum_{n=1}^{N_j} P_{nj}(i) \ ,i = 1, \dots, j$$

(W_j: アンケート調査時の実際のマーケットシェア, Nj : 選択肢iのサンプル数, $P_{n\mathrm{j}}(i)$: 現在選択肢jを選択している個人nの選択肢iの選択確率)

(2)代表的個人法

母集団の「代表的」な仮想の個人の説明変数値を計算して、その個人の選択確率をセグメントにシェアすること

- 「代表的」な変数値➡モデル推定に用いたサンプル数の平均値

母集団の中で説明変数のバラつき大きいと予測精度は低下

⇩

説明変数がばらつかない同質的なグループ毎に「代表的個人」を立てる

(マーケットセグメントといい、年齢層,性別等のセグメント分けが容易なグループを用いる)

※グループ内が同質的になる保証がないので、あまり使われない

6.3 離散選択モデルによる予測

- ロジットモデルによる弾性値

Aに対するBの弾力性➡ Aの単位変化比率に対するBの変化比率

「直接弾性値」➡ある選択肢の属性変化させた時のその選択肢の需要変化
「交差弾性値」➡ある選択肢の属性変化させた時の他の選択肢の需要変化

※交差弾性値は選択確率に影響を受けない→すべての選択肢で同じ値

IIA特性の一つの表現方法でもある

6.4 IIA特性を持たない離散選択モデル

- ロジットモデルとIIA特性

ロジットモデルの2つの選択肢の選択確率比 $\rightarrow \frac{P_{in}}{P_{jn}} = \exp[V_i - V_j]$

選択肢i, jの効用の確定部分だけ影響を受けている(他の選択肢の影響を受けない)

IIA特性(independence from irrelevant alternatives)
[無関係な選択肢から選択確率の独立性]

- IIA特性の長所

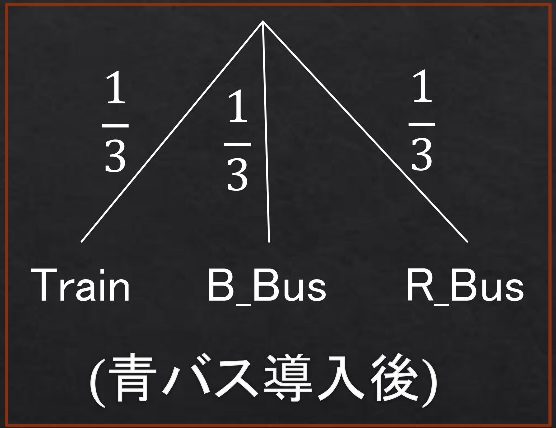
全ての選択肢を知る必要がない(選択肢集合の部分集合を用いて推定できる)
➡選択肢集合が極めて大きい, 不確定の時に推定してもパラメータ推定値にバイアスが生じない

6.4 IIA特性を持たない離散選択モデル

- IIA特性の短所(赤バス-青バス問題)

効用の確定部分が同じ電車と赤バスがあると考える。
ここで新たなバス(青バス)を導入するとする。サービスは同一だが色が異なるものとして考える。

赤バスと青バスの効用は同じ



バス2/3, 電車1/3となる

しかし.....
青バスを導入してもバス効用は変わらない
⇩
電車の選択確率=1/2
青バス=赤バス=1/4

原因: <u>選択肢間の確率項(誤差項) が独立という仮定が間違いで、相関が生じているため</u>

➡類似した選択肢の選択確率が過大に.....

この特性を緩和するためにネスティッドロジットモデルを導入

6.4 IIA特性を持たない離散選択モデル

- 選択肢間の確率項(誤差項)に相関が生じる例

(1)交通手段選択(車,バス,鉄道)

➡ バスと鉄道は「公共交通機関」という表現しにくい共通項がある

(2)経路選択

➡ 一部のリンクを共有している経路間には相関が生じやすい

(3)目的地選択

➡近い目的地同士には相関が生じやすい

(4)統合モデル(例えば、目的地選択と手段選択)

➡共通の目的地や手段を持つ選択肢間では相関が生じやすい

6.4 IIA特性を持たない離散選択モデル

- GEV(General Extreme Value)モデル

一般化極値分布モデルともいい、一般的な極値分布を仮定

個人nがJ個の選択肢を持ち、$B_n^1, \ldots\ldots, B_n^k$というk個のサブグループに分かれていると考える

GEVモデルの選択確率は、

$$P_n(i) = \frac{\exp(\frac{V_{in}}{\lambda_k})\left(\sum_{j\in B_n^k} \exp(\frac{V_{jn}}{\lambda_k})\right)^{\lambda_k - 1}}{\sum_{k=1}^{K}\left(\sum_{j\in B_n^k} \exp(\frac{V_{jn}}{\lambda_k})\right)^{\lambda_k}}$$

λ_k: サブグループ内の相関の程度を表すパラメータ($0 \leq \lambda_k \leq 1,\ \lambda_k$=1は無相関でロジットモデル)

➡GEVモデルから多項ロジットモデルとネスティッドロジットモデルを導出することができる

6.4 IIA特性を持たない離散選択モデル

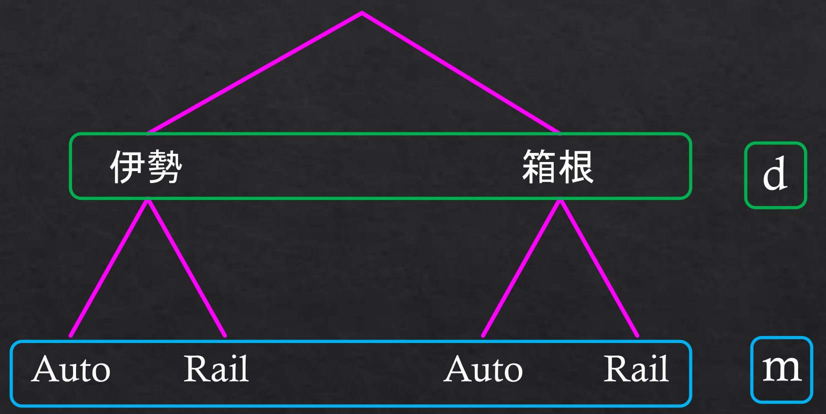
- ネスティィッドロジット(NL)モデル

IIA特性を緩和したモデル、GEVモデルからも導出できる

(効用関数)

$$U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$$

V_d:目的地選択肢dに特有な効用の確定項
V_m:手段選択肢mに特有な効用の確定項
V_{dm}:目的地選択肢dと手段選択肢mの組み合わせで決まる効用の確定項
ε_d:目的地選択肢dに特有な効用の確率項
ε_{dm}:目的地選択肢dと手段選択肢mの組み合わせで決まる効用の確率項



(選択確率[条件付き確率と周辺確率の積])

$$P(d,m) = P(m|d)P(d)$$

$P(m|d)$→条件付き確率 (目的地決定後の手段の選択確率)
$P(d)$→周辺確率 (目的地の選択確率)

6.4 IIA特性を持たない離散選択モデル

(周辺確率)

$$P(d) = \frac{\exp\{\mu^d (V_d + V'_d)\}}{\sum_{d' \in \{I,H\}} \exp\{\mu^d (V_{d'} + V'_{d'})\}}$$

V_d: 目的地dが持つ効用の確定項, V'_d:車と鉄道の両方を考えた交通アクセスの合成効用

確定項は$V_d + V'_d$、確率項は$\varepsilon_d + \varepsilon'_d$の離散選択モデル　(スケールパラメータ$\mu^d$のガンベル分布)

V'_d⇒ログサム変数(ツリー下位レベルの合成効用)

$$V'_d \equiv \frac{1}{\mu} \ln \sum_{m \in \{A,R\}} \exp\{\mu (V_m + V_{dm})\}$$

(条件付き確率)
目的地dが与えられたときの手段選択では選択肢間の共通項V_d, ε_dは関与しない

$$P(m|d) = \frac{\exp\{\mu (V_m + V_{dm})\}}{\sum_{m' \in \{A,R\}} \exp\{\mu (V_{m'} + V_{dm'})\}}$$

従ってNLの選択確率は、

$$P(d,m) = P(m|d)P(d) = \frac{\exp\{\mu (V_m + V_{dm})\}}{\sum_{m' \in \{A,R\}} \exp\{\mu (V_{m'} + V_{dm'})\}} \cdot \frac{\exp\{\mu^d (V_d + V'_d)\}}{\sum_{d' \in \{I,H\}} \exp\{\mu^d (V_{d'} + V'_{d'})\}}$$

6.5 複数データに基づくモデル推定

- 複数データソースの考え方

データは採取方法によって異なるため、1種類のデータでモデル推定ができるとは限らない

(1)集計データと非集計データ

➡ 大まかだが精度が良いか、説明要因が分かるがモデル化しにくいか

(2)カウントデータとアンケートデータ

➡ 属性が分からないが物理量のバイアスがない、属性が分かるが回答バイアスあり

(3)RPデータ(実行動データ)とSPデータ(アンケート調査等)

➡実行動と属性値の関係が不明、属性値をコントロールできるが実行動との関係が不明

(4)異なる地域データ(地域の特徴が反映)　　　　(5)異なる時点データ(時点の特徴とつながり)

〇モデル化の枠組み

- あるデータソースから推計されたモデルを別のデータによって更新
- 複数のデータソースの特徴を考慮してモデル推定時に同時に利用

6.6 離散選択モデルの応用

- <u>個人の嗜好性</u>

効用関数中の係数パラメータは母集団の中で同一と仮定
→しかし、個人ごとに係数の真値が異なると考えることが自然

(1)個人属性の導入

<u>性別,職業,所得等の社会経済変数を変数として導入</u>

$$U_{in} = \alpha_i + \underbrace{\boxed{\beta' X_{in}}}_{\text{LOS変数}} + \underbrace{\boxed{\gamma_1 s_n + \gamma_2 l_n}}_{\text{社会経済変数}} + \varepsilon_{in}$$

(α_i: 定数項, s_n: 性別ダミー, l_n: 免許保有ダミー)

LOS変数にかかる係数を個人属性によって変化させることも可能
例)単位所要時間が性別によって効用に影響が異なる場合➡ $(\beta_1 + \beta_2 s_n) t_{in}$と表してもよい

ただし、どのような属性によって異質的であるかを先験的に、試行錯誤に決定する必要あり

6.6 離散選択モデルの応用

(2)マーケットセグメンテーション

嗜好がおおよそ均一となるサブグループ(マーケットセグメント)をいくつか準備して、そのグループ毎に推定

ただし、どのような属性でグループを定義するか決定する必要がある

限られたサンプル数でグループを多くすると、グループ内のサンプル数が少数で推定に支障がでる

(3)潜在クラスモデル

個人がどのセグメント(グループ)に属するかを確率モデルで表す方法

$$P_n(i) = \sum_{s=1}^{S} P_n\,(i|s)Q_n(s) \qquad i = 1, \ldots, j$$

($P_n(i|s)$:個人nがセグメントsに属している時の選択肢iの選択確率, $Q_n(s)$:個人nがセグメントsに属する確率[帰属確率])

セグメントの総数Sを定める

セグメント毎の係数パラメータと帰属確率のモデルは最尤法で推定

6.6 離散選択モデルの応用

(4)個人別パラメータモデル

個人ごとにかなりの数の観測値が必要 (例)パネル調査, SPデータ

(5)確率係数モデル

効用関数の係数パラメータを以下の様に確率変数として記述

$$U_{in} = \beta' X_{in} + \varepsilon_{in}$$

$$\beta \sim MVN(\overline{\beta}, \Omega)$$

($\overline{\beta}$:平均値ベクトル, Ω:共分散行列)

ε_{in}が多変量正規分布⇨多項プロビットモデル(MNP)

ε_{in}がIIDガンベル分布⇨mixed logitモデル(MNLとMNPの融合型)

6.6 離散選択モデルの応用

- <u>選択肢集合の考え方</u>

(1)選択肢集合があらかじめ与えられている場合

通常の離散選択モデルでは、個人の選択肢集合にどのような選択肢が入っているかが分かっている

例.運転免許を持ってない⇒自動車の選択肢を除外　4km以上離れた目的地を「徒歩」⇒一般的でないので除外

※誤った選択肢集合を仮定にした場合パラメータ推定でバイアスが生じる等、<u>選択肢集合は重要</u>

(2)IIA特性は成立だが選択肢が極めて大きい場合

「選択肢の集計化」・・・例えば、目的地選択をゾーンとする

真の目的地(要素選択肢)の効用値がすべて同一の確定項効用を持つとき選択確率は、

$$P_n(i) = \frac{exp(V_{in} + \frac{1}{\mu}\ln M_i)}{\sum_{j=1}^{J} exp(V_{jn} + \frac{1}{\mu}\ln M_j)}$$

V_{in}:選択肢iの確定効用値
M_i:選択肢iの要素選択肢数

「選択肢のサンプリング」・・・IIA特性の長所の活用が可能

ただし、無作為に抽出する際に実際に選択した選択肢を含めることに留意

6.6 離散選択モデルの応用

(3)選択肢集合形成過程のモデル化

選択肢集合が分析者にとって不確定の場合に確率的な選択肢集合を考えた分析が有効

$$P_n(i) = \sum_{C \in G} P_n(i|C)\, Q_n(C)$$

($P_n(i|C)$:個人nの選択肢集合がCの選択肢iの選択確率, $Q_n(C)$:個人nの選択肢集合がCである確率, G:選択肢全ての部分集合[空集合以外])

選択肢集合の形成と選択肢集合が所与の下での選択行動の2段階構成

➡選択肢集合の不確実性を考慮した離散選択モデルを構成

まとめ

◈ 離散選択モデルについて再度振り返ることができた
→特に二項選択モデル,多項ロジットモデルについて理解を深めた

◈ 離散選択モデルは他にもいくつかある
→クロスネスティッドロジットモデル(CNL), network-GEV等

(追加) 6.4 IIA特性を持たない離散選択モデル

- ネスティッドロジットモデルにおける注意点

$\frac{\mu^d}{\mu}\left(=\sqrt{\frac{Var(\varepsilon_{dm})}{Var(\varepsilon_d)+Var(\varepsilon_{d'})}}\right)$ [Var: 分散]の値において、

$\frac{\mu^d}{\mu}$の値	特徴
1以上	ツリーの上下関係が逆 ⇒ツリーの構造を変えて再推定
1に近い値	ε_dの影響がない⇒MNLで表現可
0に近い値	ε_dの影響がε_{dm}より大きい ⇒下位ツリー変化が上位の選択に影響をしない（赤バス-青バス問題に近い）

※$\mu = 1$としてμ^dを推定することが多い

※基本的に$\frac{\mu^d}{\mu}$の値は1以下である必要がある

(追加) 6.4 IIA特性を持たない離散選択モデル

- ミックスドロジットモデル(MXL)

多項ロジットモデルと多項プロビットモデル(MNP)の融合型モデル

(効用関数)

$$U_{in} = V_{in} + \eta_{in} + \xi_{in}$$

η_{in}: 共分散行列Ωを持つ多変量正規分布に従う確率項
ξ_{in}: IIDガンベル分布に従う確率項

(条件付き確率)

$$P(i|\eta_n) = \frac{\exp(V_{in}+\eta_{in})}{\sum_{j=1}^{J} \exp(V_{jn}+\eta_{jn})}$$

選択確率の計算はJ次元の多重積分を行う
↓
モンテカルロ・シミュレーションから近似
(ギブスサンプリング法etc……)

(選択確率)

$$P_n(i) = \int P(i|\eta_n)\, f(\eta_n|\Omega) d\eta_n = \int \frac{\exp(V_{in} + \eta_{in})}{\sum_{j=1}^{J} \exp(V_{jn} + \eta_{jn})} f(\eta_n|\Omega) d\eta_n$$

(追加) 6.5 複数データに基づくモデル推定

- 新たなデータによるモデルパラメータの更新

◈ 一部パラメータの再推定

例. 交通手段選択で時間価値は等しいが、地域特性が大きく影響する手段別定数項が異なる場合

$$P_n(i) = \frac{\exp(\alpha_i + \mu\widehat{V_{in}})}{\sum_{j=1}^{J} \exp(\alpha_j + \mu\widehat{V_{jn}})}$$

μ, αを再推定
$\widehat{V_{jn}}$:は移転元で推定された効用関数のパラメータと移転先の説明変数値で計算した効用確定項

◈ ベイズ法によるパラメータの更新

ある確率変数の事前分布をデータを得て事後分布に更新する方法

$$p(\theta|X) = \frac{f(X|\theta)\pi(\theta)}{\int f(X|\theta)\pi(\theta)d\theta}$$

$f(X|\theta)$:θが与えられた時のXの出現確率(尤度)
$\pi(\theta)$: θの事前分布
$p(\theta|X)$: Xが与えられた時のθの事後分布

- あるデータから推定したパラメータ推定量を事前情報として、他のデータから推定した推定量の分布を尤度関数として両推定量の重み付き平均と考える (藤原・杉恵,1990)
- 非集計モデルから推定されたパラメータを選択肢のシェア(集計データ)を用いて更新する方法(森地,1987)

(追加) 6.5 複数データに基づくモデル推定

- RP/SPモデル

(考え方)

RPとSPの利害得失は互いに補完的、RPで推定できなかったパラメータをSPで補強

SPに含まれるバイアスやランダムエラーを修正

(特徴)

1.**バイアスの修正**: 2つのデータ発生過程を別々にモデル化することでSPバイアスの影響を需要予測から除去が可能

2.**統計的有効性の増大**: 属性間のトレードオフを表すパラメータを両データから同時推定で増大

3.**パラメータの同定**: RPからだけでは同定できないパラメータを推定

(追加) 6.6 離散選択モデルの応用

- 便益指標としてのログサム変数

各選択肢の確率項に$E(\varepsilon) = 0$とすると、ログサム変数は全選択肢の最大効用の期待値
個人n, 選択肢集合Cnとすると、

$$E\left[\max_{i \in C_n} U_{in}\right] = \frac{1}{\mu} \ln \sum_{i \in C_n} \exp(\mu V_{in})$$

ログサム変数には選択肢集合の合成効用として望ましい性質がある
(1)新しい選択肢を加えるとログサム変数の値が大きく増加

$$E\left[\max_{i \in C_n} U_{in}\right] \leq E\left[\max_{i \in C'_n} U_{in}\right]$$

C’n: 選択肢集合Cnの部分集合

全選択肢の効用の平均値を合成効用とした場合、効用値の低い選択肢が選択肢集合に加えると効用値が減少

(追加) 6.6 離散選択モデルの応用

(2)<u>選択肢の確定効用値が上がるとログサム変数の値が大きく増加</u>

$$\frac{\partial}{\partial V_{jn}} E\left[\max_{i \in C_n} U_{in}\right] = P_n(j) > 0, \forall j \in C_n \qquad \frac{\partial P_n(j)}{\partial V_{in}} = \frac{\partial^2 E\left[\max_{i \in C_n} U_{in}\right]}{\partial_{V_{in}} \partial_{V_{jn}}} = \frac{\partial P_n(i)}{\partial V_{jn}}$$

これらの式から政策施行前後のログサム変数変化が消費者余剰の変化に等しいことを示した(Williams(1997))

消費者余剰の変化を政策施行前後の確定効用値をV_n^1, V_n^2として表すと、

$$\Delta CS_n = \sum_{i \in C_n} \int_{V_n^1}^{V_n^2} P(i|V) dV = \frac{1}{\mu} \ln \sum_{i \in C_n^2} \exp(\mu V_{in}^2) - \frac{1}{\mu} \ln \sum_{i \in C_n^1} \exp(\mu V_{in}^1)$$