McFadden, D., Train, K., Mixed MNL models for discrete response, Jounal of Applied Econometrics, pp.447-470, 2000.

離散選択のMixed MNLモデル

M1　瀧口洋平

発表の構成

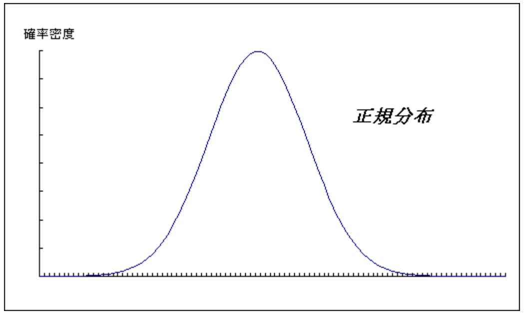
1. MMNL(Mixed MNL)とはなにか
2. 近似：確率効用最大化(RUM)理論に従う離散選択モデルの選択確率の近似解を得られる.
3. 推定：シミュレーションによる推定手法
4. 実験：車種選択問題への適用

Behavior in Networks Research Unit

1-1 MMNLについて

- Mixed Multinomial Logit Model(MMNL)とは？
  - 嗜好の異質性，系列相関などの制約を全て緩和したモデル．

$$U_{nj} = \beta'_n x_{nj} + \varepsilon_{nj}$$ ・・・(1)



β_n：サンプルnについて確率分布$f(\beta)$に従うパラメータ
x_{nj}：サンプルn，選択肢jに関する特性変数ベクトル
ε_{nj}：IID(独立で同一な)ガンベル分布
→MNLではβは全サンプルで一定の値をとる

パラメータβが1つ選ばれて与えられたときの選択確率はロジット型であり

$$L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^{J} e^{V_{nj}(\beta)}}$$ ・・・(2)

1-1 MMNLについて

- βが確率的に分布する時，標準選択確率は

$$P_{ni} = \int L_{ni}(\beta) f(\beta)\, d\beta = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) f(\beta)\, d\beta \quad \cdots (3)$$

- パラメータβを与える分布$f(\beta)$を推定することになる.
  - 一般的には分布の平均値，共分散を調べる.
  - 分布は正規分布，一様分布，ガンマ分布など様々
  - 適切な分布を用いて様々な効用最大化行動を表現可能
- 積分が残るのでMNLのように簡単に推定できない
  - シミュレーションを用いる

1-2 RUMについて

- RUM（Rundam Utility Maximization）モデルとは，個人の効用を確率変数として捉え，その効用を最大化すると考えるモデル．
  - 選択結果も確率的に（選択確率として）出る．
- 効用関数は次のように書ける．

n：選択者
j：選んだ選択肢
α：パラメータ
z：説明変数

$$U_{nj} = \alpha'_n z_{nj} \quad \cdots(4)$$

または

$$U_{nj} = \beta'_n x_{nj} + \varepsilon_{nj}$$

ここで $z'_{nj} = \langle x'_{nj}, d_j \rangle, \alpha' = \langle \beta'_n, \varepsilon_{nj} \rangle$

$\cdots(5)$

x，d：説明変数の要素
β：確定項の係数
ε：確率項の係数

Behavior in Networks Research Unit

1-2 RUMについて

- 選択肢iの条件付き選択確率
  - パラメータαが与えられたとき，選択確率は0か1

$$q_{ni}(\alpha) = I(\alpha_n' z_{ni} > \alpha_n' z_{nj} \ \forall j \neq i) \quad \cdots(6)$$

I：条件を満たすとき1，その例外は0を取る

- αの確率分布を考慮すると選択確率は

$$Q_{ni} = \int I(\alpha_n' z_{ni} > \alpha_n' z_{nj} \ \forall j \neq i) f(\alpha)\, d\alpha \quad \cdots(7)$$

RUMモデル中で確率項を考慮することでロジットモデル（ガンベル分布），プロビットモデル（正規分布）などが導出される.

bin Behavior in Networks Research Unit

1-3 MMNLとの違い

- MMNLは式(4)の確定項の一部をIID誤差項と考える
  - IID（independently and identifically, 独立で同一な）
  - ガンベルに従う誤差項を確率分布から分離する
  - 平均値0で適宜設定する分布を持つ誤差項

$$U_{nj} = \alpha_n' z_{nj}' + \text{IID誤差項} \quad \cdots(8)$$

z'_{nj}:z_{nj}からIID誤差項を分離した変数

- スケールパラメータλを導入した選択確率

$$P_{ni} = \int \left(\frac{e^{(\alpha/\lambda)' z_{ni}}}{\sum_j e^{(\alpha/\lambda)' z_{nj}}} \right) f(\alpha)\, d\alpha \quad \cdots(9)$$

λを変化させて式(7)と式(9)を等価にしたい

bin Behavior in Networks Research Unit

1-4 誤差項を無視するという事

- 例
  - λを小さくするとα/λは大きくなる
  - すると式(9)のP_{ni}は1か0に近づくので　ex)$\int e^{ax}\mathrm{dx} = (1/\mathrm{a})\ e^{ax}$+C

$$P_{ni} = \int \left(\frac{e^{(\alpha/\lambda)' z_{ni}}}{\sum_j e^{(\alpha/\lambda)' z_{nj}}} \right) f(\alpha)\, d\alpha$$

$$\fallingdotseq\ Q_{ni} = \int I(\alpha_n' z_{ni} > \alpha_n' z_{nj}\ \forall j \neq i) f(\alpha)\, d\alpha$$

- (8)式は次式で表され，IID誤差項は確率に影響しない

$$U_{nj} = \frac{\alpha_n}{\lambda} z_{nj}\ +\text{IID誤差項} \qquad \cdots(10)$$

$U_1(\alpha/\lambda)$より$U_2(\alpha/\lambda)$が5大きい時，誤差による変化が±3とすると，選択結果は変わりうる．λ を10倍した時両者の差は5になり，誤差によって選択結果は変わらず

1-5 近似のまとめ

- つまりスケールλを調整することで実質的にRUMモデルとMMNLは等価なモデルになる

- だが普通はあえてこの等価性を示すことは無く，誤差項に分布系を仮定することががそのままRUMとの近似と考えることもある
  - RUMの効用にもIID誤差は存在するため
  - 確定項と確率項を分離するのが難しいため
  - 経験的にはスケールをIID誤差項の分散と考える

1-6 近似の例

- Nested Logitとの近似
  - NLは，入れ子内の選択肢間にだけ相関がある
  - MMNLの効用を次のように書く

$$U_{nj} = \alpha' x_{nj} + \mu'_n z_{nj} + \varepsilon_{nj} \quad \cdots (11)$$

α : 固定のパラメータ
μ_n : サンプルnのパラメータで平均値ゼロの確率項
x_{nj} , z_{nj} : サンプルn，選択肢jに関する説明変数の値ベクトル
ε_{nj} : IID(独立で同一な)ガンベル分布

  - 誤差項を次式で考えるとNLと似た階層構造になる

$$\mu'_n z_{nj} = \sum_{k=1}^{K} \mu_{nk} d_{jk} \quad \cdots (12)$$

d_{jk}はjが入れ子内の時1，
それ以外で0のダミー変数とする
μ_{nk}は正規分布$\mathrm{N}(0, \sigma_k)$とする

1-6 近似の例

- 式(4)に式(5)を代入

$$U_{nj} = \alpha' x_{nj} + \sum_{k=1}^{K} \mu_{nk} d_{jk} + \varepsilon_{nj} \quad \cdots (13)$$

μ_{nk} : 正規分布の分散σ_kはkごとに決まる

- NLの効用関数
  - サンプルnが入れ子B_k内の選択肢iを選ぶ

$$U_{nj} = W_{nk} + Y_{nj} + \varepsilon_{nj} \quad , \quad j \in B_k \quad \cdots (14)$$

W_{nk} : 入れ子k内で共通の属性（相関を表す項）
Y_{nj} : 選択肢jに固有の属性
ε_{nj} : IID(独立で同一な)ガンベル分布

2-1 最尤推定法(MLE)

- ロジットモデルのパラメータ推定方法
  - ロジットモデルの選択確率は

$$P_{ni} = \frac{\exp(V_{in})}{\sum_{j=1}^{J} \exp(V_{in})}, i = 1, \ldots, J \qquad \cdots(15)$$

  - データの対数尤度は

$$\ln L = \sum_{n=1}^{N} \sum_{j=1}^{J} d_{in} \ln P_{ni} \qquad \cdots(16)$$

  - これを最大化するようにパラメータを計算する.
    - 未知パラメータで微分して極値0として連立方程式を解く

bin Behavior in Networks Research Unit

2-2 最尤推定法の限界

- しかしプロビットモデルやMMNLのように選択確率に積分が残されている場合は計算が困難

$$P_{ni} = \int L_{ni}(\beta) f(\beta)\, d\beta = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) f(\beta)\, d\beta \quad \cdots (3)$$

MMNLの選択確率（再掲）

- シミュレーションにより近似的に選択確率を求める
  - 繰り返し計算結果を平均したものとして求める
  - 求めた確率からパラメータを推定する
  - シミュレーション最尤推定法という

2-3 シミュレーション最尤推定法(MSLE)

1. $f(\beta|\theta)$からβをR回抽出する.（β^r, $r = 1 \ldots \mathrm{R}$）
   - $f(\beta|\theta)$:パラメータ密度関数
   - θ:fの分散や平均を表すパラメータ
2. β^rを用いて$L_{ni}(\beta)$をR回繰り返し計算する.
3. $L_{ni}(\beta)$の平均値$\hat{P}_{ni} = \frac{1}{R}\sum_{r=1}^{R} L_{ni}(\beta^r)$…(16)を求める
4. シミュレーション対数尤度（SLL）は次式で表される

$$\mathrm{SLL} = \sum_{n=1}^{N}\sum_{j=1}^{J} d_{nj} \ln \hat{P}_{ni}$$

・・・(17)

   - d_{in}:選択者nが選択肢jを選ぶとき1，それ以外は0
   - これを最大化するθ の値を推定する.
   - 計算回数はR×J×Nになると考えられる

bin Behavior in Networks Research Unit

2-3 シミュレーション最尤推定法

- ここで$\hat{P}_{ni}$はP_{ni}の不偏推定量であるが，$\ln\hat{P}_{ni}$は$\ln P_{ni}$ の不偏推定量ではない．
  - 偏り：推定する量（分布）と異なる平均値を持つ統計量（分布）を推定量（分布）として用いてしまう事
- $\ln\hat{P}_{ni}$の偏りはドロー回数（抽出回数R）を増やすほど減っていく．そのためドロー回数は偏りを減らすような条件を満たす数となる．
- ここでドロー（抽出）手法にも何種類かある．

2-4 ドロー(抽出)手法

- ランダムドロー法
  - 正規分布や一様分布から変数を抽出する
  - 独立かつランダムにドローする
  - カバー範囲が狭くなる，多くのドロー回数が必要

- ハルトンシークエンス法（ハルトン数列）
  - 一連のドローが負の相関を持つ素数を用いたアルゴリズム
  - ベースとなる素数を選び，(0,1)区間をその数に分割することを繰り返す.
  - 1/3，2/3，1/9，4/9，7/9，2/9，5/9，8/9
  - 広範囲を早くカバーできる優れた準乱数を発生する.

2-5 サンプルサイズとドロー回数

- サンプルサイズN, ドロー回数Rとする
  - ドロー回数とは, サンプル1個あたりの繰り返し計算数
- RはNによらない固定値とすると, $\ln \hat{P}_{ni}$が不連続になってしまう.
- MSLEがMLE（最尤推定法）と漸近的に等価になる条件は, $\sqrt{N} < \mathrm{R}$である.

3-1 実験結果

- 6肢車種選択(サンプル数4654)にMMNLを適用
  - Sports utility vehicle, Sports car, Station wagon, Truck, Van, EV
- 確定項を21，確率項を4種入れてMLSEで推定
- これを同一の確定項だけを入れたMNLモデルと比較

	MMNL	MNL
確定項	21種	21種
確率項	4種	なし
最終尤度	-7375.34	-7391.83

- 初期尤度は等しいので尤度比は最終尤度が大きい方(＝MMNL)が高くなる→説明力が高い

3-2 人工変数の追加

- MNLにmixingの妥当性を判定するための人工変数を適当に12種類追加する.
  - 人工変数は次の式で定義される. tは選択肢iの属性ベクトルx_i中の要素, Cは選択肢集合とする.

$$z_{ti} = \tfrac{1}{2}(x_{ti} - x_{tC})^2 \quad with \quad x_{tC} = \sum_{j \in C} x_{tj} \cdot L_C(j; \mathbf{x}, \hat{\alpha}) \quad \cdots(18)$$

L：MNLの選択確率

- この変数が有意であれば, 選択要素tはmixingを持つと考えられる（mixingがないという帰無仮説を棄却する）.
  - ただし変数の独立性が無く, 非観測属性間にも相関があると考えられるためmixingがあることを必ずしも保証するわけではない.

3-3 人工変数を加えたMNL推定結果

人工変数	パラメータ	標準誤差	t値
Price/log(income)	0.0019	0.0927	0.02
Range	-0.0349	0.0551	-0.63
Acceleration	-1.3728	2.1388	-0.64
Top speed	-0.2071	0.6383	-0.32
Pollution	0.0977	0.6764	0.14
Size	21.5773	9.5	2.27*
"Big enough"	0.2837	0.3832	0.74
Luggage space	3.8731	3.4638	1.12*
Operating cost	4.2245	0.8369	5.05*
Station availability	0.6741	0.3781	1.78*
EV	2.3476	0.5704	4.12*
CNG	1.2364	0.4798	2.58*
最終尤度	-7356.61	*は|T|>1の変数	

- 尤度比は上昇し，6種類の変数がT＞1．つまりmixingが生じていると考えられる．これを確率項としてMMNLに導入して再び推定する．

3-4 改良MMNL推定結果

- 確定項は同じなので省略する．推定結果は以下．

Random Effects	Parameter Estimates	SE
Non-EV	3.3802	0.7647
Non-CNG	1.1042	0.4990
Size	8.0788	2.7021
Luggage space	7.6220	1.7153
Operating cost	4.4532	0.8014
Station availability	1.3987	0.5730

- 確率項のパラメータから相関はあると考えられる

	MMNL	MNL	改良MMNL
確定項	21種	21種	21種
確率項	4種	なし	6種
最終尤度	-7375.34	-7391.83	-7358.93

- 尤度を比較すると高くなることがわかる
  - Mixingは有効と考えられる．

Behavior in Networks Research Unit

3-5 推定回数による結果の違い

- 2節で述べたようにMLSEで推定すると標準誤差にシミュレーションノイズが影響してしまう
  - 繰り返し計算回数が増えると一般的に誤差は減る
  - しかし普通の誤差$\Delta_N(\theta_N)$では誤差が増えてしまう

- Newey & McFadden(1994)で提案された，Robustな共分散行列の推定値を用いる

$$\Gamma_N(\theta_N)^{-1}\Delta_N(\theta_N)\Gamma_N(\theta_N)^{-1} \quad \cdots(19)$$

bin Behavior in Networks Research Unit

3-6 実験結果（MMNL）

- 50回→250回で通常(Asymptotic)の誤差は増えているが，共分散行列の推定値(Robust)は全部減っている.
- 共分散行列の推定値で誤差を表現するのは有効である.

Table 2: Mixed Logit for Alternative-Fueled Vehicle Choice

Variables	Parameter Estimates	Standard Error 250 replications		Standard Error 50 replications	
		Asymptotic	Robust	Asymptotic	Robust
Price/log(income)	-0.264	0.0435	0.0452	0.0412	0.0525
Range	0.517	0.0581	0.0685	0.0511	0.1022
Acceleration	-1.062	0.1859	0.1990	0.1738	0.2519
Top speed	0.307	0.1150	0.1184	0.1131	0.1188
Pollution	-0.608	0.1392	0.1420	0.1357	0.1546
Size	1.435	0.5082	0.4991	0.4945	0.5156
"Big Enough"	0.224	0.1126	0.1166	0.1113	0.1220
Luggage Space	1.702	0.4822	0.5854	0.4314	0.8971
Operating Cost	-1.224	0.1593	0.2069	0.1393	0.2998
Station availability	0.615	0.1452	0.1536	0.1410	0.1757
Sports utility vehicle	0.901	0.1484	0.1486	0.1482	0.1493
Sports car	0.700	0.1625	0.1513	0.1626	0.1518
Station wagon	-1.500	0.0674	0.0645	0.0674	0.0659
Truck	-1.086	0.0556	0.0520	0.0555	0.0556
Van	-0.816	0.0558	0.0468	0.0557	0.0471
EV	-1.032	0.4249	0.5022	0.3777	0.6035
Commute < 5 & EV	0.372	0.1660	0.1763	0.1608	0.1927
College & EV	0.766	0.2182	0.2374	0.2073	0.2796
CNG	0.626	0.1482	0.1670	0.1391	0.2139
Methanol	0.415	0.1464	0.1474	0.1440	0.1534
College & Methanol	0.313	0.1243	0.1256	0.1223	0.1308
Random Effects					
Non-EV	2.464	0.5414	0.7184	0.4428	1.0252
Non-CNG	1.072	0.3773	0.4109	0.2781	0.5711
Size	7.455	1.8194	2.0408	1.5538	2.4734
Luggage Space	5.994	1.2483	1.6617	1.0483	2.7719
Log Likelihood	-7375.34				

Note: Parameter estimates are from Brownstone & Train (1996); standard error estimates are from this study.

まとめ

- MMNLという概念についての解説を行った
- MMNLの普遍的な近似性を示した
- 推定手法であるMLSEを解説した
- MMNLを用いる妥当性を示した
- 実データを用いた実験から有効性を示した