離散選択モデルと選択肢集合の特定化

愛媛大学
倉内慎也
kurauchi@cee.ehime-u.ac.jp

プレゼンテーションの構成

- 離散選択モデルの概略
- 離散選択モデルのフレームワーク
- 選択肢集合形成過程のモデル化
- おわりに

離散選択と連続選択

- 連続量の選択
  例）活動時間（施設での滞在時間），一週間のトリップ頻度

- 離散選択
  例）ブランドの選択，携帯電話会社

交通分野での選択肢は離散選択が多い

いつ？	出発時刻選択
どこへ？	目的地選択
どの交通手段で？	交通手段選択
どの経路で？	経路選択
どれぐらいの頻度で？	トリップ頻度の選択

合理的選択と効用最大化

合理的選択
　完備性：*{車，鉄道}→（車≧鉄道）and/or（鉄道≧車）*
　推移性：*（車＞バス）&（バス＞鉄道）⇔（車＞鉄道）*

複数の選択肢を選好（望ましさ）の順に並べることができる

例）{A,B,C,D,E}
(A>B), (B>C), (C>D), (D<E), (C>E)　　完備性
↓
(A>B>C>E>D)　　推移性

D　E　C　B　A　→効用

効用最大化：「人は最大の効用を与える選択肢を選ぶ」
Aさん：車を選択⇔U（車）＞U（バス），U（鉄道）

ランダム効用(1)

<u>効用を構成する要因</u> (例)交通手段選択

- **代替案の属性**:料金，所要時間，乗換え回数etc.
- **個人属性**:性別，年齢，免許の有無etc.
- **トリップ属性**:トリップ目的，時間帯etc.

$$
\begin{aligned}
U(car) &= \beta_1 + \beta_3 * time_{car} + \beta_4 * \cos t_{car} + \beta_5 * carown + \varepsilon_{car}\\
U(bus) &= \beta_2 + \beta_3 * time_{bus} + \beta_4 * \cos t_{bus} + \beta_6 * age60 + \varepsilon_{bus}\\
U(rail) &= \beta_3 * time_{rail} + \beta_4 * \cos t_{rail} + \varepsilon_{rail}
\end{aligned}
$$

確定項(V)　　誤差項

<u>分析者にとって意思決定者のもつ真の効用は不明</u>

→ランダム(誤差)項を用いて効用を確率的に表す

ランダム効用(2)

<u>誤差項に含まれるもの</u>

- **非観測属性**:快適性，移動の自由度etc.
- **測定誤差**:駅までのアクセス時間etc.
- **情報の不完全性**:認知所要時間と実際の所要時間のずれetc.
- Instrumental (proxy) variables :「快適性」の代わりに「座席数」を代理変数として用いたときの差異etc.
- **異質性**:所要時間や費用の重み(効用パラメータ)の個人差etc.
- **効用最大化以外の意思決定ルールによる影響**:所要時間や費用の重み(効用パラメータ)の個人差etc.

誤差項の分布とモデル(1)

$$U(car) = V_{car} + \varepsilon_{car}$$
$$U(rail) = V_{rail} + \varepsilon_{rail}$$

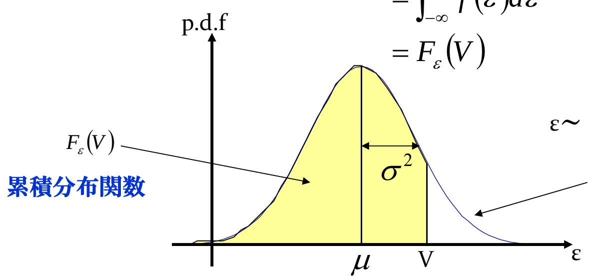
誤差項は確率的に変動
→分析者から見て効用が最大となる選択肢は確率的
→分析者から見た意思決定者の選択行動は確率的

$$
\begin{aligned}
choice = car &\Leftrightarrow U(car) > U(rail)\\
&\Leftrightarrow V_{car} + \varepsilon_{car} > V_{rail} + \varepsilon_{rail}\\
&\Leftrightarrow \varepsilon_{rail} - \varepsilon_{car} < V_{car} - V_{rail}\\
&\Leftrightarrow \varepsilon < V
\end{aligned}
$$

誤差項の分布とモデル(2)

$$
\begin{aligned}
\Pr ob(choice = car) &= \Pr ob(\varepsilon < V)\\
&= \int_{-\infty}^{V} f(\varepsilon) d\varepsilon\\
&= F_\varepsilon(V)
\end{aligned}
$$

$\varepsilon \sim N(\mu, \sigma^2)$



選択確率はεとVに依存

多項プロビットモデル

$$U(car)=\beta X_{car}+\varepsilon_{car}$$
$$U(bus)=\beta X_{bus}+\varepsilon_{bus}$$
$$U(rail)=\beta X_{rail}+\varepsilon_{bus}$$

ε～多変量正規分布

$$P(i)=\int_{\varepsilon_1=-\infty}^{\varepsilon_i+V_i-\varepsilon_1}\cdots\int_{\varepsilon_i=-\infty}^{\infty}\cdots\int_{\varepsilon_J=-\infty}^{\varepsilon_i+V_i-\varepsilon_J}\phi(\varepsilon)d\varepsilon_J\cdots d\varepsilon_1$$

$$\phi(\varepsilon)=\frac{1}{\left(\sqrt{2\pi}\right)^{J-1}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}\varepsilon\Sigma^{-1}\varepsilon'\right)$$

多項ロジットモデル

$$U(car)=\beta X_{car}+\varepsilon_{car}$$
$$U(bus)=\beta X_{bus}+\varepsilon_{bus}$$
$$U(rail)=\beta X_{rail}+\varepsilon_{bus}$$

ε～IIDガンベル
独立で(Independently)
同一(Identically)の分散を持つ
分布(Distributed)

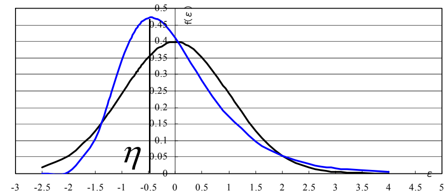


図2.1　正規分布とガンベル分布の確率密度関数

$$V(\varepsilon)=\frac{\pi^2}{6\mu^2}$$

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

多項ロジットモデルと多項プロビットモデル

$$U(car)=V_{car}+\varepsilon_{car}$$
$$U(bus)=V_{bus}+\varepsilon_{bus}$$
$$U(rail)=V_{rail}+\varepsilon_{bus}$$

多項ロジットモデル

- closed-formであるため計算が容易
- 便益計算が簡便

$$P(i)=\frac{\exp(\mu V_i)}{\sum_{j\in C}\exp(\mu V_j)}$$

多項プロビットモデル

- 中心極限定理より誤差項の仮定は尤もらしい
- open-formであるため計算負荷が大きい(J-1重積分)

$$P(i)=\int_{\varepsilon_1=-\infty}^{\varepsilon_i+V_i-\varepsilon_1}\cdots\int_{\varepsilon_i=-\infty}^{\infty}\cdots\int_{\varepsilon_J=-\infty}^{\varepsilon_i+V_i-\varepsilon_J}\phi(\varepsilon)d\varepsilon_J\cdots d\varepsilon_1$$

$$\phi(\varepsilon)=\frac{1}{\left(\sqrt{2\pi}\right)^{J-1}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}\varepsilon\Sigma^{-1}\varepsilon'\right)$$

標準的ロジットモデルのフレームワーク

意思決定タイミング
外生的に与える

$$U_{in}=V_{in}(X_{in};\beta_n)+\varepsilon_{in}(\theta_n)\qquad i\in C_n$$
$$Choice_n=D_n(U_{in})$$

効用の時間的変化
不変

社会的相互作用
なし

確定項の関数型
線形効用関数

説明変数
客観的確定変数

嗜好
母集団で同一

誤差構造
IIDガンベル分布

選択肢集合
所与

意思決定ルール
効用最大化

選択肢集合の特定

- 離散選択モデルは「互いに排反（mutually exclusive）かつ網羅的（collectively exhaustive）な代替案の集合」からの選択を前提（Ben-Akiva & Lerman, 1985）
- 各個人がどのような代替案を選択肢として認知しているのかを分析することが不可欠
- 通常の離散選択モデルは，個人が選択可能な選択肢集合を分析者が確定的に与える必要がある．（Swait & Ben-Akiva, 1986）

$$P(i) = \frac{\exp(\mu V_i)}{\sum_{j \in C} \exp(\mu V_j)}$$

離散選択モデル固有の問題

選択肢集合の問題

- 誤った選択肢集合の特定化はパラメータ推定値にバイアスをもたらす（Swait & Ben-Akiva, 1986）

交通手段	所要時間（分）	費用（円）
A	100	8,000
C	90	5,000

- nさん：所要時間を重視（所要時間のパラメータの影響が大），ただし，交通手段Bが利用不可能→交通手段Cを選択
- 分析者：universal choice setを用いて分析
- コンピュータ：nさんは費用が一番安いからCを選んだ→費用のパラメータを大きく負に推定（バイアス）

選択肢集合形成過程のモデリングアプローチ

網羅的（完全情報）
例）universal choice setを用いたロジットモデル

選択的（不完全情報）

- 暗黙的
  Dial (1971)
- 明示的
  - 確定的（ルールベース）
    D'Este (1997), Hato et al. (2000)
  - 確率的
    - Single stage
      Cascetta et al. (2002)
    - Two stage
      - GEV型
        Swait (2001)
      - Manski型
        Swait et al. (1987)

繰り返し選択状況における確定的アプローチ

- プローブデータの活用
  - ✓ 1回でも利用実績がある経路を選択肢集合とする

- ただし，交通ネットワークの構造やサービスレベルが変化した場合の選択肢集合の特定が課題

確率的選択肢集合モデル

◆Probabilistic Choice Setモデル（Manski, 1977）

例）鉄道（R） v.s. 自動車（C）

$$
\begin{aligned}
P_n(R) &= P_n(R|\{R,C\})\times Q_n(\{R,C\}) \\
&+ P_n(R|\{R\})\times Q_n(\{R\}) \\
&+ P_n(R|\{C\})\times Q_n(\{C\})^{\gamma_0} \\
&+ P_n(R|\{\phi\})\times Q_n(\{\phi\})
\end{aligned}
$$

◆Independent Availabilityモデル（Swait & Ben-Akiva, 1987）

代替案が利用可能性は代替案間で独立であると仮定

$$Q_n(\{R\}) = q_n(R)\times(1-q_n(C))$$

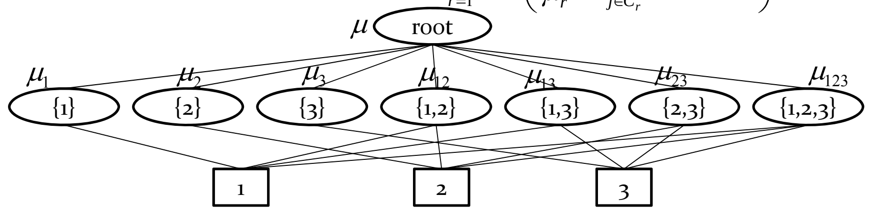
$$q_n(R) = \frac{1}{1+\exp(-Y_n(R))},\quad Y_n(R) = \gamma_0 + \gamma_1 z_{1n} + \gamma_2 z_{2n} + \cdots + \varsigma_{nR}$$

GenL(choice set Generation Logit)モデル(Swait, 2001)

- GEVモデルに則した内生的割り当てモデル：効用が選択肢の利用可能性に影響

$$P(i) = \sum_{k\in K_i} P(i|C_k)Q(C_k)$$

$$P(i|C_k) = \frac{\exp(\mu_k V_i)}{\sum_{j\in C_k}\exp(\mu_k V_j)},\quad Q(C_k) = \frac{\exp\left(\frac{\mu}{\mu_k}\ln\sum_{j\in C_k}\exp(\mu_k V_j)\right)}{\sum_{r=1}^{K}\exp\left(\frac{\mu}{\mu_r}\ln\sum_{j\in C_r}\exp(\mu_r V_j)\right)}$$



- ただし，推定すべきスケールパラメータ数は2^I-1となり不安定

single stageモデル

- Cascetta et al. (2002)モデル

$$U_{in} = V_{in} + \ln Q_{in} + \varepsilon_{in}$$

$$\begin{cases} Q_{in}\to 0 \;\Rightarrow\; \ln Q_{in}\to -\infty \\ Q_{in}\to 1 \;\Rightarrow\; \ln Q_{in}\to 0 \end{cases}$$

- PCSモデルから導出されるモデル (Ben-Akiva & Lerman, 1985)

$$U_{in} = V_{in} - \ln Q_{in} + \varepsilon_{in}$$

誤差項のスケールパラメータを含めれば実証的には問題ない？

ロジットモデルによる利用可能性の考慮

◆選択肢集合Aと選択結果の同時確率

$$P(i,A) = P(A|i)\times P(i) = P(i|A)\times P(A)$$

◆ベイズの定理

$$P(i|A) = \frac{P(A|i)\times P(i)}{\sum_{j\in A}P(A|j)\times P(j)} = \frac{\exp[\mu V_i + \ln P(A|i)]}{\sum_{j\in A}\exp[\mu V_j + \ln P(A|j)]}$$

◆Independent Availabilityを仮定

$$P(A) = \prod_{j\in A} q_j \times \prod_{j\notin A}(1-q_j) \Rightarrow P(A|i) = \prod_{j\in A, j\neq i} q_j \times \prod_{j\notin A}(1-q_j) = \frac{P(A)}{q_i}$$

◆利用可能性を考慮したロジットモデル

$$P(i|A) = \frac{\exp[\mu V_i - \ln q_i]}{\sum_{j\in A}\exp[\mu V_j - \ln q_j]}$$

ただし，Vとqに含まれるパラメータを選択結果のみから推定するのは困難．．．

おわりに

- ネットワーク分析における経路選択肢集合の列挙の問題は積年の課題
  - ✓ 計算コスト，再現性，政策評価（特に情報提供）
- ビッグデータによって様々な展開
  - ✓ 繰り返し観測データ，交通状態データ
- まずは演習で色々と試しましょう