選択行動モデル
プロビット・ロジット・判別分析

Moshe　Ben-Akiva
June 19, 2007
Lecture 1

2010/05/26(水)
M1 柿元淳子

目次

1. ランダム効用モデルの復習
2. 2つ以上の選択肢への拡張
3. プロビットモデル
4. ロジットモデル
5. ロジットモデルの性質
6. 判別分析

1.ランダム効用モデルの復習

- 選択ルール：効用最大化
  -個人nは選択肢C_nのなかでもっとも効用の大きい選択肢を選択する。
- 効用： $U_{in} = V_{in} + \varepsilon_{in}$

  V_{in}：観測可能な変数によって表現された効用

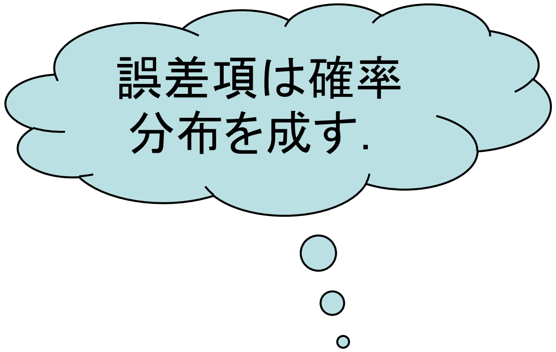
$$V_{in} = \sum_{k}^{K} \beta_k X_{ink}$$

  ε_{in}:誤差項

1.ランダム効用モデルの復習

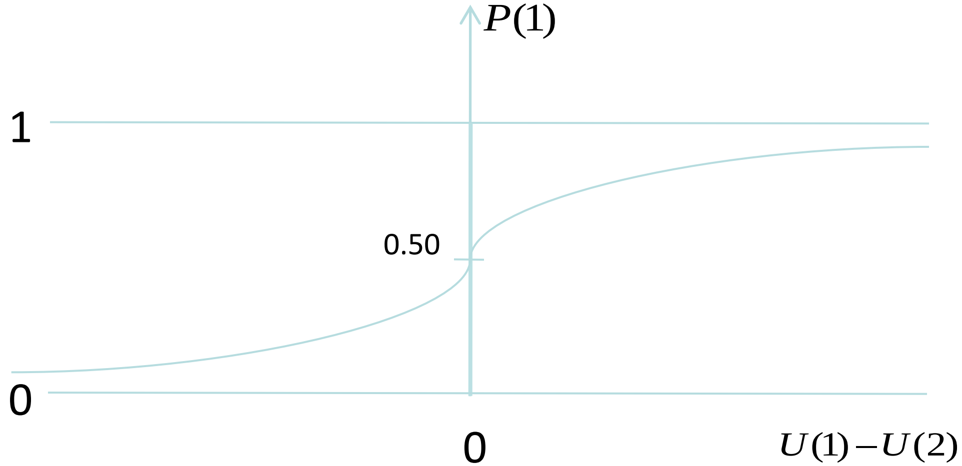
- 選択確率

$$P(i|C_n) = P(U_{in} \geq U_{jn}, \forall j \in C_n)$$
$$= P(U_{in} - U_{jn} \geq 0, \forall j \in C_n)$$
$$= P(U_{in} = \max_j U_{jn}, \forall j \in C_n)$$



- 選択肢2つの場合

$$P(1) = P(U_{1n} \geq U_{2n})$$
$$= P(U_{1n} - U_{2n} \geq 0)$$
$$= P(\varepsilon_{2n} - \varepsilon_{1n} \leq V_{1n} - V_{2n})$$
$$= F_{\varepsilon_2 - \varepsilon_1}(V_{1n} - V_{2n})$$



ε_2-ε_1を変数とする1変数累積分布関数になる。

２．２つ以上の選択肢への拡張

- 選択肢集合C_n : J_n 選択肢, J_n≧２

$$P(i|C_n) = P(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n)$$

$$= P(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}, \forall j \in C_n)$$

２．２つ以上の選択肢への拡張

- 例えば,
  選択肢集合C_n ={1,2,3}∀n

$$P(1)=P(1|C_n)=P(U_{1in}\geq U_{2n}\text{かつ}U_{1in}\geq U_{3n})$$

$$=P(V_{1n}+\varepsilon_{1n}\geq V_{2n}+\varepsilon_{2n}\text{かつ}V_{1n}+\varepsilon_{1n}\geq V_{3n}+\varepsilon_{3n})$$

$$=P(\varepsilon_{2n}-\varepsilon_{1n}\leq V_{1n}-V_{2n}\text{かつ}\varepsilon_{3n}-\varepsilon_{1n}\leq V_{1n}-V_{3n})$$

$$=\boxed{F_{\varepsilon_2-\varepsilon_1,\varepsilon_3-\varepsilon_1}(V_{1n}-V_{2n},V_{1n}-V_{3n})}$$

$$=\int_{-\infty}^{V_{1n}-V_{2n}}\int_{-\infty}^{V_{1n}-V_{3n}} f_{\varepsilon_2-\varepsilon_1,\varepsilon_3-\varepsilon}(q_1,q_2)dq_1dq_2$$

ε_2-ε_1とε_3-ε_1を変数とする2変数累積分布関数

- ε= (ε_1 , ε_2,…, ε_J)に対する（同時）確率分布の仮定によって選択行動モデルは異なる。

3. プロビットモデル

- プロビットモデル：εが多変量正規分布に従う、つまり、

$$\varepsilon \sim MVN(0,\Sigma),\quad f(\varepsilon)=(2\pi)^{-\frac{J}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{2}{2}(\varepsilon'\Sigma^{-1}\varepsilon)}$$

ここで**Σ**は分散-共分散行列で、

誤差項が平均0で,
共分散行列に従うと
する.

$$\Sigma=\begin{bmatrix} Var(\varepsilon_1) & Cov(\varepsilon_1,\varepsilon_2) & \cdots & Cov(\varepsilon_1,\varepsilon_J) \\ Cov(\varepsilon_2,\varepsilon_1) & Var(\varepsilon_2) & \cdots & Cov(\varepsilon_2,\varepsilon_J) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\varepsilon_J,\varepsilon_1) & Cov(\varepsilon_J,\varepsilon_2) & \cdots & Var(\varepsilon_J) \end{bmatrix}$$

である。

- プロビットモデルの性質
-柔軟な代入パターン？(flexible substitution pattern)
-ランダムな係数に拡張（ランダムな異質性を考慮できる）
-連続的に相関関係のあるパネルデータに使える。

3つの選択肢を持つプロビットモデル

- モデル：

$$U_{1n} = V_{1n} + \varepsilon_{1n}$$
$$U_{2n} = V_{2n} + \varepsilon_{2n} \quad \text{ここで} \quad \begin{pmatrix} \varepsilon_{1n} \\ \varepsilon_{2n} \\ \varepsilon_{3n} \end{pmatrix} \sim N\left(0, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}\right)$$
$$U_{3n} = V_{3n} + \varepsilon_{3n}$$

- 選択肢1との効用の差

$$U_{1n} - U_{2n} = V_{1n} - V_{2n} - (\varepsilon_{2n} - \varepsilon_{1n})$$
$$U_{1n} - U_{3n} = V_{1n} - V_{3n} - (\varepsilon_{3n} - \varepsilon_{1n})$$

- 選択確率

$$P_n(1) = P(\varepsilon_{2n} - \varepsilon_{1n} \le V_{1n} - V_{2n} \text{かつ} \varepsilon_{3n} - \varepsilon_{1n} \le V_{1n} - V_{3n})$$

$$\begin{bmatrix} \varepsilon_2 - \varepsilon_1 \\ \varepsilon_3 - \varepsilon_1 \end{bmatrix} \sim N\left(0, \Sigma = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \sigma_1^2 + \sigma_{23} - \sigma_{12} - \sigma_{13} \\ \sigma_1^2 + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_1^2 + \sigma_3^2 - 2\sigma_{13} \end{bmatrix}\right)$$

$$P_n(1) = \int_{-\infty}^{V_{1n} - V_{2n}} \int_{-\infty}^{V_{1n} - V_{3n}} n(q;0\Sigma) dq_1 dq_2$$

3つの選択肢を持つプロビットモデル

- 一般的な場合の選択肢1を選択する確率は

$$P_n(1) = \underbrace{\int_{-\infty}^{V_{1n}-V_{2n}} \int_{-\infty}^{V_{1n}-V_{3n}} \cdots \int_{-\infty}^{V_{1n}-V_{Jn}}}_{\text{J-1次元積分}} n(q;0,\Sigma)dq$$

- $n(q;0\Sigma)$

  :平均値0で分散-共分散マトリックスΣの多変量正規分布

4．ロジットモデル

- 選択肢集合： $C_n = \{1,2,3\} \forall n$

$$P(1) = F_{\varepsilon_2 - \varepsilon_1, \varepsilon_3 - \varepsilon_1}(V_{1n} - V_{2n}, V_{1n} - V_{3n})$$

$$= \frac{1}{1 + e^{-\mu(V_{in} - V_{2n})} + e^{-\mu(V_{in} - V_{3n})}}$$

$$= \frac{e^{\mu V_{1n}}}{e^{\mu V_{1n}} + e^{\mu V_{2n}} + e^{\mu V_{3n})}}$$

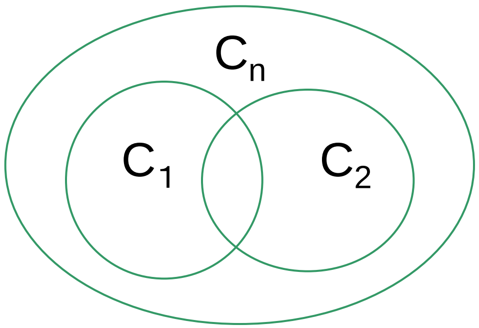
5．ロジットモデルの性質

- IIA特性：相関のない選択肢からの独立
(Independence from Irrelevant Alternatives)

$$P(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{in}}}$$

- オッズ比：

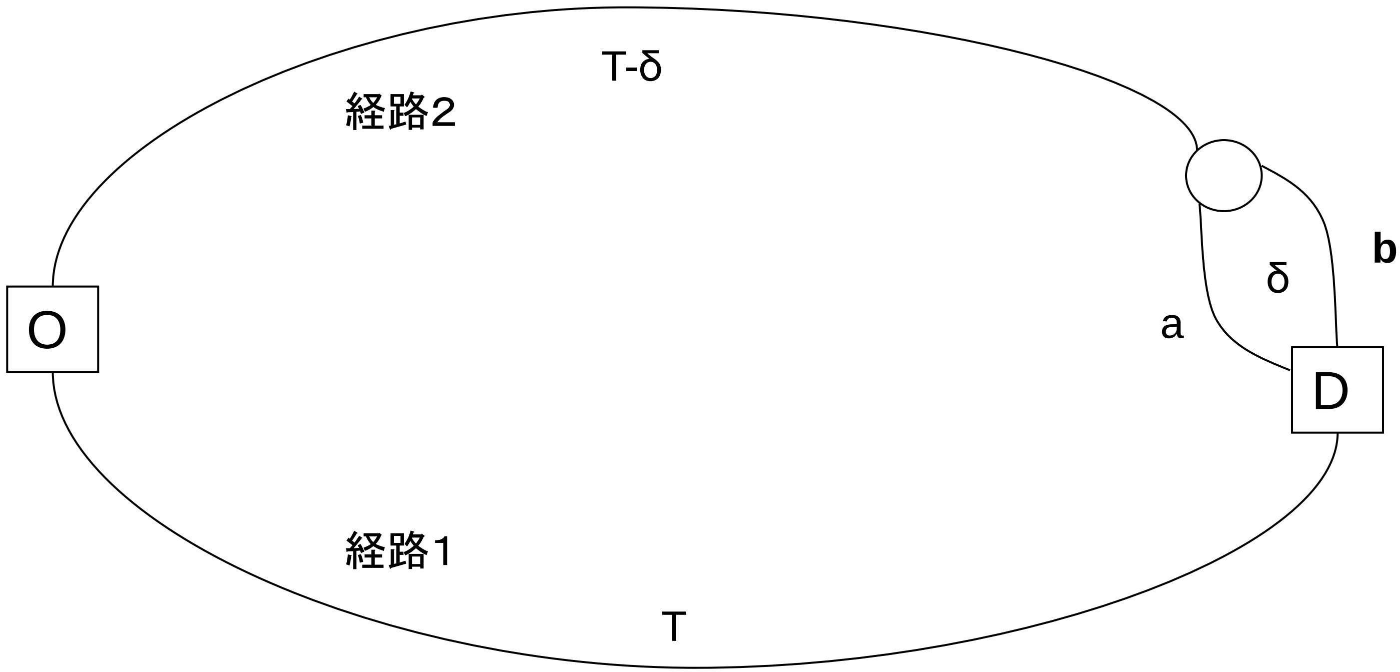
$$\frac{P(i|C_{1n})}{P(j|C_{1n})} = \frac{P(i|C_{2n})}{P(j|C_{2n})}$$



$i, j \in C_{1n}, i, j \in C_{2n}, C_{1n} \subseteq C_n$かつ$C_{2n} \subseteq C_n$

５．IIA特性の例

- 重複するセグメントを持つ経路選択



$$P(1|\{1,2a,2b\}) = P(2a|\{1,2a,2b\}) = P(2b|\{1,2a,2b\}) = \frac{e^{\mu T}}{\sum_{j \in \{a,2a,2b\}} e^{\mu T}} = \frac{1}{3}$$

５．赤バス-青バス問題

- 同等の効用を持つ乗用車とバス

$$C = \{auto, bus\} \text{かつ} V_{auto} = V_{bus} = V$$

$$P(auto) = P(bus) = 1/2$$

ここで、現状のバスと塗装以外まったく同質な、新しいバスサービスの導入を考える。

$$C = \{auto, red, blue\} \text{かつ} V_{auto} = V_{red} = V_{blue}$$

ロジットモデルによると、

$$P(auto) = P(red) = P(blue) = 1/3$$

であるが、直観的には、

$$P(auto) = 1/2, P(red) = P(blue) = 1/4$$

5. IIA特性と集団

- 母集団を同じ大きさの2つの集団に分割する。一方は自家用車を好み、一方は公共交通を好む。
- 青バス導入前の各交通手段のシェア

集団	自家用車利用率	赤バス利用率	
自家用車を好む	90%	10%	P(auto)/P(red)=9
公共交通を好む	10%	90%	P(auto)/P(red)=1/9
合計	50%	50%	

- 青バス導入後（自家用車と赤バスの利用確率の比は変わらないとする）

集団	自家用車利用率	赤バス利用率	青バス利用率
自家用車を好む	81.8%	9.1%	9.1%
公共交通を好む	5.2%	47.4%	47.4%
合計	43.5%	28.25%	28.25%

→Nested Logit Modelの必要性。

6. 判別分析

- $f(i,x)$ =同時生起確率:従属変数と独立変数の密度

  x（ex.男女、学歴etc.）のとき、選択肢iを選ぶ

- 離散選択モデル: $f(i,x) = P(i|x)f(x)$
  - $P(i|x)$:選択モデル（iは離散的）
  - $f(x)$:集合におけるxの密度（xは連続または離散的）

- 判別分析: $f(i,x) = Q(x|i)w(i)$
  - $Q(x|i)$:iによって定められる部分母集団におけるxの確率分布
    たとえば=N(μ_i, Σ)
  - $w(i) = \int f(i,x)dx$

選択肢iを選択した人が、
xという母集団に属している確率は？

全体で選択肢i
を選ぶ確率

６．判別分析

- $Q(x|i)$ と $w(i)$が与えられたとき、$P(i|x)$は以下のようになる。

$$P(i|x) = \frac{Q(x|i)w(i)}{f(x)} = \frac{Q(x|i)w(i)}{\sum_j Q(x|j)w(j)}$$

$$= \frac{e^{\ln Q(x|i) + \ln w(i)}}{\sum_j e^{\ln Q(x|j) + \ln w(j)}}$$

- $Q(x|i) = n(\mu_i, \Sigma)$ のときは、

$$P(i|x) = \frac{e^{a_i + \beta'_i X}}{\sum_j e^{aj + \beta'_j X}}$$