

半順序に基づいた非集計モデルの推定法

鉄道総合技術研究所 野末 尚次 (Naotugu Nozue)

1 はじめに

意識データ (stated preference data) に基づいた非集計型モデルの推定が研究されているが、現在の主流は、代替案の順序付けに基づいた方式 (Ranked Logit) である。

しかし、意識データは未体験事項に対する質問であるため、多数の代替案の間に完全な順序を設定することは被験者にとって非常に困難であり、不正確な調査データとなりやすい。

この問題を解決するために、被験者に対して完全な順序付けを要求するのではなく、相対的な評価を要求することにより得られたデータから、Logitモデルを推定する方式 (Poset Logit) を開発した。

2 全順序と半順序

複数の代替案群 (例えば {A, B, C, D, E}) に対して Ranked Logit では次のような問をする。

[問 I] 次の代替案に対して、望ましい順に番号を記入して下

さい。

A:□位、B:□位、C:□位、D:□位、E:□位

この様な問は、代替案の全順序を規定することになる。

A:1位
B:5位
C:2位
D:4位
E:3位

Ⓐ → ⓒ → ⓔ → ⓓ → Ⓑ

(回答例)

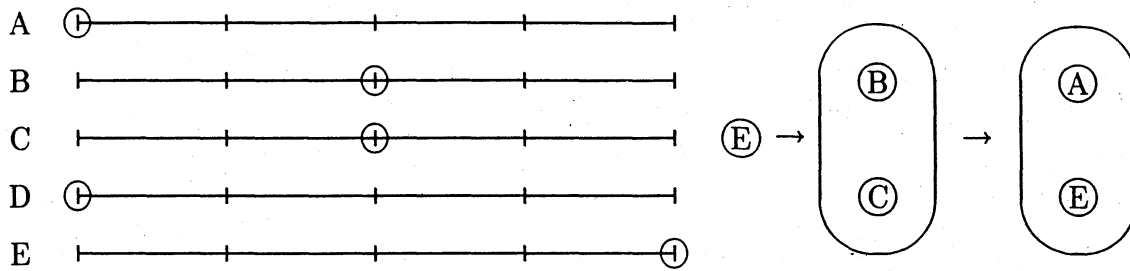
(束構造)

上記の質問は全体の順序付けを要求しているため非常に回答が難しい。同じ様な代替案もあるため、次の様な設問を考える。

[問Ⅱ] 次の代替案に対して、Cを基準とした場合の評価を記入して下さい。

非常に悪い	悪い	同じ位	良い	非常に良い
A:	<div style="position: absolute; top: -5px; left: 10%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 30%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 50%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 70%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div>			
B:	<div style="position: absolute; top: -5px; left: 10%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 30%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 50%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 70%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div>			
C:	<div style="position: absolute; top: -5px; left: 10%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 30%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 50%; width: 10%; border-left: 1px solid black; border-right: 1px solid black; text-align: center;">⊕</div> <div style="position: absolute; top: -5px; left: 70%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div>			
D:	<div style="position: absolute; top: -5px; left: 10%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 30%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 50%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 70%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div>			
E:	<div style="position: absolute; top: -5px; left: 10%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 30%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 50%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div> <div style="position: absolute; top: -5px; left: 70%; width: 10%; border-left: 1px solid black; border-right: 1px solid black;"></div>			

後者のタイプの問いに対しては、アンケート結果は全順序とならず、次の様な半順序となる。



(回答例)

(束構造)

さらに、次の様な完全一対比較を求める調査ではより複雑な束構造となる。

[問Ⅲ] 次の問いの□の中に、下記の記号から適当なものを選んで入力して下さい。

0:比較できない 1:劣っている 2:同じ 3:優れている

A案の方がB案より

C案より

D案より

E案より

B案の方がC案より

D案より

E案より

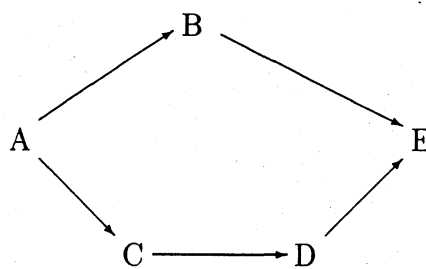
C案の方がD案より

E案より

D案の方がE案より

	B	C	D	E
A	3	3	3	3
B		0	0	3
C			3	3
D				3

(回答例)



(束構造)

従来の推定方式は、設問 I の全順序を前提とした Ranked Logit であり、上記の設問 II や III に対しては Logit モデルの構成ができなかった。

今回、前記のようなアンケート調査結果の半順序構造の発生確率を簡便に計算する方式を開発し、Logit モデルの構築を可能とした。

3 半順序構造に対する発生確率(設問 II 型)

複数の代替案の中から、確率効用モデルにより選択が行われる場合を考える。

各代替案の効用 U_i は、計測可能な効用 V_i と各人に固有の観測外の確率的な効用 ε_i からなる。

$$U_i = V_i + \varepsilon_i \quad : \varepsilon_i \in 2 \text{重指数分布 } (F(y) = e^{-e^{-y}})$$

実際の選択では、「最大の効用を持つ代替案 i_0 が選択される」と仮定する。

$$U_{i0} = \text{Max}(U_i)$$

いま、このモデルに対する各種の確率を計算する。

$$U_{max} = \text{Max}(U_1, \dots, U_n)$$

$$U_{min} = \text{Min}(U_1, \dots, U_n)$$

計算は、次の公式を繰り返し利用することにより簡略化される。

$$F(y+a) = F(y)e^{-a} \quad (1)$$

[補題 1]

$$P\{U_{max} < t\} = \prod_{i=1}^n F(t - V_i) = \prod_{i=1}^n F(t)e^{V_i} = F(t)^{\sum e^{V_i}} \quad (2)$$

$$P\{U_{min} > t\} = \prod_{i=1}^n \{1 - F(t - V_i)\} = \prod_{i=1}^n \{1 - F(t)e^{V_i}\} \quad (3)$$

$$P\{U_{min} < t\} = 1 - \prod_{i=1}^n \{1 - F(t)e^{V_i}\} \quad (4)$$

[補題 2]

$$P\{t|U_t = U_{max}\} = \frac{e^{V_t}}{\sum e^{V_i}} \quad (5)$$

[補題 3] $S = \{1, \dots, n\}$ とおくとき、

$$P\{t|U_t = U_{min}\} = \sum_{J \subseteq S - \{t\}} (-1)^{|J|} \frac{e^{V_t}}{\sum_{k \in J} e^{V_k} + e^{V_t}} \quad (6)$$

(証明)

$$t \text{ が } k \in S - \{t\} \text{ に負けない確率 } \frac{e^{V_t}}{e^{V_k} + e^{V_t}}$$

$$t \text{ が } k, l \in S - \{t\} \text{ に負けない確率 } \frac{e^{V_t}}{e^{V_k} + e^{V_l} + e^{V_t}}$$

従って、principle of inclusion and exclusion により

$$\begin{aligned}
 & P\{t \text{ が } S - \{t\} \text{ の少なくとも1つに勝つ}\} \\
 &= \sum_{k \in S - \{t\}} \frac{e^{V_t}}{e^{V_k} + e^{V_t}} - \sum_{k,1} \in S - \{t\} \frac{e^{V_t}}{e^{V_k} + e^{V_1} + e^{V_t}} + \cdots \\
 &= \sum_{\substack{J \subseteq S - \{t\} \\ J \neq \phi}} (-1)^{|J|+1} \frac{e^{V_t}}{\sum_{k \in J} e^{V_k} + e^{V_t}}
 \end{aligned}$$

$$\begin{aligned}
 P\{t|U_t = U_{\min}\} &= 1 - P\{t \text{ が } S - \{t\} \text{ の少なくとも1つに勝つ}\} \\
 &= \sum_{J \subseteq S - \{t\}} (-1)^{|J|} \frac{e^{V_t}}{\sum_{k \in J} e^{V_k} + e^{V_t}}
 \end{aligned}$$

[補題 4] $T = \{1, \dots, n\}$ とおくとき、

$$P\{U_{\max} \leq t_1, U_{\min} \leq t_2\} = F(t_1)^{K_T} - \prod_{i=1}^n \{F(t_1)^{e^{V_i}} - F(t_2)^{e^{V_i}}\} \quad (7)$$

但し、 $K_T = \sum_{t \in T} e^{V_t}, t_1 \geq t_2$

(証明)

$$\begin{aligned}
 & P\{t_1 \geq U_{\max}, U_{\min} \geq t_2\} \\
 &= \prod_{i=1}^n P(t_1 \geq U_i \geq t_2) \\
 &= \prod_{i=1}^n \{F(t_1 - v_i) - F(t_2 - v_i)\} \\
 &= \prod_{i=1}^n \{F(t_1)^{e^{v_i}} - F(t_2)^{e^{v_i}}\}
 \end{aligned}$$

$$\begin{aligned}
 \therefore P\{U_{\max} \leq t_1, U_{\min} \leq t_2\} \\
 &= P\{t_1 \geq U_{\max}\} - P\{t_1 \geq U_{\max}, U_{\min} \geq t_2\} \\
 &= F(t_1)^{K_T} - \prod_{i=1}^n \{F(t_1)^{e^{V_i}} - F(t_2)^{e^{V_i}}\}
 \end{aligned}$$

[補題 5] $T = \{t_1, \dots, t_n\}, S = \{s_1, \dots, s_n\}, T \cap S = \phi$ とおくとき、

$$\begin{aligned}
 & P\{U_t \geq U_s, \forall t \in T, \forall s \in S\} \\
 &= \sum_{J \subseteq T} (-1)^{|J|} \frac{K_s}{K_s + \sum_{j \in J} e^{V_j}}, \text{ 但し、 } K_s = \sum_{k \in S} e^{V_k} \quad (8)
 \end{aligned}$$

(証明)

$$P\{t \text{ が } S \text{ の max に負ける}\} = \frac{K_s}{K_s + e^{V_t}}, t \in T$$

$$P\{t, r \text{ が } S \text{ の max に負ける}\} = \frac{K_s}{K_s + e^{V_t} + e^{V_r}}, t, r \in T$$

⋮

$$\begin{aligned} & P\{T \text{ のいずれかが } S \text{ の max に負ける}\} \\ &= \sum_{t \in T} \frac{K_s}{K_s + e^{V_t}} - \sum_{\substack{t, r \in T \\ t > r}} \frac{K_s}{K_s + e^{V_t} + e^{V_r}} + \dots \\ &= \sum_{\substack{J \subseteq T \\ J \neq \emptyset}} (-1)^{|J|+1} \frac{K_s}{K_s + \sum_{j \in J} e^{V_j}} \end{aligned}$$

$$\begin{aligned} & P\{T \text{ の min } \geq S \text{ の max}\} \\ &= 1 - P\{T \text{ のいずれかが } S \text{ の max に負ける}\} \\ &= \sum_{J \subseteq T} (-1)^{|J|} \frac{K_s}{K_s + \sum_{j \in J} e^{V_j}} \end{aligned}$$

この公式により2分割の場合の確率は容易に計算が可能となったが、複雑な束構造を分解して確率を計算するために、次のExplosion公式を利用する。

[定理 6: Explosion formula]

$$\begin{aligned} & P\{U_1 \geq U_2 \geq \dots \geq U_n\} \\ &= \prod_{i=1}^{n-1} \frac{e^{V_i}}{\sum_{j=i}^n e^{V_j}} \end{aligned} \tag{9}$$

[定理 7: 順位の独立性]

$$\begin{aligned} & P\{U_k \geq U_{k+1} \geq \dots \geq U_n | U_1 \geq \dots \geq U_{k-1} \geq U_j, j = k, \dots, n\} \\ &= P\{U_k \geq U_{k+1} \geq \dots \geq U_n\} \end{aligned} \tag{10}$$

この定理を利用することにより、次の分割公式を証明することができる。

[定理 8: 分割公式]

$S = S_1 + \cdots + S_m$ とする。

$$\begin{aligned} & P\{U_{i_j} \geq U_{i_{j+1}}, \forall i_j \in S_j, j = 1, \dots, m-1\} \\ &= P\{U_1 \geq U_j, \forall i \in S_1, \forall j \in \overline{S_1}\} \\ &\quad \times P\{U_{i_j} \geq U_{i_{j+1}}, \forall i_j \in S_j, j = 2, \dots, m-1\} \end{aligned} \quad (11)$$

但し、 $\overline{S_1} = S - S_1$

(証明) S_i の全ての置換を $\pi(S_i)$ とする。

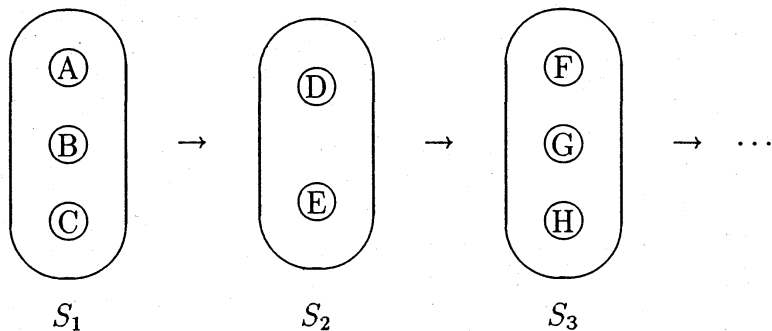
$$\begin{aligned} & P\{U_{i_j} \geq U_{i_{j+1}}, \forall i_j \in S_j, j = 1, \dots, m-1\} \\ &= \sum_{\rho_1 \in \pi(S_1)} \cdots \sum_{\rho_m \in \pi(S_m)} P\{U_{\rho_1(1)} \geq U_{\rho_1(2)} \geq \cdots \geq U_{\rho_1(n_1)} \geq U_{\rho_2(1)} \\ &\quad \geq U_{\rho_2(2)} \geq \cdots \geq U_{\rho_m(1)} \geq \cdots \geq U_{\rho_m(n_m)}\} \\ &= \sum_{\rho_1 \in \pi(S_1)} \cdots \sum_{\rho_m \in \pi(S_m)} P\{U_{\rho_1(1)} \geq \cdots \geq U_{\rho_1(n_1)} \geq U_k, \forall k \in \overline{S_1}\} \\ &\quad \times P\{U_{\rho_2(1)} \geq \cdots \geq U_{\rho_m(n_m)} | U_{\rho_1(1)} \geq \cdots \geq U_{\rho_1(n_1)} \geq U_k, \forall k \in \overline{S_1}\} \\ &= \sum_{\rho_1 \in \pi(S_1)} \cdots \sum_{\rho_m \in \pi(S_m)} P\{U_{\rho_1(1)} \geq \cdots \geq U_{\rho_1(n_1)} \geq U_k, \forall k \in \overline{S_1}\} \\ &\quad \times P\{U_{\rho_2(1)} \geq \cdots \geq U_{\rho_m(n_m)}\} \quad (\because \text{Explosion formula}) \\ &= \left[\sum_{\rho_1 \in \pi(S_1)} P\{U_{\rho_1(1)} \geq \cdots \geq U_{\rho_1(n_1)} \geq U_k, \forall k \in \overline{S_1}\} \right] \\ &\quad \times \left[\sum_{\rho_2 \in \pi(S_2)} \cdots \sum_{\rho_m \in \pi(S_m)} P\{U_{\rho_2(1)} \geq \cdots \geq U_{\rho_m(n_m)}\} \right] \\ &= P\{U_i \geq U_j, \forall i \in S_1, \forall i \in \overline{S_1}\} \\ &\quad \times P\{U_{i_j} \geq U_{i_{j+1}}, \forall i_j \in S_j, j = 2, \dots, m-1\} \end{aligned}$$

この分割方式を利用することにより、束を分解して発生確率を計算することが可能である。

[定理 9] 設問 II 型で $S = S_1 + \cdots + S_m$ に分解される場合の確率は次式となる。

$$\begin{aligned} & P\{U_{i_j} \geq U_{i_{j+1}}, \forall i_j \in S_j, j = 1, \dots, m-1\} \\ &= \prod_{j=1}^{m-1} \left\{ \sum_{L \subseteq S_j} (-1)^{|L|} \frac{K_{j+1}^*}{K_{j+1}^* + \sum_{p \in L} e^{V_p}} \right\} \end{aligned} \quad (12)$$

但し、
$$K_j^* = \sum_{i=j}^m \sum_{p \in S_i} e^{V_p}$$



この定理により、設問IIのグループ間の順位付けが与えられた場合の発生確率が計算できる。

4 一般的な半順序に対する発生確率(設問III型)

設問IIIの一般的な半順序に対する計算を考える。

与えられた束 S を次の形式に分割する。

$$S = S_1 + \cdots + S_m$$

$$\forall x \in S_i, \forall y \in S_{i+1} \quad x \succ y \quad (i = 1, \dots, m-1)$$

いま $R_i = \cup_{k=i}^m S_k,$

$$V_i^* = \ln \left[\sum_{k \in R_i} e^{V_k} \right]$$

とおく。

この時、各 $S_i (i = 1, \dots, m-1)$ に対して、最小元として効用 V_{i+1}^* の元を付加した束を S_i^* とする。

[定理 10] 束 S が上に述べた分割を持つ時、束 S の発生確率 $P(S)$ は次式で計算される。

$$P(S) = \prod_{i=1}^m P(S_i^*)$$

但し、 $P(S_i^*)$ は束 S_i^* の発生確率。

証明は定理 9 と同様なので省略する。

この定理に従って分割された最小単位の半順序 S_i^* に対する発生確率を束構造、又は、対応した有向グラフから簡潔に表現することは困難なため、数え上げにより計算を行う。

いま、束 S から全順序 $N = \{1, \dots, |S|\}$ の上への準同型写像 ρ の集合 Γ を考える。

$$\forall s_1, s_2 \in S, s_1 \succ s_2 \longrightarrow \rho(s_1) \geq \rho(s_2)$$

このとき、発生確率 P_s は次式で与えられる。

$$P_s = \sum_{\rho \in \Gamma} P\{U\rho^{-1}(1) \geq U\rho^{-1}(2) \geq \dots \geq U\rho^{-1}(n)\}$$

従って、 Γ をシステマティックに生成する方法を見いだせば良い。

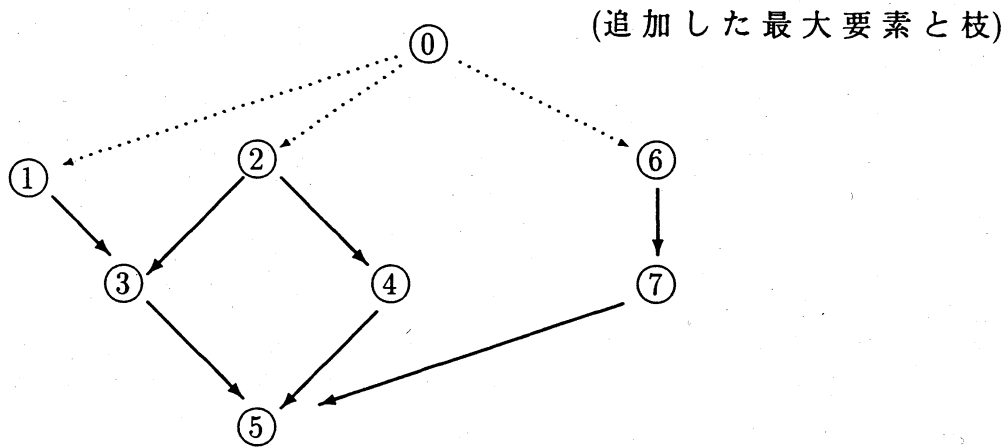
まず、束 $S = \{1, \dots, n\}$ に対して、最大要素 0 を追加する。

この 0 を追加した束のグラフ $G(S)$ に対して、次のアルゴリズムが成立する。

$$n = |S|$$

Suc(i) = ノード i の successor のノード集合

$\text{Pred}(i)$ = ノード i の predecessor の ノード 集合



$\text{Suc}(2) = \{3, 4\}$

$\text{Pred}(3) = \{1, 2\}$

[Γ の生成アルゴリズム]

step 1: (初期化)

$\text{stack1} = \phi.$

$\text{stack2}[i] = \phi. \quad (i = 0, \dots, n).$

$\text{link}[j] = |\text{Pred}(j)| \quad (j = 0, \dots, n).$

Push 0 into stack1 and stack2 [0].

$i = 0$

step 2: (新ノードの選択)

if $\text{stack2}[i] = \phi$ then go to step 5.

$s_1 = \text{Pop}(\text{stack2}[i]).$

Push s_1 into stack1.

if $i = n$ then go to step 4.

step 3: (極大ノード抽出)

```

 $i = i + 1$ 
copy stack2[ $i - 1$ ] to stack2[ $i$ ]
for each  $s_2$  in Suc( $s_1$ ).
    link [ $s_2$ ] = link [ $s_2$ ] - 1
    if link [ $s_2$ ] = 0 then
        Push  $s_2$  into stack2[ $i$ ].
end.
go to step2.

```

step 4: (生成出力)

```

print stack1.
go to step 2.

```

step 5: (バックトラック)

```

if  $i = 0$  then stop.
 $s_3 = \text{Pop}(\text{stack1})$ 
for each  $s_4$  in Suc( $s_3$ )
    link [ $s_4$ ] = link [ $s_4$ ] + 1
end.
 $i = i - 1$ 
go to step2.

```

5 パラメータ推定法

アンケート調査等により意識データが収集されると、それに基づいた効用関数のパラメータ推定が、前述の発生確率により最尤推定法で可能となる。

いま、個人*i*の第*j*代替案の特性値ベクトルを

$$X^{(i)j} = (1, X_{j1}^{(i)}, \dots, X_{jn}^{(i)}) \quad j = 1, \dots, m$$

パラメータベクトルを

$$\beta = (\beta_0, \beta_1, \dots, \beta_n)$$

とすれば、

$$U_j^{(i)} = \beta' X_j^{(i)} = \beta_0 + \beta_1 X_{j1}^{(i)} + \dots + \beta_n X_{jn}^{(i)}$$

$$P(U_k^{(i)} = \max(U_1^{(i)}, \dots, U_m^{(i)})) = \frac{e^{\beta' X_k^{(i)}}}{\sum_j e^{\beta' X_j^{(i)}}}$$

となる。

個人毎の発生確率は、設問Iの場合には 定理6、設問IIの場合には定理9、設問IIIの場合には定理10に従って計算される。

従って、各個人の発生確率の積により尤度関数が求まる。

6 おわりに

従来の全順序に基づくアンケート調査から、今回提案した半順序に基づく方式への拡張により、非集計型モデルの適用範囲の拡大、及び、推定の信頼度の向上が可能となった。

このモデルを利用して北海道の札幌～千歳空港間に超電導磁気浮上鉄道を建設した場合のアンケート調査に基づくシェア

の推定を行なっているが、従来の方式に比べて安定した推定結果が得られている。

今後の課題としては定理10で分割された部分束の発生確率の計算を束構造から導出すること、及び、このモデルの実際問題への適用による有効性の確認がある。

参考文献

- [1] 屋井鉄雄 オペレーションズ・リサーチ, vol 38, no 2, pp73-78, 1993
- [2] Ogawa Marketing Science, vol 6 no 1, pp 66-81, 1987
- [3] Ben-Akiba Journal of Business Research 23, pp253-268, 1991
- [4] Chapman Journal of Marketing Research, vol XIX pp 288-292, 1982