

資料

集合論的潜在特性モデルの応用**

椎 名 乾 平*

AN APPLICATION OF SET THEORETIC LATENT TRAIT MODEL

Kanpei SHIINA

The set theoretic latent trait model (STLTM) proposed by Shiina (1990) is applied to Bergan's (1988) math-items data. The MSTLTM, a marginalized version of STLTM, which can be considered as a latent class model in which response probabilities are reparametrized by a set function of subject set structure and item set structure, is proposed. In applying the model, it is argued that : 1) in some situation the MSTLTM and the latent class model are equivalent, but the former utilizes a convenient graphical representation for items and subjects which will be helpful for explorative model construction, and 2) a major advantage of the STLTM lies in that it permits a statistical comparison among models by assuming various developmental paths.

Key words : Latent class analysis, Item response theory, abilities, skills, developmental paths

椎名 (1990) の集合論的潜在特性モデル (Set Theoretic Latent Trait Model, 以下 STLTM と略す) では, 被験者はおのものが持つ (離散的) 能力の集合として, また各テスト項目はそのテスト項目に正答するのに必要な (離散的) 能力の集合として表現される。被験者を表現する集合の集合 (集合族) は, 包含関係によって半順序集合となり, また項目を表現する集合も同様の半順序集合となる。そして, 被験者のある項目に対する正答確率は被験者を表現する集合と項目を表現する集合との確率的集合関数によって決定されるとする。

本稿の目的は集合論的潜在特性モデルの有効性を, Bergan のデータについて検証することである。この

データの詳細については, Bergan (1988), Bergan & Stone (1985), Bergan, Stone, & Feld (1984) を参照されたい***。

Bergan のデータ

Bergan (1988) のデータは, 子供に 4 つの算数の問題 (A, B, C, D と書くことにする) を解かせその正答・誤答のパターンを注目したものである (TABLE 1)。問題は 4 問あるので可能なパターンは 16 となり, 各パターンを示した被験者数が左端の列に表示されている。下位問題の具体的内容は

A : 数字「3」が同定できる。

B : 数字「4」が同定できる。

C : 数字「3」と 3 つのブロックが対応づけられる。

D : 数字「4」と 4 つのブロックが対応づけられる。

である。Bergan (1988) はこのデータに対して, パラメー

* 日本学術振興会特別研究員 (早稲田大学)
(Fellowships for Japanese Junior Scientists)
(Department of Psychology, School of Letters, Waseda University)

** 本研究は早稲田大学情報科学研究教育センター 昭和63年度特定研究 (A)「認知・運動機能障害検査のコンピューター化」(研究代表者 早稲田大学教授 西本武彦) の援助をうけた。

*** なお, 問題には Set 1 と Set 2 があり, 各 set が 4 つの下位問題より形成されている。本稿では Set 1 のみを扱う。

TABLE 1 Bergan のデータ (Bergan, 1988, p.240, Tab. 1
より引用)

回答パターンの人数	問 題			
	A	B	C	D
197	1	1	1	1
19	0	1	1	1
15	1	0	1	1
38	0	0	1	1
25	1	1	0	1
15	0	1	0	1
15	1	0	0	1
30	0	0	0	1
30	1	1	1	0
4	0	1	1	0
13	1	0	1	0
34	0	0	1	0
43	1	1	0	0
13	0	1	0	0
30	1	0	0	0
71	0	0	0	0

ターに制約を加えた様々な潜在クラスモデルを当てはめてその優劣を比較している****。そこで、STLTMの当てはまりをそれらのモデルと比較してみると、そのような比較の過程を通じて STLTM の特性を明らかにするのが本稿の目的である。なお、Bergan の研究ではモデル比較に尤度比検定を用いているが、本研究では AIC(坂元, 石黒, 北川, 1983)を用いることにする。

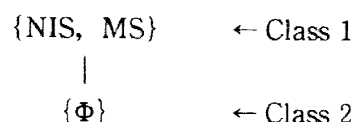
Bergan のデータに対する様々なモデル

Bergan の方法も STLTM も、課題の分析を通じて被験者の反応のモデルを仮定することから始まる。まず Bergan が提案した3つのモデルについて簡単に紹介する。

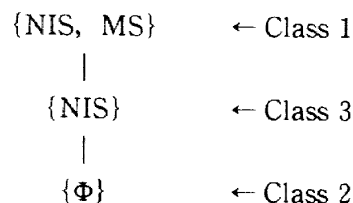
H1: このモデルは、潜在クラスが2つで、クラス1は学習完了(mastery)クラスであり、クラス2は学習非完了クラス(nonmastery)クラスであると仮定する。

H2: このモデルでは、4つのテスト問題が、2つのスキルより成り立っており、1つのスキル(数字の同定)がもう1つのスキル(数字とブロックの対応づけ)の前提(prerequisite)になっていると想定する。このモデルでは、H1での2つのクラスに加えて、クラス3(数字

H1



H2



H3

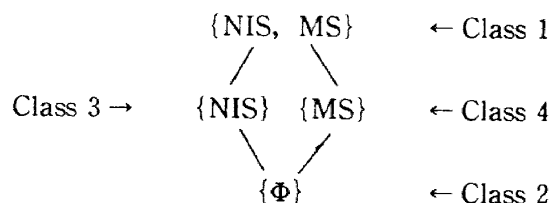


Fig. 1 Bergan (1988) の H1, H2, H3 のハッセ図による表現

MS=Matching Skill

NIS=Numeral Identification Skill

の同定はできるが、対応づけはできない)を想定する。

H3: まず、課題が「数字の同定」、「数字とブロックの対応づけ」という異なる知識領域に含まれる、異なる2値変数によって表現されるものとする。そして4つのクラス

クラス1: 数字の同定可能, 数字とブロックの対応づけ可能

クラス2: 数字の同定不能, 数字とブロックの対応づけ不能

クラス3: 数字の同定可能, 数字とブロックの対応づけ不能

クラス4: 数字の同定不能, 数字とブロックの対応づけ可能を想定する。

STLTM の考え方をを用いると、H1—H3 は、1)潜在クラスをその潜在クラスが持つ能力の集合として表わし、2)集合間の包含関係より成立する半順序構造(これは同時に、学習・発達の順序と解釈できる)をハッセ図を用いて表わすことにより (Fig. 1 参照)、わかりやすく表現することができる。

H1—H3 を仮定すると、あるパターンが生起する確率の評価式と、このような確率の間で満たされなければならない制約条件が明らかになる (例えば TABLE 2)。Bergan はこれらを用いてモデルの対数尤度を計算し

**** このような方法論を論じた研究は数多いが Goodman (1975), Rindskopf (1983), Formann (1985), Langeheine (1988) 等が代表的と言える。

TABLE 2 Bergan によって予想された各クラスの被験者が各問題に正答する確率

問 題	A	B	C	D
Class 1	P5	P6	P7	P8
Class 2	P1	P2	P3	P4
Class 3	P5	P6	P3	P4
Class 4	P1	P2	P7	P8

(P1—P8 は確率の値を示し、同じ記号によって表現されたセルは同じ確率にならなければならない。問題 AB は NIS を、問題 CD は MS を持てば正答できると仮定すると、Class 1 の被験者の P5, P6, P7, P8 は高い数値になり、Class 2 の被験者の P1, P2, P3, P4 は低い数値になることが仮定できる。Class 3 の被験者は問題 AB に対しては Class 1 の被験者と同じ正答確率(すなわち P5, P6)を持つが、CD に対しては Class 2 の被験者と同じ正答確率(すなわち P7, P8)を持つと予想できる。同様に、Class 4 の被験者は問題 CD に対しては Class 1 の被験者と同じ正答確率(すなわち P7, P8)を持つが、AB に対しては Class 2 の被験者と同じ正答確率(すなわち P1, P2)を持つと予想できる。

尤度比検定を行ってモデル間の優劣を検証した。その結果 H3 が最も当てはまりがよいことがみいだされた。

集合論的項目反応理論による分析

STLTM では、Fig. 1 のような被験者の能力の構造があると仮定する。また問題の構造は、問題 A と B は {NIS}, 問題 C と D は {MS} で表現できる。STLTM によれば正答および誤答の確率は被験者を表現する集合と項目を表現する集合との集合関数であるとされるが具体的な関数形には様々な可能性がある(椎名, 1990 参照)。ここでは

$$\begin{aligned} P(\text{正答} | S_i, Q_j) \\ = \exp\{\alpha_j(S_{ij} - q_j) - \beta_j\} \\ P(\text{誤答} | S_i, Q_j) = 1 - P(\text{正答} | S_i, Q_j) \end{aligned} \quad (1)$$

と仮定する。ここで

S_i : 被験者 i を表現する能力集合

Q_j : 問題項目 j を表現する能力集合

$\alpha_j, \beta_j \geq 0$: 項目に与えられるパラメーター

$S_{ij} = \text{Card}(Q_j \cap S_i)$

(Card は集合の濃度、すなわち要素数を算出する関数)

$q_j = \text{Card}(Q_j)$

である。

Bergan は潜在クラス分析を用いているので、これとの対応をとるために、本稿では(1)式を用いた潜在ク

ラス分析流モデルを考える。すなわち $P(S_k)$ を母集団の中で能力集合 S_k を持つ者の事前確率とし、

$$P(S_k, Q_j) = \exp\{\alpha_j(S_{kj} - q_j) - \beta_j\}$$

とすると被験者 i の 1—0 の回答パターンベクトル u_i の生起する確率は

$$P(u_i) = \sum_k [P(S_k) \prod_j P(S_k, Q_j)^{u_{ij}} \{1 - P(S_k, Q_j)\}^{1-u_{ij}}] \quad (2)$$

と書ける。ここで

$$\sum_k P(S_k) = 1$$

である。このモデルを周辺化された (Marginalized) STLTM という意味で MSTLTM と呼ぶことにする****。このモデルによる全データの対数尤度 L は、 $n(u_i)$ を u_i と同じパターンが観察された頻度として

$$L = \sum_i n(u_i) \ln P(u_i) \quad (3)$$

で与えられる(3)の添え字 i は被験者ではなく回答パターンを示すのに注意されたい。以下同様)。最尤推定量を求めるためには $\alpha_j, \beta_j, P(S_k)$ を未知パラメーターとし

$$\alpha_j, \beta_j \geq 0$$

$$1 \geq P(S_k) \geq 0$$

$$\sum_k P(S_k) = 1$$

を制約条件として L を最大化すればよい。この制約条件をみたすために付加変数 $-\infty \leq a_j, b_j, s_k \leq \infty$ を導入し

$$\alpha_j = \exp(a_j)$$

$$\beta_j = \exp(b_j)$$

$$P(S_k) = \frac{\exp(s_k)}{\sum_m \exp(s_m)}$$

と定義する。付録に示した一次微分係数を用いて準ニュートン法と最急降下法を組み合わせ、 L の最大化を行うプログラムを作成した。

Bergan の分析では H3 が最も当てはまりがよかったので、手始めに Bergan の H3 と Fig. 1 の H3 の構造を用いる MSTLTM との比較をすることにしたい。Bergan の H3 では、基本的には、

$$\begin{aligned} P(\text{正答} | \text{Class } k, \text{問題 } j) \\ (k=1, 2, 3, 4; j=A, B, C, D) \end{aligned}$$

$$P(\text{Class } k)$$

$$(k=1, 2, 3, 4)$$

という 2 種類の確率を未知パラメーターとする。クラス数が 4, 問題が 4 問であるので、全部で 20 個のパラ

**** この周辺化はいわゆる Marginal Maximum Likelihood 法 (Bock & Aitkin, 1981) とは異なる。なぜなら、不要のパラメーターが尤度関数から除去されていないからである。ここでの周辺化はデータ生成プロセスの中に周辺化が含まれているという意味であり、潜在クラス分析の基本モデルに組み込まれている仮定を再び述べたにすぎない。

メーターがあるが、TABLE 2 のような制約条件(この制約条件が問題構造に対するモデルを表現している)をおき、また $\Sigma P(\text{Class } k) = 1$ を考慮すると、推定すべきパラメーターは11個となる。また H3 の最大対数尤度は 1382.98 となるので(この値は Bergan (1988, p.241, Tab. 2) の数値から逆算した)、 $AIC = 2787.96$ となった。

MSTLTM の場合は、例えば $\text{Class } 1 = \{\text{NID}\}$ 、問題 $A = \{\text{NID}\}$ の場合を考えると

$$S_1 = \{\text{NID}\}$$

$$Q_A = \{\text{NID}\}$$

$$S_{1A} = \text{Card}(Q_A \cap S_1) = 1$$

$$q_A = \text{Card}(Q_A) = 1$$

となるから

$$P(\text{正答} | S_1, Q_A) = \exp\{-\beta_A\}$$

となり、また $\text{Class } 2 = \{\phi\}$ 、問題 $A = \{\text{NID}\}$ の場合

$$S_2 = \{\phi\}$$

$$Q_A = \{\text{NID}\}$$

$$S_{2A} = \text{Card}(Q_A \cap S_2) = 0$$

$$q_A = \text{Card}(Q_A) = 1$$

となるから

$$P(\text{正答} | S_2, Q_A) = \exp\{-\alpha_A - \beta_A\}$$

となる。同様に、全ての場合を列挙すると、TABLE 3-1 のようになる。TABLE 3-1 と TABLE 2 は本質的に同じであり、TABLE 3-1 は TABLE 2 のパラメー

ターを別の形式で書き表わしたものになっており、パラメーター数も同一である。従って、FIG. 1 の H3 の構造を用いる MSTLTM と Bergan の潜在クラス分析の結果は全く同一の当てはまりを与えると予想できる。実際に計算してみると TABLE 3-2、TABLE 4 のような結果がえられ、両者の実質的な同一性が確かめられた。

制約されたモデル 上の結果では、 $\exp\{-\beta_A\}$ 、 $\exp\{-\beta_B\}$ 、 $\exp\{-\beta_C\}$ 、 $\exp\{-\beta_D\}$ の値が 1 に近くなっている。そこで $\beta_A = \beta_B = \beta_C = \beta_D = 0$ と制約したモデルを考えることができる。これは能力集合と問題を表現する集合が等しい場合、正答確率は 1 となるという仮説を検討することに相当する。計算の結果 $AIC = 2788.16$ となり H3 ($AIC = 2787.96$) とほとんど同等の当てはまりを示した。

以上の適用例の場合 MSTLTM とは、潜在クラス分析で推定される、 $P(\text{正答} | \text{潜在クラス, 問題})$ という確率を、(1)式のような集合関数を用いてリパラメトライズしたものと解釈できる。その際に、FIG. 1 で示したような発達・学習の仮説的構造が上記の確率の中に反映されるのである。従って、一般に MSTLTM が使用可能な場合には、潜在クラス分析が原理的に使用可能である。ただし潜在クラス分析を用いる場合には、TABLE 2 のようなパラメーターの制約条件を定めな

TABLE 3-1 MSTLTM によって予想される各クラスの被験者が各問題に正答する確率

問 題	A={NIS}	B={NIS}	C={MS}	D={MS}
Class				
1{NIS, MS}	$\exp\{-\beta_A\}$	$\exp\{-\beta_B\}$	$\exp\{-\beta_C\}$	$\exp\{-\beta_D\}$
2{ ϕ }	$\exp\{-\alpha_A - \beta_A\}$	$\exp\{-\alpha_B - \beta_B\}$	$\exp\{-\alpha_C - \beta_C\}$	$\exp\{-\alpha_D - \beta_D\}$
3{NIS}	$\exp\{-\beta_A\}$	$\exp\{-\beta_B\}$	$\exp\{-\alpha_C - \beta_C\}$	$\exp\{-\alpha_D - \beta_D\}$
4{MS}	$\exp\{-\alpha_A - \beta_A\}$	$\exp\{-\alpha_B - \beta_B\}$	$\exp\{-\beta_C\}$	$\exp\{-\beta_D\}$

TABLE 3-2 BERGAN (1988) および MSTLTM によって推定されたクラスの事前確率と各クラスの被験者が各問題に正答する確率 (+は Bergan (1988), p.241, Tab. 3 より引用)

	P(Class)	P(A に正答)	P(B に正答)	P(C に正答)	P(D に正答)
Class					
1	.363+	.944+	.999+	.973+	.944+
	.369	.940	1.000	.971	.943
2	.428+	.296+	.166+	.300+	.344+
	.418	.295	.158	.298	.331
3	.139+	.944+	.999+	.300+	.334+
	.142	.940	1.000	.298	.331
4	.070+	.296+	.166+	.973+	.944+
	.070	.295	.158	.971	.943

TABLE 4 様々なモデルのAICの一覧

		最大対数尤度	df.	AIC
H3	Berganの結果	1382.98	11	2787.96
H3		1383.34	11	2788.68
	(制約付き)	1387.08	7	2788.16
S1		1389.62	13	2805.48
	(制約付き)	1395.42	9	2808.84
S2		1384.07	13	2794.13
	(制約付き)	1387.23	9	2792.46
S3		1394.96	12	2813.92
	(制約付き)	1395.53	8	2807.06
S4		1387.92	12	2799.84
	(制約付き)	1393.28	8	2802.56
S1'		1388.94	12	2801.88
	(制約付き)	1395.20	8	2806.40

ければならない。項目や潜在クラス数が増えた場合、これは大変な作業になる。MSTLTMではモデルを作る際にハッセ図を用いることになるが、この図は発達・学習のパスを表わすのに大変便利である。しかしながら、これだけでは、ハッセ図のような図式表現を用いて被験者と項目をわかりやすく表現できる以外には、たいした前進ではないという評価もなりたとう。しかし、MSTLTMでは、潜在クラス分析で扱うことが可能な、1) 正答確率が同じ値になるという制約、2) 正答確率がある定数になるという制約、に加えて、3) 正答確率間の大小関係に関する制約を簡単にモデル化することができるという大きな利点がある。またクラス数や項目数が多い場合にはパラメーター数も節約できることになろう。この利点をはっきり示すことができる例を次節で示したい。

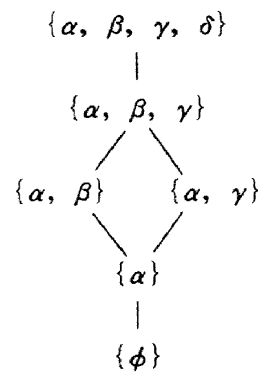
複雑なモデルの検討比較

前節の例は、MSTLTMと潜在クラス分析が結果的に等価となる場合であった。そこでFIG. 1の構造以外にも様々な構造モデルを考えてみることにする(FIG. 2-1)。

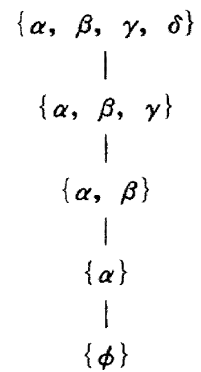
S1では、数字「3」の同定がまずできるようになり、その後「4」の同定または、「3」とブロックの対応づけが可能になると仮定する。その後、「3」と「4」の同定及び「3」とブロックの対応づけができるようになり、最後にすべての能力を獲得するというモデルを想定している。S1'はS1から $\{\alpha, \gamma\}$ を抜いたものであり、線形のパスで発達が表現されることになる。

S2は、H3をもうすこし細かくしたものであり、同定にしろ対応づけにしろまず「3」について可能にな

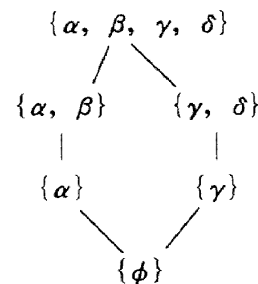
S1



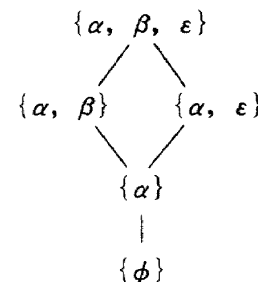
S1'



S2



S3



S4

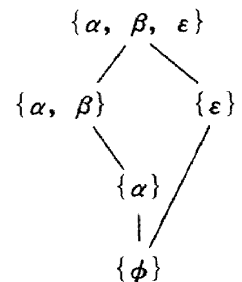


FIG. 2-1 様々な能力集合の構造

α :「3」の同定スキル, β :「4」の同定スキル,
 γ :「3」とブロックの対応づけスキル, δ :「4」
 とブロックの対応づけスキル, ε :対応づけスキル
 ($\gamma\delta$ を区別しない)

り、次に「4」について可能になるという順序性の仮説を盛り込んでいる。以上述べたS1, S1', S2ではFIG. 2-2のテスト構造を想定する。この構造は4つのテスト項目が相互に独立であると主張している。

S1, S1', S2の場合

	α	β	γ	δ
A	1	0	0	0
B	0	1	0	0
C	0	0	1	0
D	0	0	0	1

FIG. 2-2 テストの構造

S3, S4の場合

	α	β	ε
A	1	0	0
B	0	1	0
C	1	0	1
D	0	1	1

FIG. 2-3 テストの構造

S3はS1に制約をくわえたものである。すなわちS1では「3」とブロックの対応づけと「4」とブロックの対応づけは別のスキルであると想定しているが、この2つのスキルを同一のスキル、 ε 、とみなすのがS3である。またS3はH3を細かくしたものと解釈できる。すなわちH3におけるNISを α 、 β に分解しさらに α と β の間に順序性を仮定している。S4はS3とほぼ同じ構造であるが、 ε の獲得を $\alpha\beta$ の獲得と異なるパスで表現している点が違っている。

なお、S3、S4ではFIG. 2-3のテスト構造を想定する。この構造においては、 ε が α または β を前提する能力となっている。

制約されたモデル 以上のモデルに対しても前節のように、 $\beta_A = \beta_B = \beta_C = \beta_D = 0$ なる制約を与えた場合を考慮することにする。

推定の結果

パラメーター推定の結果導出されたAIC(坂元, 石黒, 北川, 1983)の値をTABLE 4に示す。最もよい当てはまりを示したのはH3であるが、S1及び制約付きのS1もほぼ同様の当てはまりを示した。この結果からBerganのモデルは本論で考察した諸モデルの中で最小AICの意味で最良と結論できよう*****。しかしながら、これらのモデルは考えるモデルのうちのほんの一部分にすぎず、さらによりモデルの可能性は常に存在することを指摘しなければなるまい。

結 語

以上よりSTLTMがいかに使うかの1つの例を示すことができたと思う。その使用上の要点をまとめると、1)ハッセ図を用いて、被験者と項目のさまざまな構造を仮定し、2)データとの当てはまりを比較するということである。ただし、1)における仮説構造の設定がなされるためには、かなり明確な仮定を持つ必要がある。これは、本手法の適用領域を制限するものである。しかしながら、1)においては、複数の発達・学習のパスを比較的容易に表現することができ、また、2)においては統計モデルの比較という形で様々な仮説を調べることが可能である。このことにより、他の手法では扱い得ないか、または扱うのが難しい問題を比較的簡単に解けるようになり、様々な分野で有効性を発揮しうのではないかと考えている。

***** 坂元, 石黒, 北川 (1983, p.63)によれば、AICの差が1以下程度であれば有意な差は無いとみなしてよい。

引用文献

- Bergan, J.R. 1988 Latent variable techniques for measuring development. In R. Langeheine & J. Rost (Eds.) *Latent trait and latent class models*. Plenum.
- Bergan, J.R., & Stone, C.A. 1985 Latent class models for knowledge domains. *Psychological Bulletin*, 98, 166—184.
- Bergan, J.R., Stone, C.A., & Feld, J.K. 1984 Rule replacement in the development of basic number skills. *Journal of Educational Psychology*, 76, 289—299.
- Bock, R.D., & Aitkin, M. 1981 Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, 46, 443—459.
- Formann, A.K. 1985 Constrained latent class models: Theory and applications. *British Journal of Mathematical and Statistical Psychology*, 38, 87—111.
- Goodman, L.A. 1975 A new model for scaling response patterns: An application of the quasi-independence concept. *Journal of the American Statistical Association*, 70, 755—768.
- Langeheine, R. 1988 New developments in latent class theory. In R. Langeheine & J. Rost (Eds.) *Latent trait and latent class models*. Plenum.
- Rindskopf, D. 1983 A general framework for using latent class analysis to test hierarchical and nonhierarchical learning models. *Psychometrika*, 48, 85—97.
- 坂元慶行・石黒真木夫・北川源四郎 1983 情報量統計学 共立出版
- 椎名乾平 1990 集合論的潜在特性モデル ——理論的考察—— 教育心理学研究, 38, 1—8.

(1989年5月20日受稿)

付 録

最大化に必要な勾配は以下の通り。

$$B = P(S_k, Q_\ell)^{u_{k\ell}} \{1 - P(S_k, Q_\ell)\}^{1-u_{k\ell}}$$

として

$$\frac{\partial L}{\partial a_\ell} = \sum_i n(u_i) \frac{\partial \ln P(u_i)}{\partial P(u_i)} \frac{\partial P(u_i)}{\partial B} \frac{\partial B}{\partial a_\ell} \frac{\partial a_\ell}{\partial a_\ell}$$

$$\frac{\partial L}{\partial b_\ell} = \sum_i n(u_i) \frac{\partial \ln P(u_i)}{\partial P(u_i)} \frac{\partial P(u_i)}{\partial B} \frac{\partial B}{\partial \beta_\ell} \frac{\partial \beta_\ell}{\partial b_\ell}$$

$$\frac{\partial L}{\partial s_\ell} = \sum_i n(u_i) \frac{\partial \ln P(u_i)}{\partial P(u_i)} \frac{\partial P(u_i)}{\partial P(S_\ell)} \frac{\partial P(S_\ell)}{\partial s_\ell}$$

ここで

$$\frac{\partial \ln P(u_i)}{\partial P(u_i)} = 1/P(u_i)$$

$$\frac{\partial P(u_i)}{\partial B} = \sum_k P(S_k) \prod_{j \neq \ell} P(S_k, Q_j)^{u_{ij}} \{1 - P(S_k, Q_j)\}^{1-u_{ij}}$$

$$\frac{\partial B}{\partial \alpha_\ell} = \begin{cases} \exp\{\alpha_\ell(S_{kj} - q_j) - \beta_\ell\} (S_{kj} - q_j), & u_{ij}=1 \\ -\exp\{\alpha_\ell(S_{kj} - q_j) - \beta_\ell\} (S_{kj} - q_j), & u_{ij}=0 \end{cases}$$

$$\frac{\partial \alpha_\ell}{\partial a_\ell} = \exp(a_\ell)$$

$$\frac{\partial B}{\partial \beta_\ell} = \begin{cases} -\exp\{\alpha_\ell(S_{kj} - q_j) - \beta_\ell\}, & u_{ij}=1 \\ \exp\{\alpha_\ell(S_{kj} - q_j) - \beta_\ell\}, & u_{ij}=0 \end{cases}$$

$$\frac{\partial \beta_\ell}{\partial b_\ell} = \exp(b_\ell)$$

$$\frac{\partial P(u_i)}{\partial P(S_\ell)} = \prod_j P(S_\ell, Q_j)^{u_{ij}} \{1 - P(S_\ell, Q_j)\}^{1-u_{ij}}$$

$$\frac{\partial P(S_\ell)}{\partial s_\ell} = \frac{\exp(s_\ell)}{\sum_m \exp(s_m)} - \left\{ \frac{\exp(s_\ell)}{\sum_m \exp(s_m)} \right\}^2$$