統計数理　第37巻　第2号（1989）

「Batchelder と Romney の正答のないテスト理論」の拡張とアンケート調査法への応用

統計数理研究所　吉　野　諒　三

（1989年11月　受付）

1. 序

社会科学における客観的データ収集の方法として，しばしばアンケート調査の手法が用いられている（Romney et al.（1986）のレビュー及び文献表を参照）．これは，特に，社会学者や人類学者がある国や社会を調査する際の基礎的手続きとして用いることが多い．そのような際には，アンケートの回答者たちは各々の国や社会に共通の文化，常識を共有していると想定される．研究者たちは，通常，調査すべき課題についてはたとえ曖昧ではあっても何等かの知識を持っていることが多い．したがって，アンケート調査項目の選択は，アンケート調査の当初においては必ずしも大きな問題ではないかもしれない(但し，アンケートを再三繰り返したり，複数の国々や文化を比較する場合等には調査項目の選択，作成自体が大きな問題となり得る)．しかしながら，アンケート調査の本来の意味を考えると，アンケートの各項目の回答の中でどれが「各文化の中での正答」，或いは「各文化を代表する回答」であるかを研究者が先験的に知っていると仮定することは，勿論できない．さらに，回答者の各々の能力には当然差違があり，これが回答に何等かの形で影響を与えることは想像に難くない．したがって，研究者の課題はそのようなデータを統合，分析して如何に合理的な結論を導くかということになる．

社会調査法としてのアンケート調査の課題は，基本的には上記の通りである．しかし，さらに，社会調査が社会を調べることを目的とすると同時に，それ故に，逆にその実施が社会と社会の変化から影響を受けるものであるということは，今日のように急速に変わりつつある世界にあっては看過できない．具体的な問題点としては，先進諸国のように流動性が激しく，またプライバシーの尊重が声高々に叫ばれる社会において，従来の全抽出，標本抽出法によるアンケート調査の回答回収率が急速に低くなりつつあり，得られたデータの偏差や信頼性の点に関して疑義を生じていることが挙げられる（林 他（1989）参照）．また，それほど大規模ではない研究機関では，人件費の高騰化傾向を考えると，フィールド・ワークを遂行する有能な調査者を確保することが困難の度合を増してきている事実も挙げられる．

社会調査法を統計数理の立場から研究する人々の直面している課題の一つは，こういった状況の中で如何に妥当性と信頼性の高いデータを得るか，或いは合理的な統計数理的分析をするかということである．本論文では，物理的に，或いは経済的に少数の回答者からのデータしか収集できない場合でも，より信頼性の高い分析結果を得る手法を開発しようとする一つの試みとして，Batchelder and Romney（1988）によるテスト理論を応用，拡張したモデルを構成する．さらに，そのモデルの一つの応用例として，内閣総理大臣官房広報室の過去14年間にわたる継続調査によって得られた日本人の国民生活に関する世論調査のデータを分析してみる．最後に，簡単ではあるが，そのモデルの修正，拡張されるべき点について議論し，このモデルの

近い将来の発展の方向を考える．なお，本論文では読者が社会調査法（例：西平（1979））とテスト理論（例：Lord and Novick（1968））の初歩的知識は持っているものと仮定する．

2. 「Batchelder と Romney のテスト理論」の要約

2.1 「Batchelder と Romney のテスト理論」の背景

理論の説明をする前に，Batchelder と Romney の研究の生まれた背景を簡単に説明しておくことが，後で彼らの研究の説明を理解するのに役に立つであろう．

Batchelder は数理心理学を代表する研究者の一人であり，1960年代の学習心理学の構築に貢献した一人であり，その他，社会科学一般にわたって，心理学的な概念を明確に定式化し，数理統計的に展開した多くの数理的モデルを構成してきたことで広く知られている．30歳代の後半から40歳代の始めにかけて *Journal of Mathematical Psychology* の編集長を務め，現在でもまだ40歳代の後半であり，幾つかの著名な数理科学系の雑誌の編集審査員となっている．Romney は多次元尺度法を始めとして，統計的手法を利用して社会科学一般，特に人類学的な研究データの解析をしてきたことで有名である．彼の場合は，新しい数理統計モデルの開発というよりは，既に知られている統計法の巧妙な応用研究が得意なようである．現在では，二人ともに University of California, Irvine（UCI）の教授であり，数理社会科学を代表する研究者である．

本論文で扱う研究課題に関していえば，彼らのこの共同研究は1986年の UCI で開催された Information Pooling and Group Decision Making をテーマとする学術会議（Grofman and Owen（1986）を参照）と前後して始まった．この会議では，各種の社会的グループ内で何等かの意志決定（例えば，政策の選択など）が必要な際に，そのグループの各構成員のそれぞれの判断を如何に統合するのが合理的であるのか等を研究することが主要なテーマであった．Batchelder と Romney は，この問題を数理心理学のパラダイムのなかで捉え，彼らに固有のアプローチを示した．即ち，(1) まず問題の本質を捉えた上で，取り扱っている問題のモデルのイメージを明確化する．(2) そして，可能な限り単純化し，しかし，なおかつ本質を表したモデルを構築する（これは，1960年以前の数理心理学的モデルが複雑な心理学的現象を解明するために，多数のパラメーターや仮定を導入した複雑なモデルを構成したが，結局，そのようなモデルでは現実の現象を解明したり，予測するのに役に立たなかったという歴史的反省に基づいたパラダイムである．この単純化を如何に巧妙に行なうかが，その研究の価値の評価基準の一つとなる）．その際，モデルの本質的な点を，例えば，公理系として明記し，その当否の議論の焦点が絞れるようにする（数理心理学では，しばしば，「公理（axiom）」という言葉を使うことがあるが，これは，モデルや理論の出発点の明確化という意味で用いられている．決して，19世紀までの数学のように，議論の余地のない自明の理という意味で用いられているわけではないので，注意していただきたい）．(3) 次に，モデルに含まれるパラメーターの推定のための統計的技法を確立する．これには，既知の技法が利用できる場合もあれば，新しい技法の開発が必要なこともあろう．モデルや理論を表面上はできる限り簡明にするために，幾つかの仮定を設け，扱える状況を制限することは多い．こういった手続きの中で，良いモデル構成とは，モデルの仮定が一つ一つ外され，より一般的な条件の場合を扱う際にも，それに対応するモデルやパラメーターの推定技法が，既に構成されているものから，比較的，直接的な拡張によって得られるようなモデルを構成することである．つまり，数学的な理論上の困難ではなく，単なる計算的な煩雑さを避けるための仮定を導入するのには，あまり躊躇する必要はない．その仮定だけを外し，より一般の場合のためのモデルを作るには，論理的議論ではなく，計算の労苦

だけが必要なのである．

Batchelder と Romney は，多肢選択テストをこの問題のモデルのイメージとして捉えた．多肢選択テストは日本でも，大学共通一次試験，或いは初等教育でもしばしば用いられ，一般の人にもよく知られている．通常は，勿論，試験を出す側は正答を知っているわけである．しかし，ここでモデルの状況設定のために，仮に正答が書かれた書類（answer key）を紛失してしまい，しかも正答を復元する手がかりは，受験者の回答データ以外は全くないと想定しよう．したがって，受験者の回答データから正答を推定し，かつ各受験者の得点を採点しなければならない．すぐに思いつく単純な解決法の一つは，各問題について最多数の受験者が選択した選択肢を正答と推定することであろう．これは，「多数決の原理」に従うルールである．しかしながら，多数者が必ずしも正しい判断をするとは限らない．烏合の衆よりも，少数の高い能力の人々の判断の方が信頼性が高いことも多いであろう．また，全ての問題が，ある程度難しいものであれば，能力の低い受験者の回答パターンはばらつきを示し，一方能力の高い受験者たちの回答パターンは一致性が高いであろう．このような回答者の能力の個人差を考慮して，全ての受験者の回答パターンから正答を合理的に推定するには，如何にすべきであろうか．

この状況設定は，Information Pooling and Group Decision Making の問題の重要な側面を捉えていることが理解できるであろう．即ち，あるグループの構成員の各々の判断データから，如何に合理的にグループとしての最終判断を導くかという問題である．アメリカ合衆国でも，大学生，もしくは大学院生の選抜試験の一部の資料として SAT や GRE と呼ばれる多肢選択テストが広く用いられているので，上述の状況設定はイメージしやすい．逆にいうと，Batchelder と Romney は教育的テスト理論そのものに特別の興味があったわけではなく，飽くまでも，彼らの理論を多くの人々が理解し，巧妙な応用を考案するための方便として，上述のようなテストの状況をイメージしたのである．

彼らの理論の展開を，一言で述べると，ある受験者がある問題の正答を知っている確率をその受験者の能力と定義して，D_i というパラメーターで表し，それによって受験者のペア毎に回答パターンの一致率の期待値の計算式を導き，その期待値に対応する実際のデータを挿入することによって，D_i を推定するのである．さらに，それらの推定値と全受験者の回答パターンから最尤法によって正答を推定するのである．具体的な数学の展開は，次節の通りである．

2.2　GCM モデルと HTM モデルの概説

Batchelder and Romney (1988) のテスト理論の詳細は彼らの原著に任せるが，彼らの GCM モデル（the general Condorcet model）と，その特殊な場合と考えられる HTM モデル（the high threshold model）を簡潔に説明しよう．この 2.2 節の内容は数学的には 3.2 節の内容に含まれるのだが，Batchelder and Romney (1988) の研究の歴史的背景を尊重することと，第 3 章での展開の説明上，彼らの用いている記号の表記に統一する便宜上，Batchelder and Romney の研究に沿って説明することから始める．以下では，便宜上「正答」等のテスト理論の用語を使うが，読者は，各具体的な応用上の状況に応じて，適切な解釈をすべきである．テスト理論の用語の本来の意味に固執して，モデルの適用価値を見誤らないように注意していただきたい．

まず質問項目が “yes-no (「はい」または，「いいえ」)” での回答の形になっていることを仮定する（この仮定は，後に，緩い制限に取り替えられる）．通常のテスト理論では，テストの遂行者（教師，または研究者）は正答を知っていることが前提とされる（Lord and Novick (1968) を参照）．しかし，GCM モデルでは，これを前提とはせず，回答者の反応データのみから「正答」と「回答者の能力」が推定される．但し，この「正答」と「回答者の能力」の意味は各場合に応じて適宜，解釈されるべきであろう．例えば，学力テストでは，テストする側に「予め

知られている正答」と，モデルによって「推定された正答」とを比較して，もし食い違いがあれば，その問題と回答を再度吟味する必要があるだろう．実際，受験者が問題文を，テストする側の意図とは異なった風に解釈するのはよくあることである．また，もしこのモデルを人類学者がある未知の文化のアンケート調査として用いれば，「推定された正答」がその文化を表す回答として解釈できよう．この場合，少数の優れた回答者の回答の方が，多数の能力の低い回答者の回答よりも信頼性が高いことがあるのに注意しよう．

N 人の回答者が，M 個の質問項目に回答するものとせよ．まず，三種類の統計量が以下のように定義される．

1. 回答データ（response profile data）行列 X は $N\times M$ 行列で，各要素 X_{ik} は，回答者 i が質問項目 k に「はい」と回答した時は 1，「いいえ」と回答した時は 0 である．
2. 正答（answer key）行列 Z は $1\times M$ 行列で，各要素 Z_k は，質問項目 k の正答が「はい」の場合は 1，「いいえ」の場合は 0 である．
3. 回答者の正答・不正答データ（performance profile data）行列 Y は $M\times N$ 行列で，各要素 Y_{ik} は，回答者 i の質問項目 k の回答が正しい場合は 1，誤りの場合は 0 である．

これらの三種類の行列要素は独立ではなく，それらの間に簡単な代数関係が成立しているのを見るのは容易であろう．

Batchelder and Romney（1988）による GCM モデルでは，各質問項目の困難度が等しいと仮定する（各質問の困難度が一様ではない場合のモデルについては，1988 年の彼らの論文の後半で議論されている）．信号検出理論（Green and Swet（1966））からのアナロジーで，GCM モデルは検出率（hit rate）P_{1i} と誤検出率（false alarm rate）P_{0i} によって，各回答者 i を質問項目 k に対して次のように特性付ける．

$$P_{1i}=\Pr(X_{ik}=1:Z_k=1)$$
$$P_{0i}=\Pr(X_{ik}=1:Z_k=0)$$

このモデルの形式的定義は，次の定義 2.1 で与えられる．

定義 2.1.　GCM モデルは次の三つの公理によって定義される．

公理 1.　共通の正答（common truth）の存在．各質問項目に対して，全ての回答者に共通の一つの正答がある．

公理 2.　質問項目間の独立性（local independence）．質問項目と回答者の回答の間に次の関係が成立する．

$$\begin{aligned}\Pr[(X_{ik})_{N\times M}=(x_{ik}):(Z_k)_{1\times M}=(z_k)]\\ =\prod_i\prod_k \Pr(X_{ik}=x_{ik}:Z_k=z_k)\end{aligned}$$

ここで (z_k) は，正答行列を表すとする（但し，正答行列は先験的には未知である）．

公理 3.　質問項目の難易度の一様性（homogeneity of items）．各回答者は各質問項目に対して次のような一定の検出率 P_{1i} と誤検出率 P_{0i} を持つ．

$$\Pr(X_{ik}=1:Z_k=z_k)=\begin{cases}P_{1i}, & z_k=1\text{ の場合}\\ P_{0i}, & z_k=0\text{ の場合}\end{cases}$$

HTM モデル（Blackwell (1963) 参照）は，GCM モデルが次の関係を満たす特殊な場合として定義される．

$$P_{1i}=D_i+(1-D_i)g_i$$
$$P_{0i}=(1-D_i)g_i$$

ここで D_i は回答者 i の能力(competence), g_i は一種のバイアスと意味付けられ, 共に[0, 1]の閉区間に属すると仮定される．この D_i は，回答者 i が「正答」を知っている確率と解釈できる．また，この g_i は，$Z_k=1$ で，回答者 i が正答を知らない場合に，選択肢「はい」を選ぶ $(X_k=1)$ バイアスと解釈できる．この比較的単純なモデルの設定に, 1960 年代の心理学における学習・記憶理論とモデル構成の影響がうかがえよう．

GCM モデル，したがってその特殊な場合である HTM モデルでは勿論，正答行列 Z は未知であり，これを各回答者の能力 D_i と共に，回答データ行列 X に基づいて推定することが課題なのである．これを，GCM（HTM）モデルにおけるパラメーター推定問題と呼ぼう．

2.3　各回答者の能力 D_i と正答行列 Z の推定

Batchelder and Romney(1988)は種々の条件を仮定し，それぞれの条件のもとで回答データ行列 X から，各回答者の能力 D_i と正答行列 Z とを推定する理論を展開している．ここでは，その中で本論文に直接関連のある HTM モデルで偏ったバイアスなしの（つまり，各選択肢に均等なバイアスがかかっていて $g_i=1/2$ を仮定できる）場合に焦点を当て簡潔に説明する．

検出率 $P_1=(P_{11},\cdots,P_{1N})$，誤検出率 $P_0=(P_{01},\cdots,P_{0N})$，及び正答行列 $Z=(Z_1,\cdots,Z_M)$ を GCM モデルのパラメーターと考えた場合，公理 2 と 3 より尤度関数 L は次式で与えられる．

$$\begin{aligned}&L(X;\ P_1,P_0,Z)\\&=\prod_i\prod_k[P_{1i}(1-P_{0i})/\{P_{0i}(1-P_{1i})\}]^{X_{ik}Z_k}\\&\quad\cdot[(1-P_{1i})/(1-P_{0i})]^{Z_k}[P_{0i}/(1-P_{0i})]^{X_{ik}}(1-P_{0i})\end{aligned}$$

GCM モデルは，モデルの中での回答者と質問項目の役割を交換すると，2 クラスの潜在構造モデル（Lazarsfeld and Henry (1968) 参照）と論理的に同型になることが知られている（Batchelder and Romney (1988)，p. 74)．しかしながら，だからといって潜在構造モデルが未知パラメーター推定のために直接利用できるわけではない．その理由は，社会科学において潜在構造モデルを利用するのは，ごく少数（高々，6 項目程度）の質問項目に対する多数（しばしば数千）の回答者からのデータを取り扱う場合が多く，特別の推定理論を発展する必要がなく，本論文で問題としているように，少数の回答者に対して中程度(40 から 100 程度)の数の質問項目がある場合の推定問題は，これまでほとんど取り扱われていないからである．

Batchelder and Romney (1988) による新しいパラメーター推定法の概要は次の通りである．但し, 以下の説明と展開には, 筆者の独自の方法が一部混じっているので注意していただきたい．

まず M_{ij} を回答者 i と j の回答の一致率と定義する．即ち，

$$M_{ij}=\left\{\sum_{k=1}^{M}X_{ik}X_{jk}+\sum_{k=1}^{M}(1-X_{ik})(1-X_{jk})\right\}\Big/M$$

ここで M は，質問項目数と定義したことを思い出そう．さらに，

$$M_{ijk}=\begin{cases}1, & X_{ik}=X_{jk}\text{ の場合}\\ 0, & X_{ik}\neq X_{jk}\text{ の場合}\end{cases}$$

と定義する．この時，期待値 $\mathrm{E}(M_{ij})$ は次式のように表される．

$$\mathrm{E}(M_{ij})=(1/M)\sum_{k=1}^{M}\Pr(M_{ijk}=1)$$

さらに，HTM モデルの定義と，これが GCM モデルの特殊な場合であることを考えて，GCM モデルの公理1と2により，次式が導かれる．

$$\begin{aligned}\Pr(M_{ijk}=1)=&D_iD_j+D_i(1-D_j)[\pi g_j+(1-\pi)(1-g_j)]\\&+(1-D_i)D_j[\pi g_i+(1-\pi)(1-g_i)]\\&+(1-D_i)(1-D_j)[g_ig_j+(1-g_i)(1-g_j)]\end{aligned}$$

となる．但し，ここで

$$\pi=\Pr(Z_k=1)$$

とする．ここでさらに，$g_i=g_j=1/2$ を仮定すると，π の値とは無関係に，

$$\Pr(M_{ijk}=1)=(D_iD_j+1)/2$$

となる．したがって

$$\mathrm{E}(M_{ij})=(D_iD_j+1)/2$$

となる．

これは便宜上，

$$M_{ij}^{*}=2M_{ij}-1$$

と定義すると

$$\mathrm{E}(M_{ij}^{*})=D_iD_j$$

となる．これによって，回答者 i, j, k に対して

$$\mathrm{E}(M_{ij}^{*})\mathrm{E}(M_{jk}^{*})/\mathrm{E}(M_{ik}^{*})=D_iD_jD_jD_k/D_iD_k=D_j^2 \tag{2.1}$$

となる．$M_{ij}^{*}, M_{jk}^{*}, M_{ik}^{*}$ は M_{ij}, M_{jk}, M_{ik} のみによって定義され，これらは反応データ行列 X からただちに求められる．したがって，上式の左辺の各期待値を対応する観測値 $M_{ij}^{*}, M_{jk}^{*}, M_{ik}^{*}$ で置き換えることによって D_j の推定値を求めることができる．各回答者 j について，上式の中で異なる i と k の可能な対が $(N-1)(N-2)/2$ 個あり，したがって D_j の推定値が $(N-1)\cdot(N-2)/2$ 個求められるが，例えば，これらの幾何平均を最終的に D_j の推定値 $\hat{D}_j$ と定義できるであろう．この点に関していえば，計算の複雑度(computational complexity)は N^2 のオーダーで増加するが，これは，いわゆる polynomial algorithm で計算でき，即ち，実践的にも問題はない（Batchelder and Romney(1988)は，この推定に因子分析法の最小余剰法（the minimum residual method）を応用している．これは，SPSS の PA2 オプションが利用できるはずである）．結論として，回答者の総数が幾つであっても，マッチング・データ（M_{ij}）があれば式(2.1)によってパラメーター D_j は，常に求められるのである．

以上のように D_j ($j=1, 2, \cdots, N$) が求められると，各質問項目の正答 Z_k は Bayse の定理により

$$\Pr(Z=z : X)=\Pr(X : Z=z)\Pr(Z=z)/\Pr(X)>0.5$$

を計算することによって次のように推定される．即ち，

$$\sum_{i=1}^{N}(2X_{ik}-1)\ln[(1+D_i)/(1-D_i)]>0$$

であり，かつその時に限り $z_k=1$ とする（Batchelder and Romney (1988), p. 79 を参照).

この推定において注意すべき点は，これが多数決の原理による推定（半数以上の回答者の一致した回答を「正答」とする推定）とは異なるということである．大多数の回答者からのデータが用いられる場合でも上述の方法が適用可能であるが（N と M の大小にかかわらず，X からマッチング・データ M_{ij} が得られている限り，この方法は適用可能である)，ほとんどの場合，この推定法は多数決の原理によるものと同じ結論を導く．しかし，少数の，しかも能力にかなり個人差がある回答者たちからのデータを扱う場合は，しばしば，ごく少数の高能力者たちの方が，多数の低能力者たちよりも「正答」の決定への寄与率が高いということが起こり得るのである．Batchelder and Romney ((1988), p. 80) はモンテ・カルロ法によるシミュレイションで，回答者数が 10 で問題数が 50 の場合を検討している．回答者の能力が高い場合ほど，推定される能力の値の標準偏差が少ない傾向が見られる．また，Maher (1987) は，同様のシミュレイションをより少ない回答者数で行ない，回答者の能力にばらつきがある時は，常に多数決の原理のルールよりも，この方法の方が正答の推定において優れていることを観察した．

3. HTM モデルの拡張

3.1 全ての質問項目が S 個の選択肢を持つ場合

GCM モデルは，各質問項目の選択肢が「はい」または「いいえ」の形ではなく，全ての項目の選択肢が S 個ずつある多肢選択の場合にも拡張されている（Romney et al. (1986) を参照．選択肢が，「はい」または，「いいえ」の場合は，$S=2$ の場合と考えられる)．その場合，HTM モデルは，偏ったバイアスなし(つまり，回答が未知の時に各選択肢を選ぶ確率が均等である）と仮定して，

$$\Pr(X_{ik}=s)=\begin{cases}\{D_i(S-1)+1\}/S, & Z_k=s \text{ の場合}\\ (1-D_i)/S, & Z_k\neq s \text{ の場合}\end{cases}$$

と表される．この時，

$$M^*_{ij}=(SM_{ij}-1)/(S-1)$$

と定義すると，単純ではあるが長い計算を経て

$$\mathrm{E}(M^*_{ij})=D_iD_j$$

が導かれ，これによって 2.3 節の式 (2.1) を導出したのと同様の方法によって各回答者の能力の推定値 $\hat{D}_i$ ($i=1, 2, \cdots, N$) を求めることができる．

Maher (1987) は，この方法を現実の多肢選択型の学力テストに適用した．この場合は，勿

論，テストをする側では正答は既知であったのだが，正答を未知と仮定して上記の推定法を用いて，少数の回答者のデータからでも正答を十分に推定できることを確認した．この研究の意味するところは，実際の学力テストのみならず，人類学や社会学の調査研究等，正答が未知の状況においても，このモデルの効用が期待できるということであろう．

3.2　各質問項目の選択肢が必ずしも等しくない場合

前節では全ての質問項目が同じ数の選択肢を持つ場合を扱った．しかしながら，通常のアンケート調査では,選択肢の数が質問項目全てを通じて同じであるとは限らない．この節では,各質問項目の選択肢数が必ずしも等しくない場合も扱えるようにするために，HTM モデルとそのパラメーター推定問題の解法を,筆者の手によって拡張する．このモデルは,第2章と3.1節のモデルを特殊な場合として含むものである．単純ではあるが詳細な計算を示すので，前記のモデルとパラメーター推定問題の解の導出を確認できるであろう．

各質問項目 k が S_k 個の選択肢を持つとせよ．この場合，HTM モデルは「偏ったバイアスなし」という仮定のもとで，次のように表される．

$$\Pr(X_{ik}=s)=\begin{cases}\{D_i(S_k-1)+1\}/S_k, & Z_k=s\ \text{の場合}\\ (1-D_i)/S_k, & Z_k\neq s\ \text{の場合}\end{cases}\tag{3.1}$$

期待値 $\mathrm{E}(M_{ij})$ は以下のように求められる．

$$\mathrm{E}(M_{ij})=(1/M)\sum_{k=1}^{M}\Pr(M_{ijk}=1)\tag{3.2}$$

一方，

$$D_i^*=\{D_i(S_k-1)+1\}/S_k \qquad (i=1,2,\cdots,N)$$

と定義すると，(2.1) より

$$\begin{aligned}\Pr(M_{ijk}=1)&=\sum_{s=1}^{S_k}\Pr(M_{ijk}=1:Z_k=s)\\&=D_i^*D_j^*+\sum_{k=1}^{M}(1-D_i^*)(1-D_j^*)/(S_k-1)^2\\&=\{S_kD_i^*D_j^*-(D_i^*+D_j^*)+1\}/(S_k-1)\end{aligned}$$

ここで

$$D_i^*D_j^*=(1/S_k)^2\{(D_i+D_j)(S_k-1)+1+D_iD_j(S_k-1)^2\}$$

$$D_i^*+D_j^*=\{(D_i+D_j)(S_k-1)+2\}/S_k$$

を用いると

$$\begin{aligned}\Pr(M_{ijk}=1)&=[D_iD_j(S_k-1)^2+S_k-1]/[(S_k-1)S_k]\\&=[D_iD_j(S_k-1)+1]/S_k\end{aligned}\tag{3.3}$$

これを，(3.2) に代入すると，

$$\begin{aligned}\mathrm{E}(M_{ij})&=(1/M)\sum_{k=1}^{M}\{D_iD_j(S_k-1)+1\}/S_k\\&=AD_iD_j+B\end{aligned}$$

ここで,

$$A=(1/M)\sum_{k=1}^{M}(S_k-1)/S_k,$$
$$B=(1/M)\sum_{k=1}^{M}1/S_k$$

である.

さらに便宜上,

(3.4)
$$M^*_{ij}=(M_{ij}-B)/A$$

と定義すると

(3.5)
$$\mathrm{E}(M^*_{ij})=D_iD_j$$

となり D_i ($i=1, 2, \cdots, N$) の推定値は, 2.3 節の式 (2.1) で記した方法で求めることができる. 即ち

(3.6)
$$\mathrm{E}(M^*_{ik})\mathrm{E}(M^*_{ij})/\mathrm{E}(M^*_{jk})=D_i^2$$

において, 左辺の各期待値を対応する観測値で置換し, その平方根を計算し D_i の推定値を求める. 上式で, 各 i に対して i と異なり, かつ互いに異なる j と k の対が $(N-1)(N-2)/2$ 個存在し, それぞれに応じて D_i の推定値が求められるので, それらの幾何平均をもって, 最終的に D_i の推定値 $\hat{D}_i$ とする.

一方, 各項目の正答 Z_k は,

$$\Pr(Z=s:(X_{ik})_{i=1,2,\cdots,N})>0.5$$

を計算することによって推定できる. Bayse の定理により, この式は

(3.7)
$$\Pr((X_{ik}):Z_k=s)\Pr(Z_k=s)\Big/\sum_{s=1}^{S_k}\Pr((X_{ik}):Z_k=s)\Pr(Z_k=s)>0.5$$

となる. 各選択肢が正答となる先験的確率が

$$\Pr(Z_k=s)=1/S_k$$

であることに注意すると, 式 (3.7) の計算には, 条件付き確率

$$\Pr((X_{ik})_{i=1,2,\cdots,N}:Z_k=s)$$

を求めればよいことが分かる. HTM モデルの定義より

$$\Pr(X_{ik}=s:Z_k=s)=D_i+(1-D_i)/S_k$$

かつ

$$\Pr(X_{ik}\neq s:Z_k=s)=(1-D_i)(S_k-1)/S_k$$

である. 一方, 公理 2 によって

$$\Pr((X_{ik})_{i=1,2,\cdots,N}:Z_k=s)=\prod_{i=1}^{N}\Pr(X_{ik}=s:Z_k=s)$$

である．便宜上，次の定義を導入する．

$$X_{iks}=\begin{cases}1, & X_{ik}=s \text{ の場合}\\ 0, & X_{ik}\neq s \text{ の場合}\end{cases}$$

これらによって，

$$(3.8)\qquad \Pr\left((X_{ik})_{i=1,2,\cdots,N} : Z_k=s\right)=\prod_{i=1}^{N}[\hat{D}_i+(1-\hat{D}_i)/S_k]^{X_{iks}}[(1-\hat{D}_i)(S_k-1)/S_k]^{1-X_{iks}}$$

が導かれる．式(3.8)を用いれば，式(3.7)が計算可能となり，したがって正答 Z_k が推定できる．但し，もし項目 k において $s=1,2,\cdots,S_k$ のいずれの選択肢も，式(3.7)を満足しない場合は，選択肢の中で，式(3.7)の左辺を最大にするものを選ぶか，或いは，推定するための情報が不十分と判断することが考えられる（本論文の第4章では，実際のデータ分析の際，後者の判断を選択している）．

本論文では，この節で展開されたモデルを便宜上，一般化されたHTMモデルと呼ぼう．即ち，一般化されたHTMモデルとは，偏ったバイアスなし（つまり，各選択肢に均等なバイアスがかかっている）を仮定できるという条件のもとで，質問項目の選択肢数が各項目によって異なる場合にも適用できるHTMモデルである．

3.3　パラメーター推定のためのコンピューター・プログラムの説明

さて，3.2節のアイデアに基づいて，実際のデータからパラメーターを推定するためのコンピューター・プログラムを作製したが，これを簡単に説明しておく．

パラメーター推定の方法は，簡単に述べると，次の2点に集約される．

(i)　X よりマッチング・データ (M_{ij}) を求め，さらに (M_{ij}^{*}) を計算し，式(3.6)により，各回答者の能力を表すパラメーターの推定値 $\hat{D}_i$ を求める．実際には，各回答者 i について，対応するマッチング・データから $(N-1)(N-2)/2$ 個の推定値が得られるので，これらの幾何平均をもって $\hat{D}_i$ を確定する．

(ii)　式(3.7)の尤度関数を，式(3.8)等を用いて計算し，正答の推定値 $\hat{Z}_k$ を確定する．

簡単なフロー・チャートを図1に示す．

次章で分析するデータ（表1を参照）を如何にして得たかを簡単に示す（データの解釈は次章を参照）．

まず上のステップ(i)により，推定値 $\hat{D}_i$ は，表2のように求められる．次に，ステップ(ii)により，各質問の各選択肢 $s=1,2,\cdots,S_k$ について，式(3.7)の左辺の尤度関数を計算する．その結果は，例えば，質問1の4個の各選択肢について，

$$\Pr(z_1=1 : X)=5.02\times10^{-6}$$

$$\Pr(z_1=2 : X)=9.99\times10^{-1}$$

$$\Pr(z_1=3 : X)=5.33\times10^{-5}$$

$$\Pr(z_1=4 : X)=5.02\times10^{-6}$$

となり，$\Pr(z_1=2 : X)>0.5$ であり，これは式(3.7)を満足するので，質問1の推定正答として選択肢の「2」を採択する．他の質問の正答も同様にして推定できる．表1の「推定正答」の欄は，このようにして得られたのである．表1のデータは，ばらつきが少ないために，多数決の

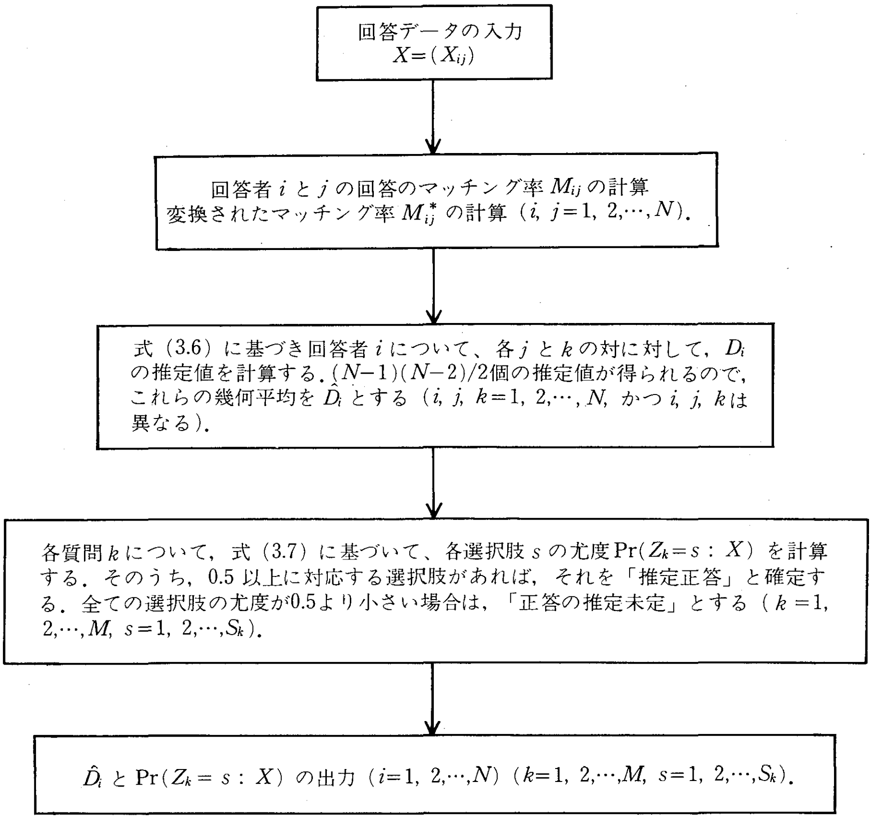

図 1．パラメーター推定のためのプログラムのフロー・チャート．

原理による決定と，このモデルによる正答の推定は完全に一致するが，後に説明するように表 3 のデータでは，これら二つの推定法は異なる結果を得る場合がある．

4. 一般化された HTM モデルの実際のデータ分析への応用

3.1 節でも述べたように，Maher (1987) は，多肢選択型の学力テストに HTM モデルを応用して，そのモデルの妥当性を確認した．また，Romney et al. (1986) は，このモデルを 40 人程度の学生の一般知識テストの回答データや，20 数名のグアテマラの人々に対する各種の病気に関する知識テストのデータを解析したりして，わずか 4 名程度の回答者の回答データからでも「全正答」が推測できる場合があることを確認した．

かなり多数の回答者からのデータがある場合は，単純集計によっても，各質問の各選択肢に対する回答者の選択数の詳細な数値が得られ，それ自体でも潜在的には豊富な情報を持っている．したがって，一見一般化された HTM モデルの適用価値はあまりないと考えられるかもしれない．しかしながら，しばしば，詳細すぎるデータは大局的な傾向を観察するには不適当なこともあろう．本章では，一般化された HTM モデルを既に多数の回答者から得られた単純集計データに適用して，一種の二次分析を試みる．

用いるデータは，内閣総理大臣官房広報室によって，1972 年（昭和 47 年）から 1985 年（昭

和60年)にわたって収集された国民の生活意識に関する調査結果である．詳細な報告は，飽戸他編（1986）を参照していただきたい．このデータは，本来は，特定のモデルや統計手法の利用を第一目的として収集されたわけではなく，各年次毎に標本抽出（層別無作為抽出）された多数の回答者（各年次毎に有効回答者数7,000から8,000名程度）から面接法によって得られたデータである．原著では，性別，世代別に年次比較が行なわれている．さらに，この継続調査データの再分析も行なわれている（鈴木（1988））．したがって，このように豊富な情報を秘めたデータに対して，一般化されたHTMモデルを直接に適用する分析の利用価値が必ずしも大きいとは思われないかもしれないが，一般化されたHTMモデルがどのように機能するかを簡単に描写するために，このデータを利用してみよう．このデータの二次分析としてデータの大局的側面の一部を捉える一つの試みとして，このモデルを利用してみよう．上記のデータのうち，特に用いるものは，1972年から1985年の14年間の調査表から，同一の質問文，同一の回答選択項目の質問で，かつ，生活意識をよく表していると思われるもので，14年間継続調査された11の共通質問項目に対応する回答データである．各質問の回答選択肢数は，質問によって異なっている．したがって，HTMモデルはそのままでは適用できず，3.2節で構成した一般化されたHTMモデルが必要である．

日本語の質問文とその選択肢の例を二つ挙げよう．

例.

質問：お宅の暮し向きは，去年の今ごろと比べてどうでしょうか．楽になっていますか，苦しくなっていますか，同じようなものですか．

1	楽になった	3	苦しくなった
2	同じようなもの	4	わからない

質問：あなたは，今後，生活のどのような面に力を入れたいと思いますか．この中ではどうでしょうか．1つだけあげて下さい．

1	食生活	5	レジャー・余暇生活
2	衣生活	6	その他
3	電気器具，家具，自動車などの耐久消費財の面	7	ない
4	住生活	8	わからない

一般化されたHTMモデルを適用するにあたり，次の2つの場合を考えよう．

(1)　1972年から1985年の間の時代の変化を調べる．
(2)　世代（年令層）による回答の差を調べる．

このモデルが少数の回答者からのデータ分析において特に威力を発揮するものであることを考えると，いま解析しようとしているデータのように多数の回答者からのデータが得られている場合には，その結果は「多数決の原理」に基づく分析と異ならないことが多いであろうから，そのデータをそのまま，このモデルに適用するのは，あまり有益ではない．しかし，ここでは一つの便宜を導入することによって，一般化されたHTMモデルで，このデータの大局的側面を簡便に捉える試みをする．その便宜とは，例えば上記(1)において，各質問について「各年次の回答者全体のうち最大多数が選んだ選択項目」を，その年次に対応する「回答」と定義し，「各年次」を「一人の回答者」と見なすのである．これは一種のデータ変換である．そうするこ

とによって，回答者数は最大 14 人（以下の実際の分析には，ほぼ等間隔の 5 ヶ年次を用いる）となり，一般化された HTM モデルを適用する意義が分かるであろう．この場合,「能力を表すパラメーター D_i」が各年次の傾向の相違を表す指標となることが期待される．つまり，D_i の数値が低いほど,他の年次からの逸脱の程度が高いと見なされ得るであろう．同様の便宜を,上記 (2) の場合にも適用する．即ち，ある特定の年次に着目して，各質問に対して各世代(20 代，30 代，40 代，50 代，60 代以上の各々）の最大選択数に対応する選択肢をもって，その世代の「回答」と定義して,「一つの世代」を「一人の回答者」と見なすのである．

各場合の詳細を以下に述べよう．

4.1　1972 年から 1985 年の間の時代の変化

前述の簡単な変換を施されたデータに，一般化された HTM モデルを適用するために，1972 年から 1985 年の間をほぼ等間隔に分ける 5 ヶ年次として，1972 年次，1974 年次，1978 年次，1981 年次，1985 年次を選択した（注：質問項目数が 11 であるので，14 の全年次のデータを用いると,「回答者数」が「質問数」を上回り，このモデルの利用価値を考えると，必ずしも得策ではない）．各年次の 11 の共通質問に対する回答は表 1 のようになる．

全体として，回答にばらつきは少なく，一致した回答が多い．不一致の場合でも，五人の回答者中，せいぜい一人または二人の「回答者（つまり年次)」が異なる回答をしているにすぎない．推定された正答は，3.2 節の式 (3.7) の尤度計算によって求めたもので，問 9 を除くと，これらの推定正答の尤度は 0.999 以上になった．問 9 では，選択肢「2」が尤度 0.985 で,「1」は尤度 0.014 を示したので,「2」を推定正答に確定した．

もともとのデータが，各質問の選択肢毎の選択率の年次の推移による微妙な変化の情報を含んでいることを考えると，この一種のデータ変換はかなりの情報を失っていることになるように見える．しかしながら，このデータに一般化された HTM モデルを適用した結果を見ると，面白いことが分かる．その結果は表 2 の通りである．

これらの年次中，1974 年次と 1985 年次の D_i の値が他の年次と比べて低いことが分かる．鈴木（1988）は，詳細な情報を持つもとの単純集計データに対して，最小次元解析法の特殊な場合（the ordered class belonging method of minimum dimensional analysis）を適用して，時代の推移の情報を解析している．その一つの結果として，最小次元解析法によって得た二次元表示の第 1 軸は「時代の流れの方向」に対応し，第 2 軸は「経済の好・不況」に対応すると

表 1．1972 年から 1985 年の間の回答の経年的変化．

質問項目 k	1	2	3	4	5	6	7	8	9	10	11
選択肢数	4	8	8	4	6	5	8	4	4	9	4
1972	2	7	7	2	3	2	4	2	2	5	1
1974	3	7	7	2	3	2	4	2	2	5	1
1978	2	7	7	2	3	2	4	2	2	5	1
1981	2	7	7	2	3	2	4	2	1	5	1
1985	2	7	7	2	3	2	4	2	1	3	1
推定正答	2	7	7	2	3	2	4	2	2	5	1

注：表中の数字は，各年次毎に各質問に対しての最多数が選択した選択肢に対応する数字である．推定正答は，一般化された HTM モデルにおいて最尤関数値が 0.5 より大きい選択肢に対応する数字である．

表 2．質問項目数 M に応じたデータから推定された各年次に対応するパラメーター D_i．

年　次	推定された D_i
1972	.961
1974	.885
1978	.961
1981	.938
1985	.861

いう考察をしている(詳細は原著を参照していただきたい). この解析結果と，表2の結果を比べると面白いことが分かる．まず1974年次と1985年次は，その最小次元解析法によって得られた二次元表示の第1軸の両極端に位置していることが分かる．この点においては，これら二つの分析結果は整合している．また，その最小次元解析法によって得られた二次元表示の第2軸の両極端には1972年次と1974年次が位置している．これは，1973年のオイル・ショックという経済上の大事件を考慮に入れると理解に難くはない．表2の1974年次の D_i の値も，この変化を捉えていることが分かる．但し，1972年次の D_i の値は他と比べても決して低いとはいえない．これだけから考えると，この時期の経済的変化はオイル・ショックの直後は負の方向(経済不況)に大きかったが，その後すぐに立ち直り，質的な差はあるかもしれないが，もとの比較的順調な経済成長の流れに戻ったといえるかもしれない．いずれにせよ，本章のようにきわめて単純化したデータに一般化されたHTMモデルを適用し分析した結果が，少なくともある程度は，もとの詳細なデータを用いた最小次元解析法による分析結果の本質的な情報を抽出していることは，このモデルの有効さを示していると考えられるであろう．

4.2 世代（年齢層）による回答の差

ここでは，特に，このデータの中では最新の1985年に着目し，20代，30代，40代，50代，60代以上の五つの世代の回答の差を見るために，前述のようにデータをきわめて単純化し，一般化されたHTMモデルを適用し分析してみる．単純化されたデータは表3の通りである．

この場合も，回答のばらつきは少ない．一つ面白い点は，問7の回答が三つに別れ，しかも回答「4」と「5」が二人ずつの同数であり，多数決の原理では一つの「正答」には決まらないことである．このデータに一般化されたHTMモデルを適用した結果は，表4の通りである(参考のために1972年次の結果も表示してある)．

ただちに分かることは，年齢に応じて D_i の値が単調に減少している点である．つまり，D_i の値は若年世代ほど高く，老年世代ほど低くなっている点である．これは，少なくとも1985年においては，若年世代が日本の世の中の意識，雰囲気を代表しているということなのだろうか．一応，参考のため1972年のデータに同様の分析をすると，30代と60代以上の世代の D_i 値が低くなっている(表4最右列を参照)．1972年と1985年の両方のデータを考えると，1972年に30代で1985年に50歳前後になった人々，つまり昭和10年前後に生まれて，したがって少年期，あるいは思春期を戦中，戦後に過ごした人々が他の年齢層の人々とはかけ離れた意識を

表3. 1985年における各世代の回答.

質問項目 k	1	2	3	4	5	6	7	8	9	10	11
選択肢数	4	8	8	4	6	5	8	4	4	9	4
20代	2	7	7	2	3	2	5	2	1	3	1
30代	2	7	7	2	3	2	4	2	1	3	1
40代	2	7	7	2	3	2	4	2	1	3	1
50代	2	7	7	2	3	2	5	2	1	3	2
60代以上	2	7	7	2	3	2	7	2	1	2	2
推定正答	2	7	7	2	3	2	4	2	1	3	1

注：表中の数字は，各世代毎に各質問に対しての最多数が選択した選択肢に対応する数字である．推定正答は，一般化されたHTMモデルにおいて最尤関数値が0.5より大きい選択肢に対応する数字である．

表4. 1985年次における推定された各世代に対応するパラメーター D_i (参考のため1972年次の推定値も記す)．

年次 \ 世代	推定された D_i	
	1985年次	1972年次
20代	.927	.961
30代	.924	.885
40代	.924	.961
50代	.906	.938
60代以上	.803	.861

持っていることが読み取れるであろう．

また，先述のように質問7の正答は，この五人の回答者による多数決の原理からは定まらないことを指摘したが，このモデルでは尤度関数の計算から「4」と判定されている．より詳しく述べると，

$$\Pr(Z_7=4 \mid X)=0.53>\Pr(Z_7=5 \mid X)=0.44$$

となる．これは，もとの全回答者 7,878 人の選んだ選択肢と比べると面白い．それらの回答者のうち最多数の 27.6% は「5」と回答し，次に多い 25.0% は「4」と回答している．これらの差は，わずか 2.6% である．ただし，1985 年次を除く他の全ての年次においては，「4」が最多数の回答率を挙げているのである．ここでは，この点については，これ以上の考察は進めないが，このような複雑な状況での「正答」の判断や解釈は慎重でなければならないであろうし，上式の尤度関数の値の比較も，この点を示していると考えられよう．即ち，単に最大尤度に対応する選択肢を「正答」と推定するだけで終わるのではなく，他の選択肢に対応する尤度が最大尤度に近い場合は，その点も慎重に考慮に入れることが必要となるであろうということである．なお，他の質問における推定正答の尤度は，質問 11 では 0.984 以上，その他では 0.999 以上となっている．

以上，簡単ではあるが，一般化された HTM モデルの適用例として，既知の社会調査データの二次分析を試みた．

5.　一般化された HTM モデルの将来の発展

Batchelder and Romney (in preparation) では，HTM モデル，或いは，その一般の場合の GCM モデルの幾つかの拡張可能性が論じられている．それには，(1) 各質問の難易度の差を考慮する場合，(2) 回答者集団が，一つの文化（或いは，社会，グループ等）ではなく，複数の文化に属する人々から構成されている可能性がある場合，(3) 回答の選択肢が，連続量，或いは，ランクを持つ尺度で与えられる場合などが含まれている．彼らは，それぞれの場合に応じて，適当なパラメーターを導入して GCM モデルを発展させ，そのパラメーター推定問題について論じている．しかし，例えば，(2) の場合も文化の数が 3 以上の場合はあまり深く考察されていない等，その他の場合も未解決の問題が多い．これら (1)，(2)，(3) の場合のいずれもが，本来の GCM モデルの仮定よりも現実的であり，無視できないのではあるが，これらの場合の条件全てを同時に満たすモデルとそのパラメーター推定問題は取り扱われていない（各個別の場合の推定問題が全て解決できても，それがただちに，全ての条件を満たす場合の推定問題の解決にはつながらないことに注意する）．しかも，現実のデータに対してのモデル選択は，AIC (Akaike Information Criterion) 等のモデル選択原理に基づいているわけではなく，その意味では，任意性が残っている．これらの点を考慮すると，未だ開発の余地が広いと思われる．

ここで，筆者の視点から，GCM (HTM) モデルを発展させるにあたって，特に三つの点を挙げておく．

一つは，面接法における面接者の回答者に対するバイアスの問題に関する．つまり，これは，面接法で回答者から回答を得る際に，面接者個人が持つ意見が意識的にせよ，無意識的にせよ，回答者に対するバイアスとなって何らかの影響を与える可能性があるが，これをどう考慮すべきかという問題である．これに対して，一つの方法として，GCM モデルのバイアス・パラメーター g_i を面接者の影響を表すものとして展開することが考えられる．例えば，「はい，または，いいえ」型のアンケート質問の場合，質問項目 k に対して，回答者 i の面接者 j に対するバイ

アス g_{kij} を

$$g_{kij}=b_i T_{jk}+(1-b_i)(1-T_{jk})$$

と展開できる．ここで，

T_{jk}：面接者 j 自身の質問項目 k に対する回答（意見）
b_i：回答者 i の面接者からのバイアスの受け易さの度合 $(0 \leqq b_i \leqq 1)$

である．$b_i=1$ であれば，回答者 i の質問 k に対する回答は，面接者 j の回答に一致するようなバイアスが最大となり，$b_i=0$ であれば面接者 j の回答の反対となるようなバイアスが最大となり，$b_i=1/2$ であればバイアスなしとなる．

このモデルは，多数の面接者が大多数の回答者からデータ収集する時は，実践的には必ずしも有益とはならないかもしれない．しかし，本論文で焦点を当てたように，能力差が大きいかもしれない少数の回答者たちから，ごく少数（場合によっては，一人）の面接者が得たデータから，より信頼性の高い情報を引き出すのには役立つかもしれない．筆者は，このためのモデルとパラメーター推定法を現在開発中である．

二つめの点は，一般化されたHTMモデルの自由回答の分析への応用に関するものである．アンケート調査においては，自由回答とは，例えば次のような質問について回答者が自由に(つまり，選択肢に拘束されずに）回答することをいう．

例.
質問：あなたにとって一番大切と思うものはなんですか．1つだけあげて下さい．

この場合は，選択肢の番号が回答となるわけではないから，HTMモデルは，そのままでは適用できない．しかしながら，自由回答問題と同時に与えられる多肢選択問題の回答パターンを考慮して，さらに，上記の例のような自由回答質問で回答者の列挙した回答（言葉）の一つ一つをあたかも選択肢（「選択肢数」は，多数となるかもしれないが，有限ではある）に対応するかのように考えると，一般化されたHTMモデルが適用できる．これは，Batchelder and Romney (1988) の扱ったGCM, HTMモデルでは不可能であるが，本論文で発展させた一般化されたHTMモデルでは，可能である．もし，さらに多文化のためのモデル（この章の冒頭の (2) の場合）が適用できると，多肢選択問題の回答パターンによって回答者をグループ分類し，それぞれのグループの回答者たちが自由回答で挙げた各語を一種のカテゴリー化することができよう．これによって，辞書的に同意である語の対（例えば，寺と寺院）が場合によっては，異なるカテゴリーに属すると判定したり，逆に辞書的には全く異なる語の対が同じカテゴリーに属すると判定することもあろう．この研究も，現在進行中である．

最後の点は，このモデルの現実的な価値のある応用に関する（注：この点は，本誌の編集長と査読者の方の貴重な助言に基づく)．第4章では，多数の回答者からのデータに単純な変換を施し，その二次分析として，一般化されたHTMモデルを適用した．しかし，本論文で繰り返し強調したように，この種のモデルは，少数の回答者しか得られない場合に本来の威力を発揮するものである．「少数の回答者しか得られない」というのは，物理的，或いは経済的コストの制約のある場合だけとは限らない．例えば，近い将来の経済動向を，優れた経済の専門家や大企業の経営者の意見を統合して判断しようとする場合を考えると，まず，そのような専門家や経営者の全体の数そのものが限られていることが分かる．そして，経済の動向の詳細な点に関しては，専門家たちの間でも意見が分かれることは少なくない．このような場合こそ，本論文で紹介したモデルが有益なものとなるのである．現在のところ，筆者はその種のデータを分

析してはいないが，多くの研究者がこれらのモデルを実践的に利用し，多様な分野で有益な情報を得ることを望みたい．

本論文では，アンケート調査における回答の回収率の低下の問題を解決する一つの試みとして，幾つかのモデルが展開された．これらのモデルは，多くの仮定のもとで構成された．推定パラメーターの信頼性を高めるのに必要とされるデータの量と，モデルの仮定（制限）の強さとの間には，一種のトレイド・オフがあり，これはやむを得ないことである．重要な問題は，如何に現実的ではない仮定を排して，合理的な仮定のみを取り入れるかということである．今後，より多くの実証的データに基づいて試行錯誤が繰り返されて，この論文で扱ったモデルが発展していくことを望みたい．

謝　辞

本論文を執筆するにあたり，ご指導くださった鈴木達三教授，および匿名の査読者の方々に感謝いたします．また，本論文の内容は，筆者がカリフォルニア大学アーヴァイン校（University of California, Irvine）に在学中にリサーチ・アシスタントとして W.H. Batchelder 教授と A.K. Romney 教授とともに研究したことが基礎となっており，両教授に感謝いたします．但し，本論文の特定のアプローチ，分析等は全て筆者に責任があり，もし，それらに誤りがあるとすれば，全て筆者の負うところである．

参 考 文 献

飽戸　弘，佐藤　彰，鈴木達三，中西尚道，林知己夫 編(1986).「国民生活に関する世論調査」の意識動向の分析視点，分析手法，適用例の研究，内閣総理大臣官房広報室研究報告，昭和 61 年 3 月．

Batchelder, W.H. and Romney, A.K. (1988). Test theory without answer key, *Psychometrika*, **53**, 71-92.

Batchelder, W.H. and Romney, A.K. (in preparation). New results in test theory without an answer key.

Blackwell, H.R. (1963). Neural theories of simple visual discriminations, *J. Opt. Soc. Amer. A*, **53**, 129-160.

Green, D.M. and Swet, J.A. (1966). *Signal Detection Theory and Psychophysics*, Wiley, New York.

Grofman, B. and Owen, G. (eds.) (1989). *Information Pooling and Group Decision Making*, JAI Press, Greenwich, Connecticut.

林知己夫，川本　勝，児島和人，佐藤　彰，鈴木達三 編(1989). 今後の世論調査の在り方についての研究，内閣総理大臣官房広報室 研究報告，平成元年 3 月．

Lazarsfeld, P.F. and Henry, N.W. (1968). *Latent Structure Analysis*, Macmillan, New York.

Lord, F.M. and Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*, Addison-Wesley, Resding, Massachusetts.

Maher, K.M. (1987). A multiple choice model for aggregating group knowledge and estimating individual competencies, Unfinished doctoral dissertation, University of California, Irvine (available upon request).

西平重喜 (1979). 統計調査法（補訂版），新数学シリーズ，8　培風館．

Romney, A.K., Weller, S.C. and Batchelder, W.H. (1986). Culture as consensus : A theory of culture and information accuracy, *American Anthoropologist*, **88**, 313-339.

鈴木達三(1988). 継続調査データの再分析 ― 時代区分の考察と意見構造の経年的推移 ―，マーケティング・リサーチャー，**49**（夏季号），62-73.

An Extension of the "Test Theory without Answer Key" by Batchelder and Romney and Its Application to an Analysis of Data of National Consciousness

Ryozo Yoshino

(The Institute of Statistical Mathematics)

Questionnaire type inquiry is often used as a simple method of social survey, especially by anthoropologists and sociologists. Recently, however, the collection rate of responses from informants is rapidly decreasing, probably because of some problems concerning the right of privacy or the rapidly changing world. In order to overcome this difficulty, one needs to develop a way to increase the collection rate or a method to obtain more reliable information from responses of a small number of informants.

The main objective of this paper is to extend the test theory by Batchelder and Romney (1988) so that one may apply the extended theory to questionnaire type inquiry even when the number of informants is not necessarily large and the competencies of those informants are not necessarily uniform. For this objective, a model called a generalized high threshold model (HTM) is constructed, and a parameter estimation problem of the model is solved. As an application of this model, it is used to analyze a set of data by a long term survey of the Japanese government on the national consciousness. Finally, some suggestions are given for the possible future development of the generalized HTM: a way to consider the biases of interviewers on informants and a way to apply this model to the analysis of responses of free-answer questionnaire.

Key words: Test theory, multiple choice questionnaire, Condorcet model, the method of maximum likelihood.