連載講座

マーケティング・サイエンスII

消費者行動のモデル化：消費者の異質性

阿部　誠

1. はじめに

マーケティングの目的は，顧客の知覚，選好を理解し，適切な商品をデザイン・販売して，ユーザーの元に届けることである．マーケターは，どのような商品・サービスを（Product）いくらで提供し（Price），どのような広告や販売促進を行って（Promotion），どの流通経路で販売するか（Place），という4Pの要素を決めなければならない．成熟した経済社会で特に重要なことは，個々の消費者（顧客）の違い―商品に関する選好やマーケティング刺激に対する反応の異質性―を十分に認識し，それに適切に対応することである．マーケティングでは，セグメンテーション，ポジショニング，ターゲティングと呼ばれる差別化された商品の提供や顧客によって異なったマーケティング活動などが，早い時期から行われてきた．

近年の情報技術の発達により1人1人の顧客データを集計せずに容易に収集，保存できるようになったおかげで，その傾向はますます重要になっている．例えば，POSシステムにフリークエント・ショッパーズ・プログラム（FSP）を組み合わせることによって，顧客の購買履歴を時系列的に収集できる．インターネットなどでは，顧客のとったアクション―カタログ請求，問い合わせ，購買―はもちろん，購入前に閲覧されたページ履歴までがログファイルに自動的に蓄積される．これらの膨大なデータが集計されずに保存されているということは，1人1人の顧客を深く理解し，より効果的なマーケティングを実践するための情報が溢れているということだ．平均的消費者を想定してのマーケティングは，今日，ほとんど無意味な概念になりつつある．今やマーケティングは情報産業であり，それを有効に生かす企業こそが競争優位に立つのである．

あべ　まこと
東京大学 大学院経済学研究科 経済学部
〒113-0033 文京区本郷 7-3-1

スキャン・パネルデータやインターネットなどから収集された個人レベルの購買履歴を分析する場合，消費者の行動を完全に説明，予測することは不可能なため，一般に確率的モデルを用いる．これらのモデルは購買行動のどの現象をモデル化するかによって大きく五つに分類できる．

(1) 購買発生，購買タイミング
あるカテゴリーの購買が一定期間に起きるか，またはいつ購買が発生するか？
(2) 店舗選択
どの店舗で購買するか？
(3) 認知集合，考慮集合，選択集合
カテゴリー内で認知，考慮，選択の対象となるブランドはどれか？
(4) ブランド選択
どのブランドを選択するか？
(5) 購入量決定
あるブランドが購買された時に，どのくらい購入するか？

上記の購買現象を複数，統合したモデルの研究も多数存在するが，本稿では消費者行動の異質性をモデル化するというトピックで，前回の講座で紹介されたロジット・モデルに代表される確率的ブランド選択モデルに焦点をあてて，最近のマーケティング・サイエンスにおける学術研究の流れを紹介する．

2. 異質性をモデル化するアプローチ

確率的効用モデルによると，消費者 n が t 回目の選択（購買）機会に集合 C_{nt} からブランド i を選択する確率は，

$$P_{nt}(i)=\Pr[U_{nti}>U_{ntj}] \quad \forall j\in C_{nt}, j\neq i \tag{1}$$

$$U_{nti}=V_{nti}+\varepsilon_{nti} \tag{2}$$

のように表される．ここで，U_{nti} は選択肢 i に対する効用，V_{nti} はそのうちの確定的部分，ε_{nti} は確率的部分である．ここで ε_{nti} が独立に同一の第1種極値分布（正規分布に似ているが左右非対称）に従うと仮定す

ると，シンプルなロジット・モデルが導かれる．

$$P_{nt}(i)=\frac{\exp(V_{nti})}{\sum_{j\in C_{nt}}\exp(V_{ntj})} \tag{3}$$

V_{nti} は通常，価格やプロモーションなどのマーケティング変数ベクトル x_{nti} とその影響度を表すパラメータ・ベクトル β_n の線形結合で式(4)のように規定される．

$$V_{nti}=\beta_n' x_{nti} \tag{4}$$

β_n は説明変数が確定的効用に与える影響度を表すため，消費者の異質性を考慮してサブスクリプト n が付随している．しかしブランド選択モデルが適用される多くの消費財カテゴリーでは，1～2年間の実験期間中に1人の消費者から観測されるブランド選択の回数は限られているため，通常，個人ごとに異なった β_n を推定することは不可能である．全ての消費者で β_n は共通だと仮定すれば，$\beta=\beta_n\ \forall n$ は最尤法によって推定できる．さらに，尤度関数は凹関数なので Newton-Raphson 法などの単純な最適化アルゴリズムで全体の最適解が求められることが保証される．しかし，これでは異質性の影響を無視していることになる．

経済学では政策変数が全体集合に与える影響を集計的に推定するのが主な興味の対象である．その際，個人間の異質性というものは推定にバイアスを与える厄介な問題として克服されなければならなかったが，それ自体は論争の中心にはなりえなかった．例えば確率変量モデル（random coefficient model）ではパラメータに分布を仮定することによって異質性を説明しているが，計量経済での関心はこの分布の中心値とバラツキ（平均と分散）を正確に推定することであり，個々人のパラメータの推定にまで足を踏み込むことは稀である．これに対して，One-to-One マーケティングや Customer Relationship Management（CRM）では1人1人個別に働きかけることも多いため，顧客ごとにユニークなパラメータの値を知ることは実務上，非常に有益である．

この場合，当然，パラメータの数に対してデータ量が激減するために，通常の手法では推定不可能であったり，可能であっても推定値の不確実性が無視できない程に大きくなる．また個人別パラメータを点推定できたとしても，この値を確率モデル式(3)に代入して選択確率を計算してしまうと，パラメータがモデルに対して非線型なためバイアスのある確率が推定されてしまう．したがって，パラメータの不確実性を考慮に入れたうえで確率を推定し，それに基づいて最適なマーケティング政策を計画・遂行する必要がある．

式(4)の効用関数を，消費者の異質性を考慮しながら推定できる形で拡張するには大きく分けて二つのアプローチが考えられる．一つはパラメータを消費者間で共通に設定して説明変数に個人の違いを表す変数を加える方法，もう一つは説明変数は共通だがパラメータは異質性によって点ではなく分布されていると仮定する方法である．前者のアプローチでは，過去のブランド選好から個人がそれぞれのブランドに対して持っている特有な選好を指標化したブランド・ロイヤルティーと呼ばれる変数や，消費者の特性（デモグラフィック変数）などを加える．後者のアプローチは，確率モデルのパラメータ自体が分布をしているのでミクスチャー・モデルとも呼ばれ，さらにパラメータの分布が離散的か連続的かにより二つに分類できる．

スキャン・パネルデータを用いたブランド選択におけるロジット・モデルの研究は1980年代の初期に始まったのだが，非集計データのメリットを生かして消費者の異質性をモデル化するために，まずは前者のアプローチが使われた[1]．その限界が見えた80年代の後半，パラメータに離散的分布を仮定したミクスチャー分布モデルが提唱された．このアプローチの特別なケースである潜在クラス・モデルは，マーケティングでのセグメンテーションの概念と解釈できるため，しばらくは異質性をモデル化する手法の主流であった[2]．パラメータに連続的分布を仮定したモデルは，計量経済では確率変量モデルとして70年代から知られていたが[3]，その推定（最尤法が使われた）は複雑でマーケティングの分野では非実用的なものであった．しかし90年代半ばのベイジアン統計学の進展から，パラメータを計算機の能力を生かしたシミュレーションによって推定する方法が可能となり，現在の研究では，階層ベイズと呼ばれるパラメータに連続的分布を仮定したミクスチャー分布モデルが主流となっている[4]．

以下の節では，マーケティング・サイエンスにおいてロジット・モデルに消費者の異質性を考慮する代表的なアプローチとして，(1)ブランド・ロイヤルティーを説明変数に使ったモデル，(2)パラメータに離散的分布を仮定した潜在クラス・モデル，(3)パラメータに連続的分布を仮定した階層ベイズ・モデル，と年代順に紹介していこう．

3. ブランド・ロイヤルティー

データベースに各消費者の特徴を示す情報が存在すれば，消費者の異質性をモデルに組み込む上でまず考えられることは，それらを説明変数として加えることであろう．これらの情報は，デモグラフィック属性，ライフスタイルや性格診断，心理学的プロファイル，過去の購買履歴（実際の行動）などに分類できる．早い時期からブランド選択データがスキャン・パネルデータで収集，分析されてきた日用消費財においては，デモグラフィック属性はブランド選択をほとんど説明できないことが多くの研究で立証されている（例えば文献[5]）．例えばプレミアム・ブランドのオレンジジュースは，収入の高低にほとんど関係なく購買されるのである．また，価格意識の高い消費者の方が特売で購買する傾向が高いなど，常識的なライフスタイルや心理学的プロファイル属性の方が，多少なりともブランド選択に影響を与えることも分かっている（例えば文献[6]）．

これらの顧客情報がブランド選択へ与える影響度の違いは，消費者行動理論で提案されているブランド選択プロセスを考慮すると理解できる．一般に，ブランド選択には図1のような因果プロセスがあると考えられており，デモグラフィック情報は実際の行動から一番離れている要因で，心理学的プロファイル，ライフスタイル，性格などを通じて，ブランドに対する知覚と態度を形成し，それが実際の購買行動に影響を与えている．

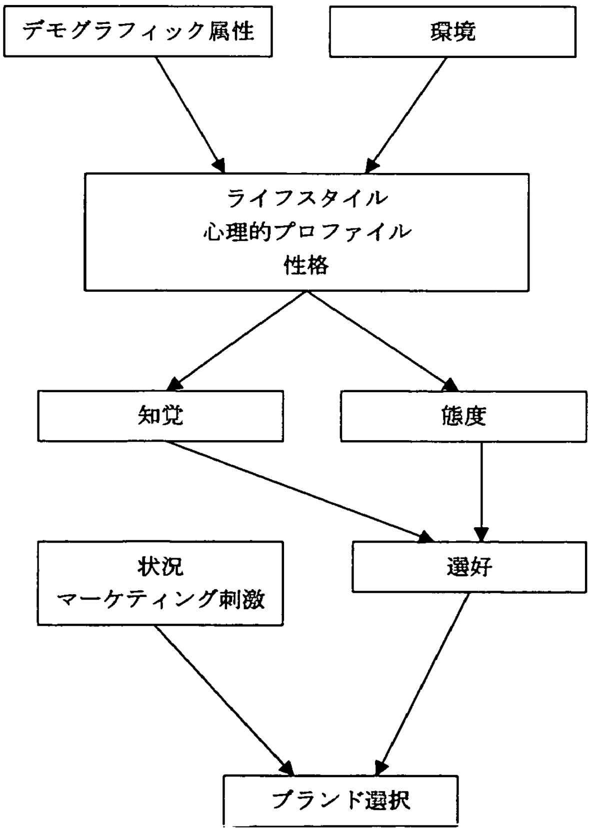

図1 消費者行動論によるブランド選択要因

顧客のメールリストからゲーム感覚的なアンケート調査をかけてライフスタイルや心理学的プロファイルを収集することも非現実的ではないが，通常，小売店のFSPなどには性別，年齢，住所などの限定されたデモグラフィック属性しか存在しないため，実務での活用はあまり現実的でない．また，図1から分かるように，ブランド選択行動は過去の購買行動やマーケティング刺激（situational constraint）に一番強く影響される．これらの情報はコンピュータに自動的に蓄積されるというデータ収集上の利点もあるため，ロジットモデルの効用関数にはマーケティング変数と過去のブランド選択を説明変数として含めるのが一般的である．

過去のブランド選択行動は，その消費者独自の個々のブランドに対する選好度の目安になるため，異質性をモデル化するのに役立つ．その中でも多くの研究者に採用された代表的な指標が，過去の購買記録をラグ項に取り入れて個々のブランドに対する選好を消費者の購買機会ごとに更新したブランド・ロイヤルティーという変数である．

これはGuadagni and Little[1]によって最初に提案され，ブランドの選好に対する消費者間の異質性とその動的変化を取り込む役割を果たす．消費者nのブランドiに対するt回目の購買機会のロイヤルティー，$loyalty_{ni}(t)$，は

$$loyalty_{ni}(t)=\lambda\, loyalty_{ni}(t-1)+(1-\lambda)y_{ni}(t-1)$$

$$\text{where}\quad y_{ni}(t)=\begin{cases}1 & \text{消費者 } n \text{ が } t \text{ 回目の購買機会にブランド } i \text{ を選択した場合}\\ 0 & \text{それ以外の場合}\end{cases}$$

と定義される．したがって，ブランドiに対するロイヤルティーは，その消費者が過去にどのブランドを選択したかを直近の購買により重みを付けた指標となっている．繰り越し係数λはデータから最尤法によって0.8と推定された．ロイヤルティー変数はラグ項を含んでいるため，通常，ロジット・モデルを推定するデータより前の期間のデータを予め確保しておいて初期化する．

パッケージ商品業界においては，ロイヤルティーの

ような過去の購買に基づいた変数は，デモグラフィック情報に比べて，消費者間の異質性をより説明できるため広く使われている．ここで紹介したブランド・ロイヤルティーはその1例であるが，シンプルな構造上いくつかの問題点も抱えている．ロイヤルティーはあくまでも過去に観察されたブランド選択行動に基づいているため，それが選好によるものなのか，あるいは値引，広告，プロモーションのようなマーケティング刺激によるものなのかが分離できない．さらに，消費者は特別な理由がなければリスクをさけて同じブランドを選択し続けるという習慣形成の要因も含めているため，これらの要因を分離させるために様々な拡張バージョンが提案されている[7].

4. 潜在クラス・モデル

4.1 モデルのフレームワーク

説明変数の影響度に対する消費者の異質性をモデル化するために，個々のパラメータに離散的分布を仮定したのが有限ミクスチャー・モデルである．その中でも，(a)全てのパラメータで確率点の数が同一（degenerate の場合も含めて S 個）で，(b)パラメータの同時密度が完全に相関している（相関行列が単位行列）場合は式(5)のように表せる．

$$P_{nt}(i|\pi,\beta)=\sum_{s=1}^{S}\pi_s P_{nt}(i|\beta_s)$$
$$\text{where}\quad \pi=[\pi_1,\cdots,\pi_S],\quad \beta'=[\beta'_1,\cdots,\beta'_S] \tag{5}$$
$$\sum_{s=1}^{S}\pi_s=1\quad \text{where}\quad \pi_s\geq 0\quad \forall s=1,\cdots,S$$

式(5)は，セグメント s ごとにパラメータ・ベクトル β_s の値が異なったロジット・モデル $P_{nt}(i|\beta_s)$ とそのサイズ π_s をデータから導き出すために，潜在クラス・モデルと呼ばれる．実務におけるセグメンテーションの概念と一致することから，マーケティングでは非常によく用いられるモデルである．アプローチが統計モデルベースなので，ヒューリスティックなクラスター分析などと違い仮説検定などが適用できるという利点がある．

モデルのパラメータ π と β が推定されたら，消費者 n が個々のセグメントに所属する確率 $p_n(s)$ を計算することができる．セグメント・サイズ π を所属の事前確率として，消費者 n のブランド選択データ履歴 $y_{ni}(t)(t=1,\cdots,T_n)$ から計算された尤度 $f_n(y_n|\beta_s)$ をベイズ定理に基づいて組み合わせることによって，所属の事後確率 $p_n(s)$ は式(6)のように求められる．

$$p_n(s)=\frac{f_n(y_n|\beta_s)\pi_s}{\sum_{u=1}^{S}f_n(y_n|\beta_u)\pi_u}$$
$$\text{where}\quad f_n(y_n|\beta_s)=\prod_{t=1}^{T_n}\prod_{i\in C_{nt}}p_{nt}(i|\beta_s)^{y_{ni}(t)} \tag{6}$$

潜在クラス・モデルではロジット・モデルのパラメータ β_s とセグメント・サイズ $\pi_s(s=1,\cdots,S)$ を最尤法によって推定するが，解を求めるには二つのアプローチがある．一つは Newton-Raphson 法などの通常の数値最適化アルゴリズムによって π_s と β_s を同時に探索するアプローチ，もう一つは Expectaion-Maximization (EM) アルゴリズムによって π_s と β_s を交互に探索するアプローチである．前者は解の収束までの反復回数が比較的少ない，バイプロダクトとしてパラメータの漸近標準誤差が得られるなどの利点があるが，尤度関数の計算に手間がかかる，尤度関数が複雑な形状をしている場合には収束しない，などの弱点がある．それに対して後者は，反復回数は多くなる傾向にあるが，計算が比較的簡単である，少なくとも局所最適解には収束するなどの利点から，前者より頻繁に使われる．本稿では主に OR 分野の読者を想定しているため，EM アルゴリズムによる最尤法[8]について少し詳細に説明しよう．

4.2 EM アルゴリズム

EM アルゴリズムは，全てのパラメータを同時に最適化するという複雑な作業をする代わりに，観測されない架空の変数を導入することによってパラメータをサブセットに分け，それぞれの最尤値が独立に容易に計算できる場合に用いられる．潜在クラス・モデルにおいては，消費者がどのセグメントに属するかは分からないが，もし分かれば，そのセグメントのロジットのパラメータの推定は簡単にできる．したがって，ここでの観測されない架空変数は消費者のセグメントに対する所属 z_{ns} を表し，消費者 n がセグメント s に属する場合は1，属さない場合は0になる．z_{ns} は欠損値として扱い，観測されたデータ（ブランド選択）と現時点でのパラメータの推定値から z_{ns} の値を推測して，その z_{ns} に基づいてパラメータの推定値を更新するという反復作業を，尤度の向上が見られなくなるまで継続するのである．E ステップでは欠損値の期待値を計算し，M ステップではこの期待値に基づいてパラメータの最尤値を求め，この最尤値から再度，欠損値の期待値を更新する（E ステップ）というプロセスを繰り返すので，EM (expectation-maximization) アルゴリズムと呼ばれている．以下では，具体

的に潜在クラス・モデルにおけるＥステップとＭステップを説明する．

まず，全ての消費者の所属セグメントが分かっていれば z_{ns} は既知と仮定できるため，全体の尤度関数とその対数尤度関数はそれぞれ式(7)と式(8)のように表せる．

$$L(\pi, \beta|y, z)=\prod_{n=1}^{N}\prod_{s=1}^{S}(f_n(y_n|\beta_s)\pi_s)^{z_{ns}} \tag{7}$$

$$\ln L(\pi, \beta|y, z)=\sum_{n=1}^{N}\sum_{s=1}^{S}z_{ns}\ln f_n(y_n|\beta_s) + \sum_{n=1}^{N}\sum_{s=1}^{S}z_{ns}\ln \pi_s \tag{8}$$

Ｅステップ

対数尤度式(8)の期待値を求めるには，欠損値 z_{ns} をその期待値で代用すればよい．現時点でのパラメータの推定値に基づいた z_{ns} の期待値は，ベイズ定理を使った式(6)の消費者がセグメントに所属する事後確率 $p_n(s)$ を用いる．対数尤度の期待値は式(9)になる．

$$E_z[\ln L(\pi, \beta|y, z)]=\sum_{n=1}^{N}\sum_{s=1}^{S}p_n(s)\ln f_n(y_n|\beta_s) + \sum_{n=1}^{N}\sum_{s=1}^{S}p_n(s)\ln \pi_s \tag{9}$$

Ｍステップ

期待対数尤度が最大になるようなパラメータ π と β を推定するのだが，式(9)で分かるように π は右辺の第２項のみに，β は第１項のみに表れるので，それぞれ個別に最大化できる．

まず π の最大化は，$\sum_s\pi_s=1$ という制約を付けてラグランジュ方程式式(10)を解くと，ラグランジュ係数 λ は N となり，式(11)の解が得られる．

$$Lagrangean=\sum_{n=1}^{N}\sum_{s=1}^{S}p_n(s)\ln \pi_s-\lambda\{\textstyle\sum_s\pi_s-1\} \tag{10}$$

$$\hat{\pi}_s=\frac{\sum_{n=1}^{N}p_n(s)}{N} \tag{11}$$

β はセグメントで独立なので，セグメントごとに式(12)を最大化する β_s を求めればよい．

$$L_s(\beta_s)=\sum_{n=1}^{N}p_n(s)\ln f_n(y_n|\beta_s) =\sum_{n=1}^{N}\sum_{t=1}^{T_n}\sum_{i\in C_{nt}}p_n(s)y_{ni}(t)\ln p_{nt}(i|\beta_s) \tag{12}$$

$L_s(\beta_s)$ は，通常のセグメント化されていないロジット・モデルに対して，ブランド選択データ $y_{ni}(t)$ の代わりにそれに消費者 n のセグメント s への所属確率で重み付けしたデータ $p_n(s)y_{ni}(t)$ を適用した尤度関数と解釈できるため，β_s の推定は容易にできる．

EM 反復プロセス

EM アルゴリズムは上記のＥステップとＭステップの反復によって，最適解に近づいていく手法である．通常は，分析者がセグメントの数 S を設定することから始める．そして初期値として，消費者がセグメントに所属する事後確率 $p_n(s)$ をランダムに発生させて，まずＭステップで β の推定値を計算し，それに基づいてＥステップで事後確率 $p_n(s)$ を再更新する，というプロセスを対数尤度の向上が基準レベル以下になるまで繰り返す．EM アルゴリズムの利点として，反復ごとに尤度が必ず向上すること，そして一般的条件の下では収束が保証されていること，が挙げられる．EM アルゴリズムの留意点としては，

(a) 局所最適解の存在：複数の適切な初期値を使うことである程度解決できる．

(b) セグメント数の決定：セグメント数の異なったモデルを推定して，AIC，CAIC，BIC など情報量の基準からデータに対して一番フィットがよいモデルを採択するのが通常の方法である．ここで注意しなければならないのは，制約付きモデルの統計的仮説検定に使われる漸近カイ２乗テストは有効ではないことだ．これは制約モデルがパラメータ・スペースの境界線にあり一般的条件を満たさないからである．

(c) パラメータの標準誤差：最尤法と違い EM アルゴリズムではフィッシャー情報行列が自動的に得られないため，尤度関数のヘシアン行列は収束解が得られた後に別途計算する必要がある．

潜在ロジット・モデルの拡張としては，消費者のセグメント所属確率 $p_n(s)$ をデモグラフィック変数に関連付ける concomitant variable segmentation などが提案されている[5]．

5．階層ベイズ・モデル

5.1 消費者異質性のベイズ的解釈

説明変数の影響度に対する消費者の異質性をモデル化するために，個々のパラメータ β_n に連続的な分布を仮定したのが確率変量モデルである．例えば線型効用関数を用いたロジット・モデルでは，β_n は平均 μ_β，分散 V_β の多重正規分布 $N(\beta_n|\mu_\beta, V_\beta)$ からランダムに抽出されたと仮定することが多い．

β_n 値が与えられた状況でのロジット・モデルによる条件付選択確率 $P_{nt}(i|\beta_n)$ を，この正規分布したパラメータで積分すれば，式(13)のように無条件な選択確

率が求められる．

$$P_{nt}(i)=\int_{\beta_n} P_{nt}(i|\beta_n)N(\beta_n|\mu_\beta,\ V_\beta)d\beta_n \tag{13}$$

経済学では平均的消費者の特性と異質性の指標が主な関心であったため，パラメータの中心値とバラツキ，つまり平均 μ_β と分散 V_β を大量のクロスセクション・データから最尤法によって点推定するのが目的であった．これに対して，1人1人の消費者に個別に働きかけることの多い新しいタイプのマーケティングでは，個人特有のパラメータの値を知ることとがより有益である．しかし，個人特有のパラメータは基本的にはその個人のデータから推定するので，データ量が絶対的に不足しておりパラメータの推定値にもそれなりの不確実性が伴う．マーケティングでは，この小サンプルによる不確実性を正確に把握して，それを意思決定に反映させることが実践で重要になる．平均や分散のような少数のパラメータを大勢の消費者のプールされたデータから点推定して漸近理論に基づいて yes/no の仮説検証を行うのとは根本的に異なり，「パラメータ自体は分布をもった確率変数」と考えるベイズ的アプローチがここでは特に有効なのである．

ベイズ統計は，β_n に事前分布 $f(\cdot)$ を仮定し，消費者 n のデータに基づいた尤度関数 $f_n(y_n|\beta_n)$ を式(14)のように組み合わせることよって，事後分布 $f(\beta_n|y_n)$ を導き出すという単純な概念に基づいている．

$$f(\beta_n|y_n)=\frac{f_n(y_n|\beta_n)f(\beta_n)}{\int_{\beta_n} f_n(y_n|\beta_n)f(\beta_n)d\beta_n}\propto f_n(y_n|\beta_n)f(\beta_n) \tag{14}$$

今回の例では，パラメータ β_n の事前分布は，$f(\beta_n)=N(\beta_n|\mu_{\beta n},\ V_\beta)$ とさらに正規分布の平均と分散パラメータによって規定されるため，μ_β と V_β はハイパー・パラメータとも呼ばれる．

ベイズ統計で議論になるのは事前分布をどのように指定するか，あるいはハイパー・パラメータをどう設定するかであるが，大きく分けて3種類の方法がある．

(a) 古典的ベイズ
分析者の経験や主観から設定するが，そういった影響を極力，避けたい場合は事前情報を含まない無情報事前分布や拡散事前分布が使われる．

(b) 実験ベイズ (Empirical Bayes)
実験ベイズでは，事前分布のハイパー・パラメータを表本理論や頻度論に基づいてデータから推定する．上記の例でいうと，消費者をプールしたデータから共通の β を最尤法によって点推定し，その値と標準誤差の倍数をそれぞれ μ_β と V_β として用いる[9]．この方法は直観的で計算も比較的容易であるが，(1)ハイパー・パラメータの推定誤差を考慮していない，(2)個人別パラメータを推定する時に事前分布と尤度にデータの重複があるため厳密な意味でベイズのフレームワークから外れている，(3)推定された事前分布の効果が漸近的にゼロに収束しない，などの問題がある．

(c) 階層ベイズ (Hierarchical Bayes)
実験ベイズの問題点は，ハイパー・パラメータ μ_β と V_β をさらにベイズによって推定するという2段階プロセスを使えば克服できる．次節では，階層ベイズの詳細をロジット・モデルを使って説明しよう．

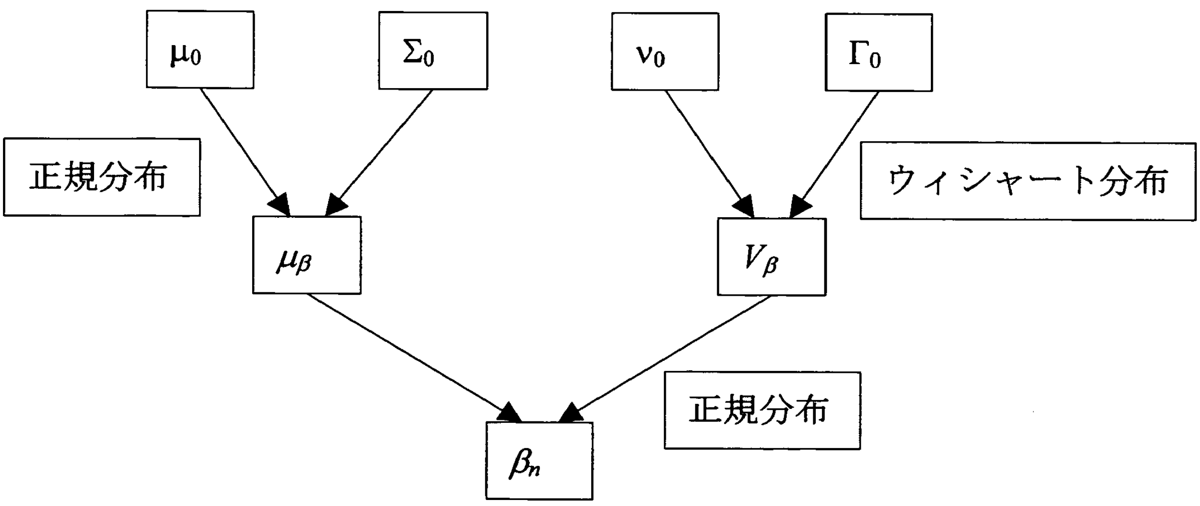

図2 階層ベイズ・モデルでのパラメータの関係

5.2 ロジット・モデルにおける階層ベイズ・モデル

モデルのパラメータは$\{\beta_n\ \forall n, \mu_\beta, V_\beta\}$で，それぞれベイズ的に推定されるため，事前分布が式(15)に，それらの関係が図2に示されている．

$$\begin{aligned} &\beta_n \sim N(\mu_\beta, V_\beta) && N=\text{normal distribution} \\ &\mu_\beta \sim N(\mu_0, \Sigma_0) && N=\text{normal distribution} \\ &V_\beta \sim IW(v_0, \Gamma_0) && IW=\text{inverse Wishart distribution} \end{aligned} \quad (15)$$

$\{\beta_n\ \forall n, \mu_\beta, V_\beta\}$は，適当な初期値から始めて，逐次的に一つずつ他のパラメータの値が与えられた条件で推定するというプロセスを繰り返す．つまり，$\{\beta_n\ \forall n, V_\beta\}$から$\mu_\beta$を推定し，その値を含んだ$\{\beta_n\ \forall n, \mu_\beta\}$から，$V_\beta$を推定し，それらの値を含んだ$\{\mu_\beta, V_\beta\}$から$\beta_n\ \forall n$を推定する，というプロセスを推定値が収束するまで反復するのである．今回の推定値は前回の推定値のみに依存し，一般的条件の下では定常分布がパラメータの同時密度に収束するため，マルコフチェーン・モンテカルロ（MCMC）シミュレーション法と呼ばれている．それぞれのステップは以下のようになる．

［a］　$\{\beta_n\ \forall n, V_\beta\} \Rightarrow \mu_\beta$

μ_βは尤度関数に対して共役な事前分布（conjugate prior），正規分布を採用しているため，事後分布も同じく正規分布になっている．

μ_βの事前分布$N(\mu_0, \Sigma_0)$：正規分布

尤度関数$f(\beta_n\ \forall n|\mu_\beta, V_\beta)$：正規分布$N(\mu_\beta, V_\beta)$から$\beta_n(n=1, \cdots, N)$が観測される確率

$$f(\mu_\beta|\beta_n\ \forall n, V_\beta) \propto f(\beta_n\ \forall n|\mu_\beta, V_\beta)N(\mu_\beta|\mu_0, \Sigma_0)$$

から，

μ_βの事後分布$N(\mu_N, \Sigma_N)$：正規分布

$$\mu_N=(\Sigma_0^{-1}+NV_\beta^{-1})^{-1}(\Sigma_0^{-1}\mu_0+NV_\beta^{-1}\bar{\beta})$$

$$\Sigma_N^{-1}=\Sigma_0^{-1}+NV_\beta^{-1}$$

$$\bar{\beta}=\frac{1}{N}\sum_{n=1}^{N}\beta_n$$

［b］　$\{\beta_n\ \forall n, \mu_\beta\} \Rightarrow V_\beta$

V_βは尤度関数に対して共役な事前分布，インバース・ウィシャート分布を採用しているため，事後分布も同じくインバース・ウィシャート分布になっている．

V_βの事前分布$IW(v_0, \Gamma_0)$：インバース・ウィシャート分布

尤度関数$f(\beta_n\ \forall n|\mu_\beta, V_\beta)$：正規分布$N(\mu_\beta, V_\beta)$から$\beta_n(n=1, \cdots, N)$が観測される確率

$$f(V_\beta|\beta_n\ \forall n, \mu_\beta) \propto f(\beta_n\ \forall n|\mu_\beta, V_\beta)IW(V_\beta|v_0, \Gamma_0)$$

から，

V_βの事後分布$IW(v_N, \Gamma_N)$：インバース・ウィシャート分布

$$v_N=v_0+N$$

$$\Gamma_N=\Gamma_0+S_\beta \quad \text{where} \quad S_\beta=\sum_{n=1}^{N}(\beta_n-\mu_\beta)(\beta_n-\mu_\beta)'$$

［c］　$\{\mu_\beta, V_\beta\} \Rightarrow \beta_n\ \forall n$

β_nの事前分布$N(\mu_\beta, V_\beta)$：正規分布

尤度関数$f_n(y_{nt}\ \forall t|\beta_n)$：$\beta_n$のロジット・モデルから$y_{nt}(t=1, \cdots, T_n)$が観測される確率

$$f(\beta_n|y_{nt}\ \forall t) \propto f_n(y_{nt}\ \forall t|\beta_n)N(\beta_n|\mu_\beta, V_\beta) \quad (16)$$

から，事後分布が求められる．

残念ながら正規分布は尤度関数に対して共役でないため，β_nの事後分布$f(\beta_n|y_{nt}\ \forall t)$は[a]と[b]のように簡単には計算できない．このような場合にβ_nを事後分布から数値的に発生させる方法として，メトロポリス・ヘースティング（MH）アルゴリズムという物理学で発案された手法が使われる．この詳細は節5.3で紹介する．

MCMCシミュレーション・プロセス

[a]～[c]の事後分布に基づいて$\{\beta_n\ \forall n, \mu_\beta, V_\beta\}$をランダムに発生させ，反復を通常，1万回ぐらい繰り返すことによって得られた乱数は，パラメータの同時密度に収束する．推定されたパラメータの不確実性は，収束したと考えられる後半50%の反復で発生された乱数値を分析することによってシミュレーション的に求められる．

5.3 メトロポリス・ヘースティング（MH）アルゴリズムとギブス・サンプリング

上記の[c]では事後分布が容易に計算できなかったが，もし全てのパラメータの事後分布を容易に計算できる場合は，その分布に基づいて乱数を逐次的に発生させるステップを十分な回数だけ反復すれば同時密度に収束する．これはギブス・サンプリングと呼ばれ，MCMCの中でも一番単純な方法である．MHアルゴリズムは，事後分布からランダムサンプルを直接発生させることが困難な場合に使われる．[c]の事後分布はロジット尤度関数と正規事前分布の積で規定された式(16)であり，解析的には求められない．

MHアルゴリズムの概念は簡単である．表記を単純にするためにサブスクリプトnを省き，前回の反復で発生されたβ_nの値をb_{old}，今回の試験的なβ_nの値をb_{new}としよう．もし，b_{new}がb_{old}に比較してデータへのフィットを向上させるならば，それを今回の

値として採用するが，逆にフィットを悪化させるのであれば，悪化の度合いに応じて確率的に棄却し前回の値を保持する．

試験的な値を発生させるには，$b_{new}=b_{old}+d$ のように前回の値 b_{old} にランダムなジャンプ・ベクトル d を加える．d は b_{old} の事前分布の不確実性に基づいて，確率的に $d \sim N(0, kV_\beta)$ から発生させる．k はジャンプの距離を調整するパラメータで，採用と棄却の割合に影響を与え，アルゴリズムの収束の速さをコントロールする．

b_{new} と b_{old} のデータへのフィットは，式(17)のような事後確率の比 r で評価され，b_{new} は $min(r,1)$ の確率で採用される．

$$r=\frac{f_n(y_n|b_{new})N(b_{new}|\mu_\beta, V_\beta)}{f_n(y_n|b_{old})N(b_{old}|\mu_\beta, V_\beta)} \tag{17}$$

つまり，b_{new} の方が b_{old} より事後確率が高ければ b_{new} を採用し，低ければ r に応じた確率で採用するのである．ジャンプ・ベクトル d が大き過ぎるとこの採用確率が下がり，逆に小さ過ぎると解の変化が小さいため，収束を速めるには k を適切な値に調整することが必要である．MH アルゴリズムの収束は採用率が約3割の時に一番速いことが経験的に知られているため（文献[10]，p. 335），採用率をモニターしながら k をリアルタイムに微調整することが望ましい．

ここでは階層ベイズをロジット・モデルに応用したが，計量経済学での状況と反して実はプロビット・モデルの推定の方がもっと簡単である．この場合，切断された正規分布の形をした効用 V_{nti} を新たにパラメータとして導入する data augumentation という手法[11]によって全てのパラメータが共役事前分布になるため，MH アルゴリズムの代わりに，計算が簡単で効率のよいギブス・サンプリングが使えるからである．潜在ロジット・モデルと同様な，消費者固有のパラメータ β_n をデモグラフィック変数に関連付ける拡張も，ベイズ階層を一つ増やすだけでモデル化が可能である[4].

6. まとめ

ブランド選択モデルにおける消費者の異質性のモデルは，過去20年の進歩をとげて現在一段落した．数万回のシミュレーションが必要なために計算時間はかかるが，個人レベルでのパラメータの推定も可能となった．計算量の問題は，コンピュータのめざましい発達が解決してくれるであろう．

近年の情報技術の発達で，大量の個人レベルのデータが自動的に蓄積される環境が急に整った．日本に POS データが一般化したのは80年代の中旬であり，FSP に代表される顧客 ID 付き購買データが収集されるようになったのも過去4，5年の出来事である．インターネットで E コマースが一般的になったのも，この2，3年だ．

しかし，このように収集された個人レベルのデータから有用な知見や知識を得なければ，これらは保存に厄介な単なるゴミであり情報にはなりえない．現在多くの企業は，この大量のデータからいかに有用な情報を抽出して，それをマーケティングに利用するかに行き詰まっている．購買金額に基づいた単純な一律還元のポイントシステム，これは全ての競合スーパー，量販店，航空会社が同等な報償を提供しているため単なる値引き合戦による過当競争を生み出している．そして，企業は「FSP を作ったのに利益があがらない」と首をかしげ，消費者は困惑し似たような競合企業のロイヤルティー・カードを数多く持ち，もはやロイヤルティーの役目をなしていない．

問題はハードの進歩にソフトの進歩がついていないことである．ハードで競合企業に追いつく，あるいは競争企業にコピーされるのは簡単である．同じ情報システム・ベンダーのシステムを取り入れればよいだけである．ハードのみに頼っていては，最新のハードを導入する東南アジアの競争企業にもすぐに追いつかれ，逆に低賃金の優位性によって追い越されてしまう．企業としての本当の競争優位はソフトで決まるといっても過言でない．製造業でさえ，これからは情報で差が生まれるのである．アマゾン・ドット・コム，TSUTAYA，デル・コンピュータ，アスクルなどは皆，良い例である．

個人レベルのデータを平均値や分散に集約してから分析をしてしまっては，非集計データのメリットを十分に生かしているとは言えない．個人の異質性を考慮したミクロの観点からの分析は，マーケティングが今後，消費者により有用なメリットをもたらし，企業，社会全体の発展を促すことを助けるであろう．その第一歩として，次回の掲載では探索的非集計データ分析，通称，データ・マイニングと呼ばれている手法を紹介する．

参考文献

[1] Guadagni, P. M. and J. D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science*, 2 (3), 203-238.

[2] Kamakura, W. A. and G. J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure", *Journal of Marketing Research*, 26 (4), 379-390.

[3] Manski, C. F. and D. McFadden (1981), *Structural Analysis of Discrete Data with Econometric Appilcation*, Cambridge, MA: MIT Press.

[4] Rossi, P., R. E. McCulloch and G. M. Allenby (1996), "The Value of Purchase History Data in Target Marketing", *Marketing Science*, 15 (4), 301-320.

[5] Gupta, S. and P. Chintagunta (1994), "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models", *Journal of Marketing Research*, 31 (1), 128-136.

[6] Mittal, B. (1994), "An Integrated Framework for Relating Diverse Consumer Characteristics to Supermarket Coupon Redemption", *Journal of Marketing Research*, 31 (4), 533-544.

[7] Fader, P. S. and J. M. Lattin (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior", *Marketing Science*, 12 (3), 304-317.

[8] Dempster, A. P., N. M. Laird and D. B. Rubin (1977), "Maximum Likelihood from Incomplete Data via the EM-Algorithm", *Journal of the Royal Statistical Society*, Series B 39, 1-38.

[9] Rossi, P. and G. M. Allenby (1993), "A Bayesian Approach to Estimating Household Parameters", *Journal of Marketing Research*, 30 (2), 171-182.

[10] Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin (1995), *Bayesian Data Analysis*, Boca Raton, Florida: Chapman & Hall.

[11] Albert, J. H. and S. Chib (1993), "Bayesian analysis of Binary and Polychotomous Response Data", *Journal of the American Statistical Association*, 88, 669-679.