

潜在セグメントモデルへの EMアルゴリズムの適用可能性について

栗山 浩一（早稲田大学）^{*1}
三谷 羊平（コロラド大学）
庄子 康（北海道大学）

2008年8月

^{*1} 〒169-8050 東京都新宿区西早稲田 1-6-1 早稲田大学政治経済学術院
TEL/FAX 03-5286-9822
kkuri@waseda.jp

潜在セグメントモデルへのEMアルゴリズムの適用可能性について

要約

環境の経済評価では、環境に対する選好の多様性をモデル化することが重要な課題の一つとなっているが、そのためのアプローチとして回答者を異なる集団（セグメント）に分類し、セグメントごとに選好パラメータを推定する潜在セグメントモデルが注目されている。ただし、潜在セグメントモデルでは、推定されるパラメータの数が多く、通常の方法では推定に失敗することが多いため、EMアルゴリズムの適用を検討した。理論モデルでは、所属確率を直接推定するだけでなく、メンバーシップ関数を用いたモデルに対してもEMアルゴリズムを適用可能であることを示した。さらにモンテカルロ・シミュレーションを用いてEMアルゴリズムの性能を確認した。その結果、所属確率の高いセグメントについては、効用関数のパラメータを正しく再現できることが示されたものの、メンバーシップ関数については再現性が低いことが明らかとなった。

キーワード：環境評価，選択型実験，離散選択モデル，モンテカルロ・シミュレーション

1. はじめに

近年、環境の経済的評価に関する研究が進められているが、その中で回答者の選好の多様性をモデル化することが重要な課題の一つとなっている^{*2}。トラベルコスト法や選択型実験では離散選択型モデルが用いられているが、標準的な離散選択モデルでは代表的な個人を想定して効用関数を推定している。しかし、環境経済学の分野では、Train (1998) や Boxall and Adamowicz (2002) が実証研究によって選好の多様性が存在することを指摘しており、すべての回答者の効用関数を同一と仮定することは現実的ではない。そこで、環境に対する選好の多様性をモデル化する方法の一つとして、潜在セグメントモデル (Latent Segmentation Model) が有望視されている^{*3}。潜在セグメントモデルは、回答者が選好の異なる複数のグループ (セグメント) によって構成されると想定し、それぞれのセグメント別に選好パラメータを推定する。ただし、潜在セグメントモデルは推定されるパラメータの数が多いため、通常の方法では推定できないことが多い。そこで本研究は、潜在セグメントモデルに EM アルゴリズム (Expectation-Maximization algorithm) を適用し、モンテカルロシミュレーションにより推定値の再現性を検討する。EM アルゴリズムは、全てのパラメータを同時に推定する代わりに、推定するパラメータをいくつかのサブセットに分けて、推定プロセスを分解し、推定を反復することで推定値を得る方法である^{*4}。

環境評価において、回答者の異質性を分析するモデルとしては、これまでにランダムパラメータロジット (Random Parameter Logit)、階層化ベイズ法 (Hierarchical Bayes)、そして潜在セグメントモデルが用いられている^{*5}。ランダムパラメータロジットは、効用関数の選好パラメータが個人間である分布を持っていると仮定し、選好パラメータの分布関数を推定することで選好の多様性をモデル化する。ランダムパラメータロジットは、従来の離散選択モデルのデータだけで推定可能であり、推定も安定しているという利点を持っているが、選好の多様性を示すだけであるため、評価結果を環境政策に適用するときに、政策の影響がどのように消費者に影響するかを詳しく見るのが難しい。一方、階層化ベイズ法は、個人別に選好パラメータを推定することができるものの、ベイズ推定を用いていることから従来の最尤推定を用いる離散選択モデルとは、背景となる考え方が大きく異なるという特徴がある。

これに対して、潜在セグメントモデルは、回答者が異なる選好を持つ複数の集団 (セグメント) によって構成されると想定し、それぞれのセグメント別に選好パラメータを推定することで選好の多様性をモデル化している。潜在セグメントモデルは、セグメント別に選好をモデル化できるので、環境政策の影響が各セグメント別に分析することが可能である。しかも、通常の方法を用いるので、標準的な条件付きロジットモデルと比較することも可能である。ただし、セグメント別に選好パラメータを推定するため、セグメント数が増加すると推定すべきパラメータが急増し、最

*2 環境の経済的評価の研究動向については栗山・庄子 (2005) が詳しい。

*3 潜在セグメントモデルは、しばしば潜在クラスモデル (Latent Class Model) と呼ばれる。環境評価に潜在セグメントモデルを用いた初期の研究には Provencher et. al. (2002), Boxall and Adamowicz (2002), Shonkwiler and Shaw (2003) が含まれる。

*4 EM アルゴリズムについては、もともとは欠損値を分析するために開発された (Dempster et al., 1977)。潜在セグメントモデルへの適用については Train (forthcoming) が詳しい。

*5 ランダムパラメータロジットや階層化ベイズ法については Train (2003) が詳しい。

尤法による推定が困難になることが多い。

最近、潜在セグメントモデルの推定困難性の問題に対して EM アルゴリズムを用いるアプローチが提唱されている (Train, forthcoming)。EM アルゴリズムは、すべてのパラメータを同時に推定する代わりに、各セグメントのパラメータを別々に推定するプロセスを反復することで最尤推定値を得る方法である。尤度関数が複雑な場合や推定するパラメータの数が多く、通常の最尤法では推定に失敗する場合でも、EM アルゴリズムを用いることで推定が可能となることが多い。ただし、EM アルゴリズムは局所的最適値に収束する可能性があることが知られており、EM アルゴリズムの推定値が必ずしも最尤推定値と一致するとは限らない。

そこで、本研究では、潜在セグメントモデルに EM アルゴリズムを適用したときの推定可能性をモンテカルロシミュレーションによって分析する。すなわち、選好パラメータとしてある想定値を設定し、この想定値に基づいて構築された仮想的な回答データに対して EM アルゴリズムを用いて推定した上で、推定結果が元の想定値を正しく再現できるか否かを見ることで、EM アルゴリズムの性能を確認する。また、先行研究では、セグメントの所属確率を直接推定する方法が用いられているが、本研究では、回答者がどのセグメントに所属するかを示すメンバーシップ関数を導入したときに EM アルゴリズムを適用する方法についても検討を行った。

本論文の構成は以下の通りである。第 2 節では、潜在セグメントモデルおよび EM アルゴリズムのモデルについて示す。第 3 節では、モンテカルロシミュレーションの分析結果を示し、最後に結論と今後の研究課題について示す。

2 . モデル

2 . 1 潜在セグメントモデル

まず潜在セグメントモデルについて示す。回答者 n は選択集合 $C = \{1, \dots, J\}$ の代替案の中から最も好ましいものを選ぶ作業を T_n 回行うとする。回答者は S 種類のセグメントのどれかに属とする。セグメント $s \in \{1, \dots, S\}$ に属する回答者 n が t 回目の選択時に代替案 j を選んだときの効用を U_{ntj}^s とする。ここでランダム効用モデルを想定し、 $U_{ntj}^s = \beta_s x_{ntj} + \varepsilon_{ntj}$ とする。ただし、 β_s はセグメント s の選好パラメータ、 x_{ntj} は代替案の属性ベクトル、 ε_{ntj} は誤差項であり独立の第一種極値分布に従うと仮定する。このとき、回答者が代替案 j を選択する確率は

$$L_{nt}(j|\beta_s) = \frac{\exp(\beta_s x_{ntj})}{\sum_{k \in C} \exp(\beta_s x_{ntk})} \quad (1)$$

となる。 T_n 回の選択作業において回答者が選んだ代替案の履歴データベクトルを $y_n = \{y_{n1}, \dots, y_{nT_n}\}$ とする。このとき、セグメント s に属する回答者が y_n の選択を行う確率は

$$K_n(y_n|\beta_s) = \prod_t L_{nt}(y_{nt}|\beta_s) \quad (2)$$

となる。

回答者がどのセグメントに属するかを調査者は直接観測できない。そこで、回答者 n が s に所属する確率を π_{ns} とすると、回答者が y_n の選択を行う確率は

$$P_n(\theta) = \sum_s \pi_{ns} K_n(y_n | \beta_s), \quad \sum_s \pi_{ns} = 1 \quad (3)$$

となる。ただし、 θ は推定されるパラメータのベクトルである。セグメント確率 π_{ns} については、すべての回答者でセグメントに所属する確率が等しい ($\pi_{ns} = \pi_s, \forall n$) と仮定して π_s を直接推定する方法と、回答者がどのセグメントに属するかをメンバーシップ関数によって推定する方法がある。メンバーシップ関数は $M_{ns} = \lambda_s z_n + \zeta_{ns}$ として表現できると仮定する。ただし、 z_n は個人属性、 ζ_{ns} は誤差項、 λ_s はパラメータである。そして回答者は M_{ns} が最大のセグメントに属すると想定する。ここで誤差項が独立の第一種極値分布に従うと仮定すると、多項ロジットモデルと同様に、回答者 n が s に所属する確率 π_{ns} は次式によって示すことができる。

$$\pi_{ns} = \frac{\exp(\lambda_s z_n)}{\sum_s \exp(\lambda_s z_n)} \quad (4)$$

このとき、回答者の選択確率は以下のとおりとなる。

$$P_n(\theta) = \sum_s \frac{\exp \lambda_s z_n}{\sum_s \exp \lambda_s z_n} K_n(y_n | \beta_s) \quad (5)$$

これを用いて、最尤法によりパラメータの推定が行われる。

2.2 EM アルゴリズム

前述のように、潜在セグメントモデルでは、推定するパラメータの数が多いため通常最尤法ですべてのパラメータを同時に推定しようとする、推定に失敗して推定値が得られないことが多い。EM アルゴリズムは、各セグメントのパラメータを別々に推定し、パラメータを更新するプロセスを反復することで最尤推定値を得る方法である。ここでは Train (forthcoming) の記述に従い、EM アルゴリズムの推定方法を示す。なお、Train (forthcoming) は回答者のセグメント所属確率を直接推定する方法のみ用いているが、本研究ではメンバーシップ関数を用いる場合についても考慮を行っている。

まず、回答者 n がどのセグメントに属するかを知っているとしよう。回答者 n がセグメント s に属するときに 1 となるダミー変数を d_{ns} とする。このとき尤度関数および対数尤度関数は以下のとおりとなる。

$$L = \prod_n \prod_s (\pi_{ns} K_n(y_n | \beta_s))^{d_{ns}}$$

$$\ln L = \sum_n \sum_s d_{ns} \ln K_n + \sum_n \sum_s d_{ns} \ln \pi_{ns} \quad (6)$$

調査者は、回答者がどのセグメントに属するのかを観測できないため、対数尤度関数の期待値を用いて推定を行う。

$$E[\ln L] = \sum_n \sum_s h_{ns} \ln K_n + \sum_n \sum_s h_{ns} \ln \pi_{ns} \quad (7)$$

ただし、 h_{ns} は回答者の選択データ y_n のもとで回答者がセグメント s に属する条件付き確率であり、ベイズの定理を用いると次式によって示すことができる。

$$h_{ns} = \frac{\pi_{ns} K_n(y_n | \beta_s)}{P_n(\theta)} \quad (8)$$

EM アルゴリズムは、回答者のセグメント所属確率 h_{ns} を所与として対数尤度 (7) を最大化し、推定されたパラメータをもとに式 (8) を用いて h_{ns} を更新するプロセスを反復することで推定値を得る。

期待対数尤度関数 (7) を見ると、 h_{ns} を所与とした場合、 $\ln K_n$ のパラメータ (β_s) と $\ln \pi_{ns}$ のパラメータ (π_s または λ_s) は独立なので、第 1 項と第 2 項それぞれ最大化することで独立に推定することができることがわかる。

まず、期待対数尤度関数の第 2 項の最大化について見てみよう。 i 回目の反復プロセスにおいて h_{ns} の値を h_{ns}^i とし、このもとで第 2 項を最大化する。最初にすべての回答者でセグメントに所属する確率が等しい ($\pi_{ns} = \pi_s, \forall n$) と仮定して π_s を直接推定する場合を考える。このとき、 π_s の合計が 1 という条件付きで第 2 項を最大化することで π_s を推定する。

$$\pi_s = \operatorname{argmax}_{\pi_s} \sum_n \sum_s h_{ns}^i \ln \pi_s + \mu \left[1 - \sum_s \pi_s \right] \quad (9)$$

ただし、 μ はラグランジュ乗数である。これを解くと π_s は次式によって更新される。

$$\pi_s^{i+1} = \frac{1}{N} \sum_n h_{ns}^i \quad (10)$$

一方、メンバーシップ関数を用いて回答者のセグメント所属確率を推定する場合を考えよう。この場合、メンバーシップ関数のパラメータは次式によって得られる。

$$\begin{aligned} \lambda_s &= \operatorname{argmax}_{\lambda_s} \sum_n \sum_s h_{ns}^i \ln \left(\frac{\exp(\lambda_s z_n)}{\sum_s \exp(\lambda_s z_n)} \right) \\ &= \operatorname{argmax}_{\lambda_s} \sum_n h_{ns}^i \ln \left(\frac{\exp(\lambda_s z_n)}{\sum_s \exp(\lambda_s z_n)} \right) \end{aligned} \quad (11)$$

ここで、パラメータ λ_s は各セグメントで独立なので、各セグメントを別々に分解して推定することができることに注目されたい。式 (11) を見ると、メンバーシップ関数の推定は、通常の多項口

ジットにウェイト h_{ns}^i を用いて推定したものと等しいことがわかる。パラメータ λ_s が推定されると、式 (4) を用いて所属確率を計算できる。

次に期待対数尤度関数 (7) の第 1 項の最大化を考えよう。 h_{ns}^i を所与とし、第 1 項を最大化することで第 1 項のパラメータ β_s を更新する。すなわち、

$$\begin{aligned}\beta_s^{i+1} &= \operatorname{argmax}_{\beta_s} \sum_n \sum_s h_{ns}^i \ln K_n(\beta_s) = \operatorname{argmax}_{\beta_s} \sum_s h_{ns}^i \ln K_n(\beta_s) \\ &= \operatorname{argmax}_{\beta_s} \sum_n \sum_t h_{ns}^i \ln L_{nt}(\beta_s)\end{aligned}\tag{12}$$

となる。ここでも、パラメータ β_s は各セグメントで独立なので、各セグメントを別々に分解して推定することができることを用いている。

以上の手順を整理すると、EM アルゴリズムの推定ステップは以下のとおりとなる。

1. 効用関数のパラメータ β_s^0 および選択確率のパラメータ (π_s^0 または λ_s^0) の初期値を設定する。ここでは、Train (forthcoming) の方法に従い、サンプルを S 個に分割し、それぞれで条件付きロジットで推定した結果を β_s^0 の初期値とした。選択確率のパラメータに関しては、選択確率を直接推定する場合は $\pi_s^0 = 1/S$ とし、メンバーシップ関数の場合は λ_s^0 をランダムに設定した。
2. 式 (8) を用いてウェイト h_{ns}^i を更新する。
3. 所属確率 π_{ns}^{i+1} を更新する。所属確率を直接推定する場合は式 (10) を用いる。メンバーシップ関数を用いる場合は、式 (11) を用いてパラメータ λ_s^{i+1} を推定し、式 (4) を用いて所属確率を更新する。
4. 効用関数のパラメータを式 (12) を用いて更新する。具体的には、 h_{ns}^i のウェイト付きで条件付きロジットを S 回実施し、各セグメントのパラメータ β_s を推定する。
5. ステップ (2) から (4) を収束するまで繰り返す。

なお、推定パラメータの標準誤差は、EM アルゴリズムによって得られた推定値を用いて対数尤度関数のヘッセ行列を計算して標準誤差を計算する方法 (Ruud, 1991) と、ブートストラップによって EM アルゴリズムの推定を繰り返すことで標準誤差を計算する方法 (Train, forthcoming) が提唱されているが、本研究ではブートストラップを用いる方法を採用した。

3 . モンテカルロ・シミュレーションによる EM アルゴリズムの検証

EM アルゴリズムは、潜在セグメントモデルのパラメータを同時に推定する代わりに、各セグメントごとに推定プロセスを分解して推定する作業を反復することで推定値を計算するため、パラメータの数が多い場合でも局所的最適値を得ることができることが知られている。ただし、EM アルゴリズムは局所的最適値に収束する可能性があることが知られているため、必ずしも最尤推定値が得られるとは限らない。そこで、モンテカルロ・シミュレーションを用いて EM アルゴリズムの性能を検証する。

まず, 100 人の回答者が存在するとする。各回答者は 8 ~ 10 回の中からランダムに選択回数 (T_n) が与えられる。代替案の数 (J) は 4 とし, それぞれの代替案は 5 種類の属性によって構成されている。各代替案の属性 (x_{mj}) は 1 ~ 10 のどれかの整数値をランダムに設定する。ただし, 第 1 属性は価格属性とし, 効用パラメータはマイナスとする。メンバーシップ関数は 4 種類の個人属性によって構成され, 個人属性 (z_n) も 1 ~ 10 のどれかの整数値をランダムに設定する。効用関数のパラメータ (β_s) およびメンバーシップ関数のパラメータ (λ_s) は表 1 の値を設定した。この想定パラメータのもとで, 第 1 種極値分布に従う誤差項 (ε_{ntj}) および (ζ_{ns}) をランダムに発生させて, 各セグメントの効用関数 (U_{ntj}^s) とメンバーシップ関数 (M_{ns}) を計算し, メンバーシップ関数が最大となるセグメントに回答者が所属し, そのセグメントの効用関数が最大となる代替案を選択するとして, 仮想的な選択データ (y_n) を作成した。この想定値における各セグメントの所属確率は表 1 のとおりであった。そして, この仮想的な選択データに対して, EM アルゴリズムを用いて推定した結果と, パラメータの想定値を比較することで, EM アルゴリズムの性能を分析した。

表 1 シミュレーションで用いた想定値

潜在セグメントモデルでは, 調査者は真のセグメント数を知らないため, セグメント数を事前に決める必要がある。セグメント数の決定には一般に AIC (Akaike Information Criterion) と BIC (Bayesian Information Criterion) が用いられる。それぞれの定義は以下のとおりである。

$$\begin{aligned} AIC &= -2LL + 2K \\ BIC &= -LL + \ln(N) \cdot K/2 \end{aligned} \tag{13}$$

ただし, LL は対数尤度, K はパラメータ数, N は回答者数である。

ここでは 2 種類のモデルを分析した。モデル 1 はメンバーシップ関数を用いずに所属確率を直接推定する方法を用いたものであり, モデル 2 はメンバーシップ関数を用いる方法を用いたものである。表 2 はセグメント数を 2 ~ 8 まで変化させたときに各モデルで AIC と BIC がどのように変化するかを示したものである。これによると, どちらのモデルの場合も, AIC と BIC はセグメント数が 4 のときに最低値となり, セグメント数が 4 つと判断されている。実際にはこのデータはセグメント数を 5 つで仮想データを作成しているが, 表 1 で示されているように s_5 の所属確率は 0.000 であることから, AIC と BIC はセグメント数を 4 と判定したと考えられる。

表 2 セグメント数の比較

表 3 はモデル 1 (所属確率を直接推定) の推定結果を示したものである。括弧内の数値は 100 回のブートストラップによって推定したパラメータの標準誤差である。セグメント数は 5 つのため推定パラメータは多いものの, EM アルゴリズムを用いると推定に失敗することはなく, 短時間で収束することができた。まず, 所属確率の推定値を見ると, s_5 以外ではすべて有意となっており, しかも表 1 の想定値と比較すると, すべてのセグメントで所属確率を正しく再現していることが分か

る。また、推定された効用関数のパラメータを見ると、s5のx5属性を除くと全てのパラメータが有意となっており、想定値と比較すると所属確率の高いs1からs3までは想定値と推定値に有意差は見られない。所属確率が低いs4とs5では想定値と推定値の差は有意となるものが見られた。つまり、仮想的データではメンバーシップ関数を用いて選択データを作成しているにもかかわらず、メンバーシップ関数を用いないモデル1であっても所属確率の高いセグメントに対しては効用関数のパラメータを正確に再現できることが分かった。

表3 モデル1（所属確率を直接推定）の分析結果

表4はモデル2（メンバーシップ関数モデル）の推定結果を示したものである。モデル2ではメンバーシップ関数も推定する必要があるためモデル1よりも推定するパラメータの数が多いが、EMアルゴリズムを用いると推定に失敗することなく、収束するまでには比較的時間を要したものの、推定結果を得ることに成功した。まず、効用関数の推定値を見ると、s4とs5の一部の属性を除くと有意となっている。そして推定値と表1の想定値を比較すると、所属確率の高いs1とs2では有意差は見られないが、s3～s5では有意差が見られた。つまり効用関数に関しては、モデル1とモデル2はどちらも所属確率が高いセグメントは正しく再現できるが、所属確率が低いセグメントに関しては再現性が低いことが分かる。一方のメンバーシップ関数について見てみると、所属確率の高いs2であっても有意ではないものが多く、所属確率の低いs5ではすべての変数が非有意となっている。また推定値と想定値を比較すると、所属確率の高いs1であっても推定値と想定値の差は有意のものがあり、メンバーシップ関数に関しては再現性が低いことが分かる。

表4 モデル2（メンバーシップ関数モデル）の分析結果

表5は二つのモデルでそれぞれ限界支払意思額を推定した結果を示している。限界支払意思額は、各属性の効用パラメータを価格属性のパラメータで割ったものの絶対値に相当する。限界支払意思額についても、所属確率の高いセグメントはすべての属性で有意だが、所属確率の低いセグメントでは有意ではないものが見られる。想定値と推定結果を比較すると、効用関数のパラメータと同様に、どちらのセグメントも所属確率の高いセグメントは想定値を正しく再現できているが、所属確率の低いセグメントでは有意差が見られた。

表5 限界支払意思額の分析結果

最後に、パラメータの想定値を変化させた場合についても検討した。これまでの分析結果は、あくまでも表1で示された想定値での結果であり、想定値を変化させると結果も変わる可能性がある。そこで、以下の手順でパラメータの想定値を変化させたときの影響を分析した。

1. パラメータの想定値を乱数で生成する。効用関数については、価格パラメータは-1から

-6, その他のパラメータは -2 から 7 の範囲の一様分布でランダムに設定する。メンバーシップ関数は -3 から 5 の範囲の一様分布でランダムに設定する。

2. 想定値をもとに 100 人の回答者の仮想データを作成する。データ作成方法はこれまでの分析と同じである。
3. 仮想データにおいて所属確率が 0.9 を上回るセグメントが存在するか、あるいは所属確率が 0 となるセグメントが存在する場合はステップ (1) に戻る。それ以外の場合は次のステップへ進む。これは所属確率の極端な偏りを排除するための処理である。
4. 仮想データに対して EM アルゴリズムによる推定を行う。試行回数 20 回のブートストラップにより推定パラメータの標準誤差や P 値を求める^{*6}。
5. ステップ (1) から (4) のプロセスを 20 回繰り返す。最後に得られた推定結果のうち有意なパラメータの比率を調べる。

表 6 は以上のプロセスで得られた結果を示している^{*7}。効用関数のパラメータは、どちらのモデルでも所属確率が上昇するとともに有意なパラメータの比率も上昇している。また、所属確率が 20 % 未満の場合でも約半分は有意な結果となっている。これに対して、メンバーシップ関数については、所属確率が 40 % 以上と高い場合でも有意なものは 36.8 % にとどまっている。このことから、表 1 の想定値で得られた結論、すなわちメンバーシップ関数の再現性が低いという結論は、想定値をランダムに変化させた場合でも支持されることが分かった。

表 6 想定値をランダムに生成したときの有意な変数の比率

4 . 結論

本研究では、潜在セグメントモデルへの EM アルゴリズムの適用可能性について検討した。これまで、潜在セグメントモデルに EM アルゴリズムを適用した研究では、所属確率を直接推定する方法が用いられてきたが、本研究ではメンバーシップ関数を用いたモデルについても EM アルゴリズムを適用することが可能であることを理論的に明らかにした。また、シミュレーション分析の結果より、セグメント数が多くて通常の方法では推定が困難な場合であっても、EM アルゴリズムを適用することで推定値を得ることが示された。また、メンバーシップ関数を用いないモデル 1 であっても、所属確率が高いセグメントに関しては効用関数を正しく再現できることが分かった。一方で、メンバーシップ関数については、有意とはならないものが多く、EM アルゴリズムを用いてもメンバーシップ関数を正しく再現することは難しいことが示された。

メンバーシップ関数を正しく再現できなかった原因としては、情報の不足が考えられる。回答者がどの代替案を選んだかという情報だけで、回答者の効用関数とメンバーシップ関数の両者を推定することに限界がある可能性がある。この場合、メンバーシップ関数を推定するための追加情報を

^{*6} Train はブートストラップの試行回数として 20 回を採用している。

^{*7} Aptech 社の統計ソフトウェア GAUSS による計算では、この結果を得るために約 10 時間を必要とした。

用いることが考えられる。たとえば、「あなたは以下のグループのどれに相当すると思いますか」というような設問を追加することで、メンバーシップ関数を別途推定することが可能となるだろう。

また、本研究ではセグメント数を固定して分析を行ったが、セグメント数を変化させたときに、EM アルゴリズムの性能を検証することも必要である。今後は、こうした方法についてシミュレーションあるいは実証データを用いた研究が必要であろう。

参考文献

- [1] Boxall, P.C. and W.L. Adamowicz. (2002) 'Understanding heterogeneous preferences in random utility models: A latent class approach' *Environmental and Resource Economics* 23 421–446.
- [2] Dempster, A., N. Laird and D. Rubin (1977) 'Maximum likelihood from incomplete data via the EM algorithm' *Journal of the Royal Statistical Society B* 39, 1–38.
- [3] 栗山浩一・庄子康編著 (2005) 「環境と観光の経済評価国立公園の維持と管理」、勁草書房
- [4] Provencher, B., K.A. Baerenklau, and R.C. Bishop. (2002) 'A Finite Mixture Logit Model of Recreational Angling with Serially Correlated Random Utility' *American Journal of Agricultural Economics* 84, 1066–1075.
- [5] Ruud, P. (1991) 'Extensions of estimation methods using the em algorithm' *Journal of Econometrics* 49, 305–341.
- [6] Shonkwiler, J.S. and W.D. Shaw. (2003) 'A Finite Mixture Approach to Analyzing Income Effects in Random Utility Models' in *The New Economics of Outdoor Recreation*, N. D. Hanley, ed. Edward Elgar, 268–278.
- [7] Train, K. (1998) 'Recreation demand models with taste variation' *Land Economics* 74, 230–239.
- [8] Train, K. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press, New York.
- [9] Train, K. (forthcoming) 'EM Algorithms for Nonparametric Estimation of Mixing Distributions', *Journal of Choice Modeling*.

表1 シミュレーションで用いた想定値

効用関数パラメータ

	セグメント				
	s1	s2	s3	s4	s5
x1	-5	-6	-3	-2	-4
x2	1	1	-1	2	3
x3	2	3	3	-1	1
x4	3	-1	2	2	-1
x5	-1	3	-1	-1	1

メンバーシップ関数パラメータ

	セグメント				
	s1	s2	s3	s4	s5
z1	3	-1	-2	0	0
z2	-1	3	2	0	2
z3	-1	2	1	0	1
z4	-1	-4	-1	0	-3

仮想データにおける想定値 (各セグメントのシェア)

	s1	s2	s3	s4	s5
所属確率	0.372	0.344	0.170	0.114	0.000

表2 セグメント数の比較

セグメント数	モデル1		モデル2	
	AIC	BIC	AIC	BIC
2	866.8	447.7	792.8	417.3
3	638.6	341.4	535.2	298.9
4	464.6	262.3	231.0	157.2
5	476.6	276.1	236.8	170.5
6	486.0	288.6	259.7	192.4
7	498.0	302.4	266.5	206.2
8	509.9	316.2	281.8	224.3

表3 モデル1（所属確率を直接推定）の分析結果

効用関数

属性	セグメント									
	s1		s2		s3		s4		s5	
x1	-4.756	***	-5.457	***	-2.771	***	-1.451	***	-0.541	***
	(1.42)		(0.85)		(0.58)		(0.42)		(0.04)	
x2	1.154	***	1.011	***	-1.008	***	1.344	***	0.111	***
	(0.34)		(0.21)		(0.25)		(0.28)		(0.03)	
x3	2.031	***	2.874	***	2.613	***	-0.707	***	0.265	***
	(0.57)		(0.52)		(0.56)		(0.18)		(0.04)	
x4	2.951	***	-0.926	***	1.838	***	1.411	***	0.210	***
	(0.88)		(0.19)		(0.41)		(0.35)		(0.03)	
x5	-0.836	***	2.812	***	-0.980	***	-0.666	***	0.019	
	(0.26)		(0.43)		(0.22)		(0.22)		(0.03)	
所属確率	0.345	***	0.305	***	0.218	***	0.131	***	0.000	
	(0.05)		(0.05)		(0.04)		(0.04)		(0.00)	

想定値との差

属性	セグメント									
	s1		s2		s3		s4		s5	
x1	0.244		0.543		0.229		0.549		3.459	***
x2	0.154		0.011		-0.008		-0.656	**	-2.889	***
x3	0.031		-0.126		-0.387		0.293		-0.735	***
x4	-0.049		0.074		-0.162		-0.589	*	1.210	***
x5	0.164		-0.188		0.020		0.334		-0.981	***
所属確率	-0.027		-0.039		0.048		0.017		0.000	

括弧内の数値は100回のブートストラップによって推定した標準誤差。

***は1%水準、**は5%水準、*は10%水準で有意を意味する。

表4 モデル2（メンバーシップ関数モデル）の分析結果

効用関数

属性	セグメント									
	s1		s2		s3		s4		s5	
x1	-4.805	***	-5.505	***	-1.439	***	-3.787	***	-2.153	**
	(1.73)		(0.89)		(0.41)		(0.89)		(1.14)	
x2	1.148	***	1.024	***	1.343	***	-1.451		-0.659	
	(0.41)		(0.22)		(0.28)		(0.81)		(0.63)	
x3	2.054	***	2.907	***	-0.710	***	3.522	***	2.053	**
	(0.69)		(0.54)		(0.18)		(0.98)		(0.81)	
x4	2.974	***	-0.934	***	1.402	***	2.378	***	1.502	**
	(1.06)		(0.19)		(0.35)		(0.67)		(0.75)	
x5	-0.838	***	2.841	***	-0.668	***	-1.051	***	-1.088	*
	(0.31)		(0.45)		(0.21)		(0.40)		(0.49)	

想定値との差

属性	セグメント									
	s1		s2		s3		s4		s5	
x1	0.195		0.495		1.561	***	-1.787	**	1.847	
x2	0.148		0.024		2.343	***	-3.451	***	-3.659	***
x3	0.054		-0.093		-3.710	***	4.522	***	1.053	
x4	-0.026		0.066		-0.598	*	0.378		2.502	***
x5	0.162		-0.159		0.332		-0.051		-2.088	***

メンバーシップ関数

属性	セグメント									
	s1		s2		s3		s4		s5	
z1	8.559	***	3.079		5.053		0.000		2.414	
	(2.42)		(2.40)		(2.41)		(0.00)		(3.26)	
z2	-3.928	**	1.715		-3.798	**	0.000		0.686	
	(1.52)		(1.46)		(1.51)		(0.00)		(2.16)	
z3	-3.047	**	0.426		-2.757	**	0.000		-1.424	
	(1.04)		(0.97)		(0.93)		(0.00)		(1.31)	
z4	-0.301		-3.608	***	1.990	***	0.000		-0.399	
	(0.81)		(0.73)		(0.68)		(0.00)		(1.14)	

想定値との差

属性	セグメント									
	s1		s2		s3		s4		s5	
z1	5.559	**	4.079	*	7.053	***	0.000		2.414	
z2	-2.928	*	-1.285		-5.798	***	0.000		-1.314	
z3	-2.047	**	-1.574		-3.757	***	0.000		-2.424	*
z4	0.699		0.392		2.990	***	0.000		2.601	**

括弧内の数値は100回のブートストラップによって推定した標準誤差。

***は1%水準, **は5%水準, *は10%水準で有意を意味する。

表 5 限界支払意思額の分析結果

モデル 1

属性	セグメント									
	s1		s2		s3		s4		s5	
x2	0.243 *** (0.02)	***	0.185 *** (0.02)	***	-0.364 *** (0.04)	***	0.926 *** (0.13)	***	0.205 *** (0.05)	***
x3	0.427 *** (0.02)	***	0.527 *** (0.03)	***	0.943 *** (0.05)	***	-0.487 *** (0.06)	***	0.489 *** (0.06)	***
x4	0.620 *** (0.02)	***	-0.170 *** (0.02)	***	0.663 *** (0.03)	***	0.972 *** (0.11)	***	0.387 *** (0.05)	***
x5	-0.176 *** (0.02)	***	0.515 *** (0.02)	***	-0.354 *** (0.04)	***	-0.459 *** (0.10)	***	0.035 (0.05)	

モデル 2

属性	セグメント									
	s1		s2		s3		s4		s5	
x2	0.239 *** (0.02)	***	0.186 *** (0.02)	***	0.933 *** (0.13)	***	-0.383 (0.22)		-0.306 (0.23)	
x3	0.428 *** (0.02)	***	0.528 *** (0.03)	***	-0.493 *** (0.06)	***	0.930 *** (0.18)	***	0.954 *** (0.24)	***
x4	0.619 *** (0.02)	***	-0.170 *** (0.02)	***	0.974 *** (0.12)	***	0.628 *** (0.10)	***	0.698 *** (0.13)	***
x5	-0.175 *** (0.02)	***	0.516 *** (0.02)	***	-0.464 *** (0.10)	***	-0.278 *** (0.10)	***	-0.505 (0.21)	*

想定値との差 モデル 1

属性	セグメント									
	s1		s2		s3		s4		s5	
x2	0.043 * (0.02)	*	0.019 (0.02)		-0.030 (0.02)		-0.074 (0.02)		-0.545 *** (0.02)	***
x3	0.027 * (0.02)	*	0.027 (0.02)		-0.057 (0.02)		0.013 (0.02)		0.239 *** (0.02)	***
x4	0.020 (0.02)		-0.003 (0.02)		-0.003 (0.02)		-0.028 (0.02)		0.637 *** (0.02)	***
x5	0.024 (0.02)		0.015 (0.02)		-0.020 (0.02)		0.041 (0.02)		-0.215 *** (0.02)	***

想定値との差 モデル 2

属性	セグメント									
	s1		s2		s3		s4		s5	
x2	0.039 (0.02)		0.019 (0.02)		1.267 *** (0.02)	***	-1.383 *** (0.02)	***	-1.056 *** (0.02)	***
x3	0.028 * (0.02)	*	0.028 (0.02)		-1.493 *** (0.02)	***	1.430 *** (0.02)	***	0.704 *** (0.02)	***
x4	0.019 (0.02)		-0.003 (0.02)		0.308 *** (0.02)	***	-0.372 *** (0.02)	***	0.948 *** (0.02)	***
x5	0.025 (0.02)		0.016 (0.02)		-0.131 (0.02)		0.222 ** (0.02)	**	-0.755 *** (0.02)	***

括弧内の数値は 100 回のブートストラップによって推定した標準誤差。

***は 1%水準, **は 5%水準, *は 10%水準で有意を意味する。

表 6 想定値をランダム生成したときの有意な変数の比率

所属確率	モデル 1		モデル 2		
	効用関数	WTP	効用関数	メンバーシップ関数	WTP
0.2 未満	50.3%	47.8%	52.5%	20.1%	48.4%
0.2 以上 0.4 未満	77.4%	76.1%	74.0%	20.0%	68.8%
0.4 以上	95.8%	92.1%	85.3%	36.8%	81.6%
全部	65.2%	62.8%	63.0%	24.1%	58.8%

少なくとも有意水準 10 % で有意となった変数の比率。

20 回のモンテカルロシミュレーションにより計測。

各シミュレーションでは試行回数 20 回のブートストラップにより推定値の標準誤差を推定。