

論 文

「家計調査」の費目間構造に関する一考察*

—文化需要費目に着目したケース—

橋 本 紀 子
荒 木 孝 治

要 旨

個別消費関数の推定では、従来、扱うパラメータ数を制限するため、段階的に個別費目への支出額が決定されると想定してきた。本稿は、この問題を克服するため、ベイジアンネットワークを利用し、より細目まで対象を広げた費目間のネットワーク構造の検出を試みた。

「家計調査」の公表データを用いた検証の結果、世帯のタイプや費目数のレベルにかかわらず、費目間の安定的な階層的構造を検出した。たとえば、食品や光熱・水道といった必需性の高い費目が始点となる費目間構造が検出された。また、文化需要費目である教養娯楽費の内容を細目まで考慮することで、より詳細な費目間の関係に関する知見が得られた。一方、住居費の位置付けや費目数を増やした場合の推定の安定性に問題が見られ、今後、データや推定法にさらなる改善の必要性が示唆された。

キーワード：家計調査；費目間構造；ベイジアンネットワーク；文化需要
経済学文献季報分類番号：15-71；15-72；16-10

1. はじめに

本稿は、「家計調査」の公表データに対し、ベイジアンネットワーク手法を利用し、二人以上世帯の支出行動における費目間のネットワーク構造を明らかにしようと試みたものである。その際、大費目のみならず、中費目さらには細目までを対象にすることができるか、分析対象とする費目のレベルが異なる場合費目構造に変化が生じるかについて、とりわけ文化需要に係わる費目（家計調査においては「教養娯楽」）に着目して分析を行った。

従来、所得（支出額）を多数の費目に配分する個別消費の推定を行う際には、多段階で意思決定が行われると仮定してきた。すなわち、家計は第一段階で可処分所得を消費と貯蓄に配分する。次に、第一段階の決定を所与として、消費額をいくつかの大きな費目（食品、住

居、被服など）に配分する。さらに第三段階では、たとえば第二段階で決定された食品支出額を食品費目（穀類、魚肉、野菜など）に分配し、次に穀類費を細目（米、麺類、パンなど）に分配し、と順次、支出が決定されると考える。個別消費関数の実証分析においては、上記の想定の下、分析対象のレベルに応じ、所与の費目グループに対する支出額がどのように各費目に配分されるかが分析されてきた。

容易に想像されるが、この仮定は支出関数の関数形やパラメータ値に大きな制約を課す。また、実際には存在するであろう費目をまたがる効果、たとえば観劇に出かける（家計調査の大費目では「教養娯楽」支出）ため弁当を買う（「食品」内の中費目“調理食品”）、外で食事を済ます（「食品」内の“外食”）といった行動を直接、分析対象とすることができない。

このような制約的なアプローチがとられてきた最大の理由は、推定可能性にある。対象とする費目数を増やしたいのはやまやまであるが、一方で、分析に利用できるデータのサンプル数は限られている。公表データを利用する場合、提供される内容に限りがあること、さらに時系列データの場合には季節性や構造変化の問題によりより推定手法に負荷がかかることなどから、モデルにおける推定すべきパラメータ数を増やせない事情があったからである¹⁾。しかし、近年、推定手法に大きな改善が見られる。

個別消費関数推定の最終目標は、所得が与えられた時、家計が直面する費目に対しどのように支出額を決定していくかを明らかにすることである。たとえば食品という費目は、どの費目と一定以上の（有意な）強さで関係しているのか、その関係はいずれが原因でいずれが結果か、その影響の大きさはいくらか等について知ることである。本稿では費目間構造について、どの費目とどの費目が関連しているか、その関連の因果の向きはどうなっているかに問題を絞り、非巡回有向グラフを描き出すことのできるベイジアンネットワークの手法を用いて、多数の費目間構造の検出を試みた。

この際、「家計調査」は10大費目分類を対象とするが、細目まで詳細に分析する費目として文化需要に係わる「教養娯楽」の動きに着目した。文化需要を特に取り上げた理由は、下記の通りである。

日本経済は第2次大戦後大きく経済成長し、生活は豊かになった。所得が増加した結果、消費は量的にも質的にも成熟し、多様化してきた。一方で、家計の在り方も大きく変わってきた。核家族化、さらに少子化が拍車を掛け、世帯人員数は減少の一途をたどっている。現在では従来「標準世帯」と呼ばれた両親と子供2人の4人家族は少数派となり、世帯数では単身世帯の比率が最大となっている。高齢化も顕著である。しかも、このような動きは今後いっそう増すと予想される。個別消費の動きを見ていく際、これらの動きに留意した分析が必要である。

では、消費がどのように豊かになってきたか、今後どのように変化していくかを検討するにはどのような切り口が適切であろうか。本稿では、文化需要、文化芸術活動への参加及び支出行動に着目する。その理由は、文化需要はまさに「豊かさ」の現れであるからである。消費の成熟化に伴い着実にその支出額は増えてきた。しかしその内容の変化について分かっていることは実はさほど多くない。たとえば、さまざまな家計の属性変化は文化需要にどのように影響してきただろうか。さらには、今後の人口減少、少子高齢化による人口構造変化はどのような影響をもたらすだろうか²⁾。

文化需要の動向については、すでに「社会生活基本調査」データを用いて、余暇時間の使い方からその特徴を捉えようとする研究が重ねられている。「社会生活基本調査」では、生活時間とともに、余暇時間をどのように用いているか生活行動に着目した調査が行われており、買い物やスポーツあるいは旅行や文化鑑賞といった趣味や娯楽などの行動にどれだけの時間を費やしたかを把握することができる。これらの行動に性別、年齢が影響することは容易に想像されるが（永山・勝浦・衛藤（2010））、それらに加えて勝浦（2012）では教育の、西郷（2012）では就業状態の影響も見られることが指摘されている。さらに、たとえ同じ性別・年齢・学歴・就業状態の人でも、1980年代と2000年代でその行動が異なると予想されることから、コーホートすなわち時代による効果の重要性も指摘されている。

このように文化需要の動きには家計の属性変化が大きく影響するが、もちろん所得の動きも大きく係わる。たとえばKatsuura（2012）は、「家計調査」の公表データにより文化需要の時系列特性の分析を行い、景気変動と文化需要の関連性を明らかにしている。

上記の研究の流れを受け、本稿では文化需要の動きについて、支出行動の側面から分析を行う。最終的な目標はマイクロデータを用い家計の属性がいかに多費目に対する支出行動に影響するかを明らかにしていくことであるが、本稿では試論として、「家計調査」の公表データを用いて、支出行動における各費目の重要度や関連、因果の方向の検出が、教養娯楽の細目データまで含む多費目モデルに拡張した場合に可能かどうか、どのような結果が得られるかについて検討していく。

本稿の構成は以下の通りである。第2節でベイジアンネットワーク手法の考え方について説明する。第3節では、分析に用いるデータを紹介する。本稿では「家計調査」の二人以上世帯を対象とした月次データを利用するが、勤労者世帯に限るか、農林漁家世帯を含むかなどから、いくつかのタイプがある。それら世帯タイプの違いや、推定期間（2000年1月から2012年11月）における特徴を概観する。第4節で推定結果を示す。各タイプの世帯が想定する費目間の構造について、総支出額が与えられた下で家計が支出配分を決めていく際、いずれの費目と費目が（相対的に）強く関連していたか、その関連の向き（因果の方向）は

いずれか等の検討結果を示す。本稿は家計の支出行動や文化需要の動きを捉えていく上での試験的な第一歩である。残された課題、今後分析していく必要のある事柄について第5節にまとめる。

2. ベイジアンネットワーク

ベイジアンネットワーク (Bayesian Network) は、様々な事象間の因果関係をグラフで表現し、それを用いて因果推論を行うための手法である。有向グラフで表現され、矢印の方向が因果関係や情報の流れを示すと解釈することにより、全体の構造を直感的に把握することができる手法である。そのため医学や工学、社会科学等様々な分野で利用されてきた (Pourret et al. (2008))。本節ではベイジアンネットワークの基本的な考え方を概説する。

2.1 ベイジアンネットワークとは³⁾

グラフ G は、 d 個の点 (ノード) の集合 $V = \{v_1, \dots, v_d\}$ と、ノードを結合する線 (エッジ) の集合 E との組 $G = (V, E)$ である。2つのノード v_i, v_j の間のエッジが有向であるとは、 $(v_i, v_j) \in E$ かつ $(v_j, v_i) \notin E$ のときをいい、これを $v_i \rightarrow v_j$ と表す。 $v_i \rightarrow v_j$ のとき、 v_j を v_i の子、 v_i を v_j の親 (parent) という。ノード v_i の親全体を $Pa(v_i) = \{v_j : v_j \rightarrow v_i\}$ と表す。親を持たないノードをルート、子を持たないノードをシンクという。また、ノード v_i から線の向きを辿って到達できるノードの集まりをノード v_i の子孫 (descendant) といい、 $D(v_i)$ と表す。ノードの集合 V に対して、 V に含まれるノードの集合 $S = \{v_k, v_l, \dots\}$ ($\subset V$) を V から除いた集合を、 $V \setminus S$ または $V \setminus \{v_k, v_l, \dots\}$ と表す。 $ND(v_i) = V \setminus \{v_i\} \cup D(v_i)$ をノード v_i の非子孫 (non-descendant) という。

ノード u からノード v への有向道とは、異なるノードの列 (v_0, v_1, \dots, v_n) において、 $v_0 = u, v_n = v$ であり、かつ、全ての $k (1 \leq k \leq n)$ に対して $v_{k-1} \rightarrow v_k$ であることをいう。また、向きの混合 ($v_{k-1} \rightarrow v_k$ または $v_k \rightarrow v_{k-1}$) を許した経路を道という。ノード v に対して、 v から出発するノードの向きを辿って v に戻ってくる有向道を巡回閉路 (サイクル) という。

有向グラフとは、全てのエッジが有向であるグラフのことであり、有向グラフでサイクルが存在しないものを、非巡回有向グラフ (DAG: Directed Acyclic Graph) という。ベイジアンネットワークでは非巡回有向グラフを考える。

2つの異なるノード u, v がノードの集合 $S \subset V \setminus \{u, v\}$ により d -分離またはブロックされるとは、次のいずれかが成立することをいう。

- (i) 3つのノードが有向道 ($v_{k-1} \rightarrow v_k \rightarrow v_{k+1}$ または $v_{k-1} \leftarrow v_k \leftarrow v_{k+1}$)、あるいは分岐道 ($v_{k-1} \leftarrow v_k \rightarrow v_{k+1}$) を構成するとき、 $v_k \in S$ である。

- (ii) 2つのノードがあるノードに合流する ($v_{k-1} \rightarrow v_k \leftarrow v_{k+1}$) とき (v_k を合流点という)、 v_k およびその子孫は S に含まれない ($\{v_k\} \cup D(v_k) \notin S$)。

グラフ G において、ノードの集合 Z がノード u と v の間の全ての道をブロックするとき、集合 Z は u と v をブロックするといひ、 $d-sep_G(u, v | Z)$ と表す。また、ブロックされない場合、接続されている (connected) といひ、 $d-con_G(u, v | Z)$ と表す。なお、記号が煩雑になるため、文脈からわかる場合、グラフの添え字 G は省略することがある。

ベイジアンネットワークでは、事象を確率変数で表現し、これらの間の独立・依存関係を用いてグラフを構成する。そのため、ここで確率変数の独立性に関する基本的な結果を見ておく。

d 個の確率変数 X_1, X_2, \dots, X_d に対して、同時分布 P を考える。ここでは確率変数の連続性を仮定して、同時密度関数 $f_{X_1 X_2 \dots X_d}(x_1, x_2, \dots, x_d)$ を考える⁴⁾。このとき、 X_2, \dots, X_d を与えたときの X_1 の条件付き (周辺) 密度関数 $f_{X_1 | X_2 \dots X_d}(x_1 | x_2, \dots, x_d)$ を、

$$f_{X_1 | X_2 \dots X_d}(x_1 | x_2, \dots, x_d) = \frac{f_{X_1 X_2 \dots X_d}(x_1, x_2, \dots, x_d)}{f_{X_2 \dots X_d}(x_2, \dots, x_d)}$$

と定義する。同様に、 $f_{X_1 X_2 | X_3 \dots X_d}(x_1, x_2 | x_3, \dots, x_d)$ 等の条件付き (周辺) 同時密度関数を考えることができる。

確率変数 X, Y に対して、同時密度関数を $f_{XY}(x, y)$ 、周辺密度関数をそれぞれ $f_X(x)$ 、 $f_Y(y)$ とする。このとき、確率変数 X, Y が独立であるとは、

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

が成立することをいい、これを $X \perp Y$ と表す。また、3つの確率変数 X, Y, Z に対して、 Z が与えられたときの X と Y の条件付き密度関数 $f_{XY|Z}(x, y | z)$ に関して、

$$f_{XY|Z}(x, y | z) = f_{X|Z}(x | z) f_{Y|Z}(y | z)$$

のとき、 X と Y は Z が与えられたときに条件付き独立であるといひ、 $X \perp_p Y | Z$ と表す。条件付き独立ではない場合、条件付き依存といひ、 $X \not\perp_p Y | Z$ と表す。なお、記号が煩雑になるため、文脈からわかる場合、分布の添え字 P は省略する。

グラフ $G = (V, E)$ のノード $V = \{v_1, v_2, \dots, v_d\}$ に確率変数 $X = (X_v)_{v \in V}$ を対応させ、ノードと確率変数とを互いに同等なものとして扱う。また、 X の同時分布を $P(V)$ とする。以下、 X の任意の要素も X と記す。このとき X に対して、 $Pa(X)$ を条件付けたとき、 X_u と $X_v \in ND(X)$ が独立、つまり、

$$X_u \perp X_v | Pa(X)$$

のとき、 $G = (V, E)$ は確率分布 P の下でマルコフ的であるといひ、 (G, P) をベイジアンネットワークという。ベイジアンネットワークに関して、次が成立する：

任意の X に対して親 $Pa(X)$ を条件付けたときの条件付き分布 $P(X | Pa(X))$ を考えると、 $P(V)$ は次の形で分解できる。

$$P(V) = \prod_{X \in V} P(X | Pa(X))$$

分布 P がグラフ G に対して忠実性 (faithfulness) 条件を満たすとは、 $X \perp_p Y | Z$ と $d\text{-sep}(X, Y | Z)$ が同値の場合をいう。これにより、条件付き独立性を確認することと d -分離性を検証することとは等しくなる。よって、ノード X と Y の間にエッジがあることは全ての $Z \subset V \setminus \{X, Y\}$ に対して $X \not\perp_p Y | Z$ となることと同値であり、エッジがないということは、ある $Z \subset V \setminus \{X, Y\}$ が存在し、 $X \perp_p Y | Z$ となることと同値である。

グラフ $G = (V, E)$ は、確率変数の集合 $X = (X_v)_{v \in V}$ に対して同時分布 P に関する統計モデルを決定する。ベイジアンネットワークは、確率変数を非巡回有向グラフのノードに対応させ、これらの確率変数間の確率構造をデータを用いて推定し、非巡回有向グラフで表現する手法である。

2.2 ベイジアンネットワークの推定⁵⁾

2.1 節で見たように、ベイジアンネットワークは、グラフ構造と条件付き分布 (パラメータ) とのペアで構成される。そのため、データからベイジアンネットワークを推定するには、グラフ構造を学習し (モデル選択ともいう)、パラメータを推定する必要がある。構造学習の際、まずは方向のないグラフ構造 (スケルトン) を定め、その後、エッジの方向を決定する形で推定を行う。その手法として、①スコアベース法、②制約ベース法、③これら両者のハイブリッド法が考えられている。本稿での分析は、データ解析環境 R のパッケージ bnlearn を利用するが、bnlearn ではスコアベース、制約ベース、ハイブリッドの代表的な手法を利用することができる。

①スコアベース法は、ネットワークを評価する基準であるスコアを設定し、そのスコアが良くなるようにネットワーク構造を逐次学習していく方法である。代表的な手法として山登り法 (Hill Climbing) がある。スコアとしては、連続型確率変数の場合、多変量正規分布の対数尤度、赤池情報量基準 (AIC)、ベイジアン情報量基準 (BIC) 等が利用できる。これに対し、②制約ベース法は、確率変数間の (条件付き) 独立性を検定により判定し、独立な変数間のエッジは削除し、依存するエッジ間のエッジは追加するという形で構造を学習していく。このとき、(条件付き) 独立性を検定するための統計量・基準が必要となる。その検定には、確率変数が連続型の場合、相互情報量 (Mutual Information) や (偏) 相関係数に基づくもの、ノンパラメトリックな並べ替え検定等、様々ある。なかでも条件付き独立性を

判定するための基準としてよく用いられるのが、確率変数 $Z = V \setminus \{X, Y\}$ を固定したときの X と Y の偏相関係数 $\rho_{XY|Z}$ の推定値 $\hat{\rho}_{XY|Z}$ である。これが t 分布に従う、あるいは、これを z 変換したものが漸近的に正規分布に従うことを利用して独立性の検定を行うことができる。偏相関係数が使われる理由は次の通りである。

ランダムベクトル (X_1, X_2, \dots, X_d) が多変量正規分布 $N_d(\mu, \Sigma)$ に従っているとす。共分散行列 Σ の逆行列 $K = \Sigma^{-1}$ （精度行列という）の要素を次のように記す。

$$K = \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1d} \\ k_{21} & k_{22} & \cdots & k_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ k_{d1} & k_{d2} & \cdots & k_{dd} \end{pmatrix}$$

このとき、 X_u と X_v 以外の変数 ($Z = V \setminus \{X_u, X_v\}$ とする) が与えられたときの偏相関係数は

$$\rho_{X_u X_v | Z} = -k_{uv} / \sqrt{k_{uu} k_{vv}}$$

となる。よって、 $k_{uv} = 0$ のとき、 $\rho_{X_u X_v | Z} = 0$ となる。つまり、多変量正規性の条件の下では、2変数の条件付き独立性と偏相関係数が0であることは同値となる。これより、偏相関係数が0かどうかを判断し、0でないなら2つの確率変数は依存であり、これらの間にノードをつける。また、0と判断できるならば、両者の間のノードは削除すればよい。

①スコアベース法、②制約ベース法はそれぞれ、利点および欠点を持つ (Wang et al., 2007)。②制約ベース法は、計算の効率が良いため、ノードの数が多い場合に適用することが可能である。しかし、独立性を検定する際に利用する有意水準に依存する。また、ノード数に応じたサンプルサイズの大きさが必要となる。ノードの探索の初期の誤りがグラフ構造の異なる推定結果をもたらすという推定の不安定さもある。これに対して一般に、①スコアベース法は制約ベース法と比べて正確なモデル選択の結果を導く。また、サンプルサイズも検定を用いる場合と比べて少なくても良い。欠点としては、計算速度が比較的遅く、特にノード数が多い場合顕著である。こうした問題に対処するため、③ハイブリッド法が提案されている。ハイブリッド法では、基本的にまず制約ベース法を用いて効率的にネットワーク構造を学習し、その後スコアベース法を用いて学習を精緻化するというアプローチが取られる。そのため、ベイジアンネットワークの構造学習における学習効率を強化し、予測精度を向上させることが期待されている。

本稿の分析で用いる `bnlearn` パッケージに実装されているスコアベース法としては、山登り法 (hc)、Tabu Search (tabu)、制約ベース法としては、Grow-Shrink (gs)、Incremental Association (iamb)、ハイブリッド法としては、Max-Min Hill Climbing (mmhc)、

Restricted Maximization 等がある。なお、括弧内は bnlearn における各手法の関数名である。

3. 利用するデータとその特徴

本稿では「家計調査 家計収支編」の時系列データを用いて分析を行って行く。利用するデータは2000年1月から2012年11月までの月次データ（期間数：155）であり、実質化している。なおこの期間を対象にしたのは、2000年（平成12年）に家計調査の調査対象が大きく変更されたからである。

この期間の経済状況をざっと概観しておく。この期間は、バブル崩壊（1990年頃）以降のいわゆる「失われた10年」と呼ばれる状態からの脱却をめざしてさまざまな試みがなされた時期に当たる。2000年にIT景気が終了した後、2002年2月から2008年2月にかけては「いざなぎ景気」と呼ばれる73か月にわたる戦後最も長い景気拡大が生じた。しかし、景気が失速するのにサブプライム問題（世界金融危機）の影響、さらには2008年9月のリーマンショックが追い打ちを掛ける形になり、日本経済も大きな打撃を受ける。2000年に509兆8600億円であった名目GDPは、いったん2002年、2003年と500兆円を割り込むものの2007年には512兆9752億円まで増加するが、その後大きく水準を下げ、2011年には470兆円まで減少している。この間、デフレーションが進行し、GDPデフレーターは一貫して下落している。

雇用状況も芳しくなく、パートタイムを含む有効求人倍率が1を超えたのは2006年と2007年のみである（パートを除いた場合は1を超える年はない）。この結果、世帯当たりの所得も伸び悩み、2011年の所得水準は1990年の水準とほぼ同じにまで落ち込んでいる。

家計調査では、総世帯を調査対象とする⁶⁾が、大きく二人以上世帯と単身世帯に分けて結果が公表される。本稿ではより詳細な情報が提供される二人以上世帯のデータに着目していくが、その内容はさらに次の4つのタイプに分けて公表されている。

- ①二人以上の全世帯
- ②二人以上の全世帯（農林漁家世帯を除く）
- ③二人以上の世帯のうち勤労者世帯
- ④二人以上の世帯のうち勤労者世帯（農林漁家世帯を除く）

第4節ではこれらの4つのタイプのデータを用いて、世帯タイプにより支出の仕方に違いがあるか、検討を行っていく。

なお、2012年11月の状況で①と③の比較を行うと、表1のようになる。ざっと特徴を見ておく⁷⁾。全世帯、勤労者世帯とも農林漁家世帯の割合は多くないこと、持ち家率は高くその広さに大きな違いは無いことがわかる。一方、消費支出額は勤労者世帯の方が若干多い。

世帯人員数は勤労者世帯の方が0.4ポイント近く多く、就学している子供の数は勤労者世帯で多い。一方、高齢者は全世帯に多く、それに応じて無職者の数も全世帯で多い。これらに対応し、勤労者世帯は全体に世帯が若く、世帯主の年齢は全世帯に比べ約10才若い。

次に、支出の対象となる費目について見ておく。「家計調査」では次のi) 10大費目を考えている。

- 1 食品、2 住居（帰属家賃は除く）、3 光熱・水道、4 家具・家事用品、
5 被服及び履物、6 保健医療、7 交通・通信、8 教育、9 教養娯楽、10 その他

ここで、教育費については留意が必要である。家計調査の教育費は授業料、教科書や学習参考教材、補習教育に対する支出からなり、いわゆる学校教育に対応するものである。これに対し、習い事に関する支出は9教養娯楽に含まれる。

表1 全世帯と勤労者世帯の特徴

	全世帯	勤労者世帯
年平均1ヶ月当たりの消費支出額	273,772円	300,181円
世帯人員数	3.06人	3.44人
18歳未満の人数	0.58人	0.96人
65歳以上の人数（無職者数）	0.73人（0.60人）	0.19人（0.13人）
世帯主の年齢	57.8才	47.9才
農林漁家世帯の割合	2.0%	0.3%
持ち家率（平均畳数）	82.8%（39.9畳）	75.3%（39.0畳）

本稿で対象としている世帯には単身世帯は含まれないが、家族構成が複数であっても上記の表から明らかなように、主に世帯主が企業等で就労している勤労者世帯に比べ、自営業あるいは無職をも含む全世帯の世帯主の年齢は高く、子供が独立している可能性が高い。公表されている指標では18歳未満の人数しか分からないが、世帯において教育費の支出対象である修学者数の比率は全世帯より勤労者世帯で高くなっている可能性が大きい。

第1節で文化需要、すなわち教養娯楽について中費目、あるいは細目まで考慮すると述べた。これらの費目内容は下記の通りである。

- ii) 教養娯楽の内容を中費目に細分化（総費目数は13費目）

- 9 教養娯楽 → 91 教養娯楽用耐久財、92 教養娯楽用品、93 書籍・他の印刷物、
94 教養娯楽サービス

- iii) 教養娯楽の内容を細目⁸⁾に細分化（総費目数は16費目）

- 9 教養娯楽 → 91 教養娯楽用耐久財、92 教養娯楽用品、93 書籍・他の印刷物、
941 宿泊料、942 パック旅行費、943 月謝額、944 他の教養娯楽サービス

なお、2012年11月段階で見た支出金額の消費支出全体に占める割合は、全世帯/勤労者世帯それぞれで、教養娯楽費は9.8%/9.4%、教養娯楽に占める比率は教養娯楽用耐久財：6.4%/8.0%、教養娯楽用品：20.1%/21.4%、書籍・他の印刷物：14.4%/13.3%、教養娯楽サービス：59.1%/57.3%（宿泊料：5.9%/4.9%、パック旅行費：13.5%/8.5%、月謝額：10.8%/14.2%、他の教養娯楽サービス：28.9%/29.7%）となっている。

次節では、これらのデータを用いて、世帯タイプの違い、費目レベルの違いにより、費目間構造にどのような違いが見られるか、検討を行っていく。

4. 世帯タイプ別に見た、異なる費目数レベルの費目間構造

本稿での分析はデータ分析環境Rのパッケージbnlearnを用いて行った。Rは、世界中の第一線の研究者や実務家が協力して開発しているオープンソースのデータ解析環境であり、bnlearnはベイジアンネットワーク分析を行うためのパッケージである。

Bnlearnには、ハイブリッド法であるMax-Min Hill Climbingを実行できる関数mmhcがあり、これを利用した。パラメータとしては、独立性の検定における有意水準 α を0.01とし、他は全てデフォルト値を用いた。

既に3節で述べたように、利用するデータは2000年1月から2012年11月までの月次データ（期間数:155）を実質化したものを用いている。連続変数のベイジアンネットワークでは、変数が独立に正規分布に従うことを仮定する。そのため、第 k 費目の第 i 月および第 $i+1$ 月のデータをそれぞれ $x_{k,i}$ 、 $x_{k,i+1}$ とするととき（ $i=1,2,\dots,114$ ）、連続した月のデータの比を対数変換したもの、すなわち、 $\log(x_{k,i+1}/x_{k,i})$ をデータとして利用する。さらに費目単位での標準化を行っている。

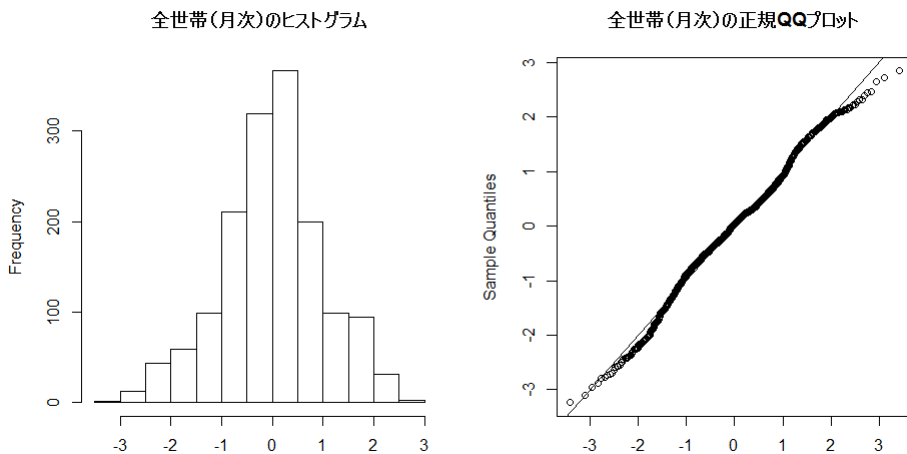


図1 全世帯の10費目の対数変換データのヒストグラム（左）およびQQプロット（右）

数値変換の効果を見るために、例として、全世帯の10費目の変換データのヒストグラムおよびQQプロットを図1に示す。ヒストグラムおよびQQプロットより、分布がほぼ標準正規分布となっていることがわかる。他の費目データに関してもほぼ同じ結果なので省略する。

次に、ベイジアンネットワークの推定を行い、全世帯、勤労者世帯それぞれについて同じ費目数レベルで描かれた費目間構造が、農林漁家世帯を含むかどうかにより異なるかどうかを検討する。

6つのケースを比較した結果、全世帯の10費目レベルと16費目レベルではまったく同じ非巡回有向グラフが得られた。他の4つのケース（全世帯の13費目、勤労者世帯の10、13、16費目）では同一ではなかったものの、ルート（始点）やシンク（終点）、分離されたノード（孤立点）がある場合にはその点は等しく、経由する点の順がまれに異なる程度の違いが見られた。

図2に勤労者世帯の10費目レベルのケースを示す。図より、農林漁家世帯の有無によらず、教育は分離されたノード（点）であること（他の費目と影響を与え合わない、あるいは非常に弱い）、食品や光熱・水道は他から影響を受けないルート（始点）であり、住居、家具・家事用品、保健医療は他に影響を与えないシンク（終点）であることがわかる。異なる点は、食品から保健医療に至る因果の経路（プロセス）に教養娯楽が関連するかどうかということと、交通・通信の位置づけの2つである。

他の4つのケースも農林漁家の有無による費目別構造への影響はさほど見られなかったため、以下、より情報の多い、農林漁家世帯を除いたデータ（②全世帯、④勤労者世帯）について検討を続ける。

全世帯と勤労者世帯での違いを見るために、費目数レベルごとにその費目別構造を検討する。図3に大費目のみに着目した10費目分類の結果、図4に教養娯楽費について中費目まで見た13費目分類の結果、図5に教養娯楽費について細目まで着目した16費目分類の結果を示す。

図3より、10費目分類に対しては、全世帯、勤労者世帯とも教育は孤立した点であり、食品や光熱・水道は始点として他の費目に影響を与える一方、住居や保健医療は終点であり、他の費目に影響を及ぼすことはない。他方、全世帯では交通・通信は終点であるが、勤労者世帯では教養娯楽から影響を受けその他の消費支出に影響を与える有向道に含まれる。また逆に勤労者世帯では、家具・家事用品は終点であり、食品、光熱・水道、被服及び履物、教養娯楽から影響を受けるが、他の費目には影響を及ぼさない。しかし、全世帯では食品、被服及び履物、教養娯楽から影響を受け、保健医療に影響を与える結合点である。全世帯にお

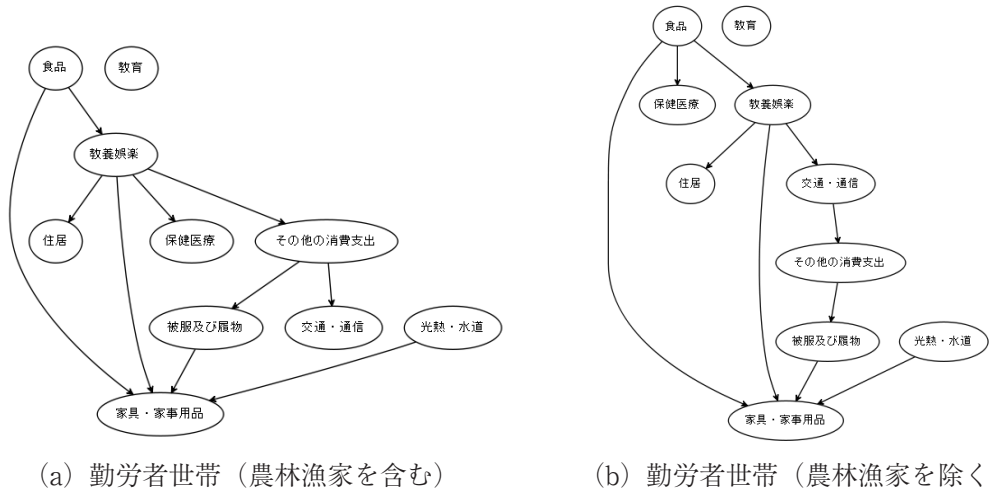


図2 農林漁家の有無による費目別構造の違い：i) 10費目分類

いては光熱・水道は、勤労者世帯と異なり住居にも影響を与え、家具・家事用品に直接影響せず被服及び履物を經由して影響する。しかし、ネットワークにおける変数の順序を考えると、ほぼ同じ位置にあると判断することができる。

13費目分類（図4）では、勤労者世帯に大きな変化が起こる。これまで孤立した点であった教育が始点となり、家具・家事用品と教養娯楽サービスに影響する。10費目分類と同じく、全世帯、勤労者世帯ともで食品、光熱・水道は始点である。ただしその影響先を見ると、10費目分類では食品は全世帯において住居、家具・家事用品、教養娯楽に影響を与えていたが、教養娯楽の内訳が分かれることにより、住居への影響は93書籍・他の印刷物を介して

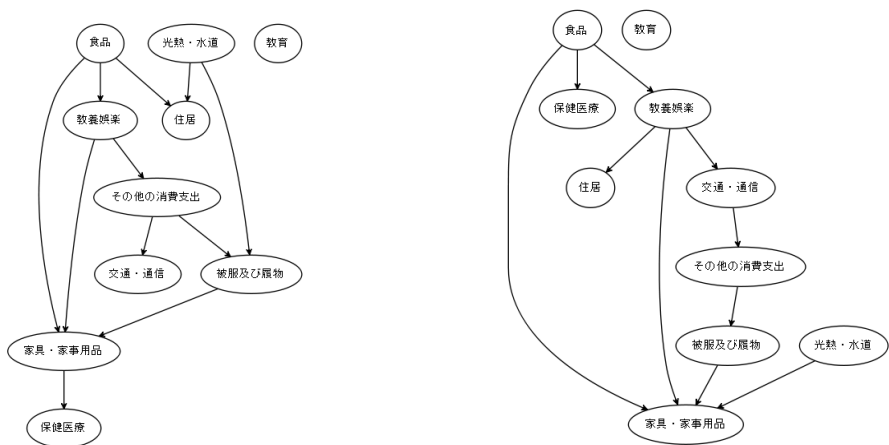
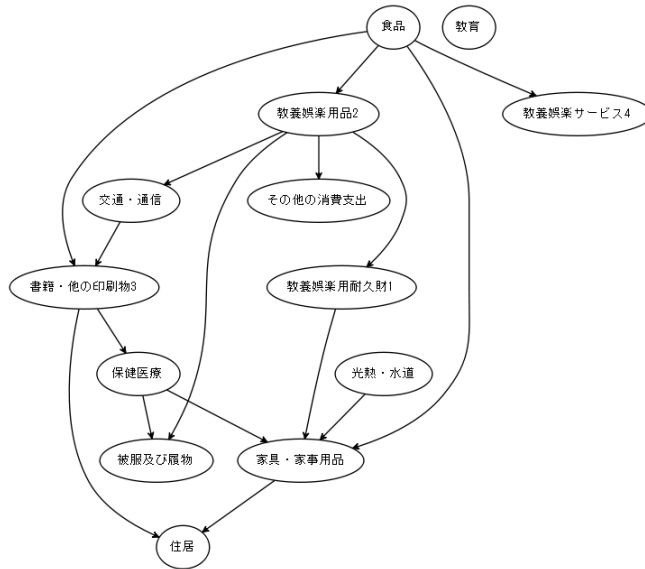
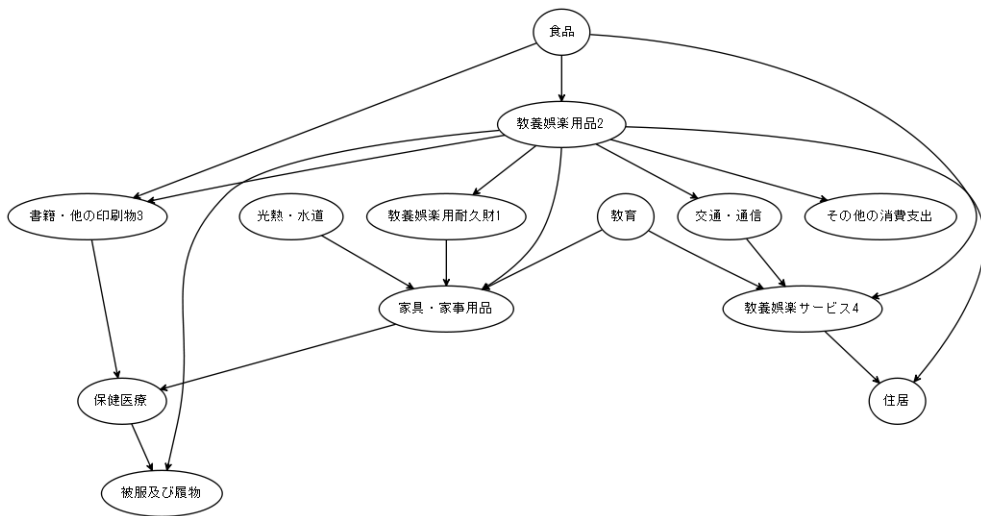


図3 世帯タイプによる費目別構造の違い：i) 10費目分類

いることが分かる。また食品から教養娯楽への影響は 92 教養娯楽用品、93 書籍・他の印刷物、94 教養娯楽サービスへと分かれ、91 教養娯楽耐久財には 92 教養娯楽用品を介しての影響となることが分かる。94 教養娯楽サービスは終点であり、他の費目に影響を及ぼさない。勤労者世帯でも教養娯楽が中費目に細分化されることにより、食品の影響は住居、92 教養娯楽用品、93 書籍・他の印刷物の 3 費目に生じる。92 教養娯楽用品から 91 教養娯楽用耐久



(a) 全世帯（農林漁家を除く）



(b) 勤労者世帯（農林漁家を除く）

図4 世帯タイプによる費目別構造の違い：ii) 13費目分類

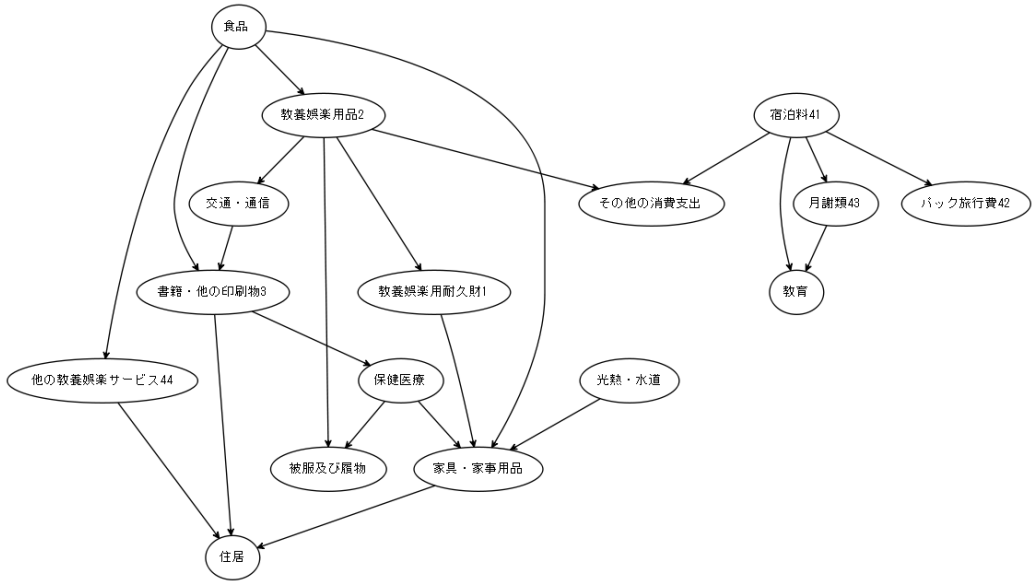
財への影響経路があるのは全世帯と同様であるが、勤労者世帯では94 教養娯楽サービスは有向路の中に入ってきており、92 教養娯楽用品から直接また交通・通信を介して影響を受け、住居に影響を与えている。

図5より教養娯楽の細目まで考えた16 費目分類では孤立点はない。教育は全世帯では終点である。勤労者世帯では始点となっており、教養娯楽サービスの細目（941 宿泊料、943 月謝類、942 パック旅行費）を経由して交通・通信へと影響する。10 費目あるいは13 費目分類と同様食品や光熱・水道は始点であり、全世帯ではそれらに加えて941 宿泊料も始点となる。因果の向きは全世帯、勤労者世帯で異なるが941 宿泊料、943 月謝類、942 パック旅行費、教育が関連する（線で結ばれる）様子が観察される。双方の世帯タイプで住居、その他の支出が終点であり、全世帯では教育、被服も終点である。他方、勤労者世帯では交通・通信、保健医療、91 教養娯楽用耐久財、943 月謝も終点であり、これらの費目が家具・家事用品と関連し合わない点が全世帯とは異なる。なお全世帯では、その他の消費支出を固定すると、右側の {941 宿泊料、943 月謝類、942 パック旅行費、教育} のグループと、左側のその他のグループとが条件付き独立になるという特徴を持つ。

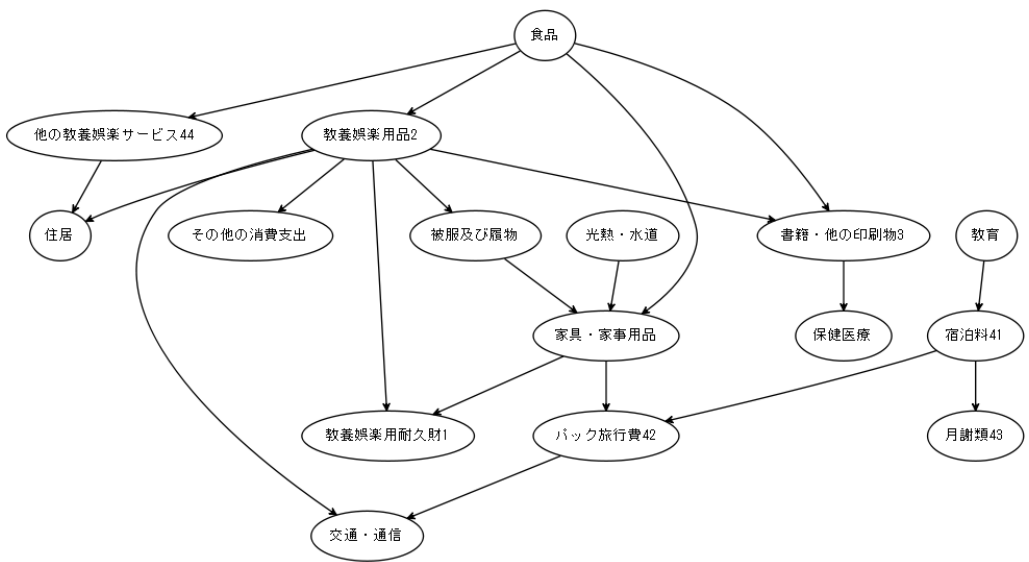
このように全世帯と勤労者世帯で費目構造は異なる。本稿で用いたデータからでは確たる予測は難しい。その原因の一つに、非巡回有向グラフを推定する際に、推定するパラメータ数に対するデータ数の問題をあげることができる。費目数を増やしたとき、細分化されることによる効果に加えて、データ数の問題が生じる。そのため、費目が増えると、推定が不安定になっている可能性が高い。さらに、より根本的な違いの原因として両世帯の人員構成、その年齢が影響していると推察される。すなわち、勤労者世帯は比較的若く、就学している子供も含まれる世帯であることから教育が他の費目に影響を与える（先に支出が決定される）費目として扱われ、一方で、全世帯には比較的高齢者が多いため、教育に支出額の比重が置かれにくい。他方、時間等が自由になる高齢者が多いことから、教養娯楽サービス（宿泊料やパック旅行費）への支出額が他の費目より先に考えられている可能性がある。今後、このような世帯の特性を考慮した詳細な検討が必要である。

さて、次に、図3、図4、図5の（a）をそれぞれ比較することにより全世帯について、（b）をそれぞれ比較することにより勤労者世帯について、費目分類を変えることによりどのような結果の違いが見られるか、検討してみよう。

全世帯、勤労者世帯ともに教養娯楽の費目をより細分化することにより、それまで分離した点であった教育を、有向グラフの中の連結された点としてとらえ、他の費目との関連をあぶり出すことができた。また、はじめにで予想した通り、教養娯楽の中費目、細目は他の費目に影響を与えたり与えられたりしており、そのメカニズムは多段階の支出プロセスで考え



(a) 全世帯（農林漁家を除く）



(b) 勤労者世帯（農林漁家を除く）

図5 世帯タイプによる費目別構造の違い：iii) 16費目分類

られているほどシンプルではないことも明らかになった。しかし、ノードを始点や終点としてとらえたとき、各ネットワークはかなり安定したものになっていることがわかる。参考までにこれらを表に示すと、表2のようになる。

表2 全世帯・勤労者世帯の10、13、16費目のベイジアンネットワークにおける
始点、終点、孤立点の関係（○は、全世帯と勤労者世帯で一致、●は異なるもの）

		全世帯（農林漁家を除く）			勤労者世帯（農林漁家を除く）		
		始点	終点	孤立点	始点	終点	孤立点
10 費目	1 食品	○			○		
	2 住居		○			○	
	3 光熱・水道	○			○		
	4 家具・家事用品					●	
	5 被服及び履物						
	6 保健医療		○			○	
	7 交通・通信		●				
	8 教育			○			○
	9 教養娯楽						
	10 その他						
13 費目	1 食品	○			○		
	2 住居		○			○	
	3 光熱・水道	○			○		
	4 家具・家事用品						
	5 被服及び履物		○			○	
	6 保健医療						
	7 交通・通信						
	8 教育			●	●		
	91 教養娯楽用耐久財						
	92 教養娯楽用品						
93 書籍・他の印刷物							
94 教養娯楽サービス		●					
10 その他		○			○		
16 費目	1 食品	○			○		
	2 住居		○			○	
	3 光熱・水道	○			○		
	4 家具・家事用品						
	5 被服及び履物		●				
	6 保健医療					●	
	7 交通・通信					●	
	8 教育		●		●		
	91 教養娯楽用耐久財					●	
	92 教養娯楽用品						
	93 書籍・他の印刷物						
	941 宿泊料	●					
	942 パック旅行費		●				
943 月謝額					●		
944 他の教養娯楽サービス							
10 その他		○			○		

5. まとめと残された課題

本稿では、「家計調査」の公表データを用いて、ベイジアンネットワーク手法を利用し、二人以上世帯の支出行動における費目間のネットワーク構造の検出を試みた。

二人以上世帯は農林漁家を含むデータ、含まないデータが利用可能であるが、その有無により得られる費目間構造に大きな違いは見られなかったため、農林漁家データを除いた全世帯と勤労者世帯において、10大費目データあるいは文化需要に該当する教養娯楽費目について細目まで考えたデータに対し、どのような費目構造が見られるか、費目の細分化のレベルにより構造に変化が生じるか、検討を行った。

いずれの世帯、費目レベルにおいても、食品および光熱・水道は必ず始点であった。これらの費目は必需性が高いことが知られており、まずこれらの費目の支出金額が決定され、その後他の費目への配分が決まっていくプロセスは妥当と考えられる。教育は孤立点、あるいは勤労者世帯で費目数が多い場合は始点として検出されたが、両者に共通するのは、他の支出額に依存せず、独立に支出額が決定される傾向が強い点である。支出額決定後、相対的に就学している人数が少ない全世帯では他の費目への影響が観察されず、就学人数が比較的多く教育支出が実質的な意味を持つ勤労者世帯では始点となり、他の費目の支出に影響が生じたと考えられる。

終点は、奢侈性とまでは言えないが、比較的所得弾力性が高いと想定される、生活のしかたにより支出額の増減が可能な費目、たとえば被服やその他の消費等が検出された。若干、違和感があるのは、住居が世帯、費目レベルによらず全てのケースで終点として検出されたことである。しかし、これは家計調査における住居費目において、支出額は家賃であり、帰属家賃が含まれていないという問題があることによると思われる。この点は、今後、たとえば消費実態調査のように帰属家賃を含む支出データを用いさらに検証していく必要がある。

費目レベルを変化させることにより、費目構造について新たな切り口を発見することができたが、16費目の結果は若干推定の不安定さを感じる結果となっている。これはデータ数(155)に比して推定すべきパラメータ数が増大することから生じたものと思われる。これについては今後、ある程度ネットワーク構造を推定した段階で、偏回帰(PLS: Partial Least Squares)手法を用いてパラメータの推定を行うことが考えられる。これは、PLSでは、変数の正規性を仮定する必要がなく、また、データ数も多くある必要がないという特徴があるからである。さらに、データを離散化することにより離散データを用いたベイジアンネットワークを推定し、変換しないデータおよびPLSを用いて支出の決定プロセスを推定することも考えることができる。また、時系列データが分析対象なので、時間変数を取り込むダイナミック・ベイジアンネットワークを利用することも考えられる。

推定結果を安定させる手立てとしてはデータ数を増加させることも考えられる。今回は時系列データを用いたが、たとえばマイクロデータを利用することができれば飛躍的にデータ数を増やすことができ、また、より詳細な家計属性(所得や貯蓄の水準、家族構成員数、年齢、

性別、住宅状況、ローンの有無等）が費目構造にどのように影響しているかについても検討を行うことができる。

このように手法、およびデータがより堅固なものとなれば、今回は教養娯楽についてのみ細目までを対象に構造を考えたが、たとえば食品の内容について素材食品と調理食品、外食、嗜好品類といった分類を採用することにより、より詳細な検討をしていくものと考えられる。記して、今後の課題としたい。

なお、変数間の依存関係の強さを線の太さで表現するネットワーク図（ストレングス図）を作成することも可能であり、例えば、全世帯（農林漁家を含まない）の16費目レベルの非巡回有向グラフを作成すると、図6のようなになる。これにより、複雑なネットワーク図であっても主要な影響関係はかなり簡単になる可能性がある。

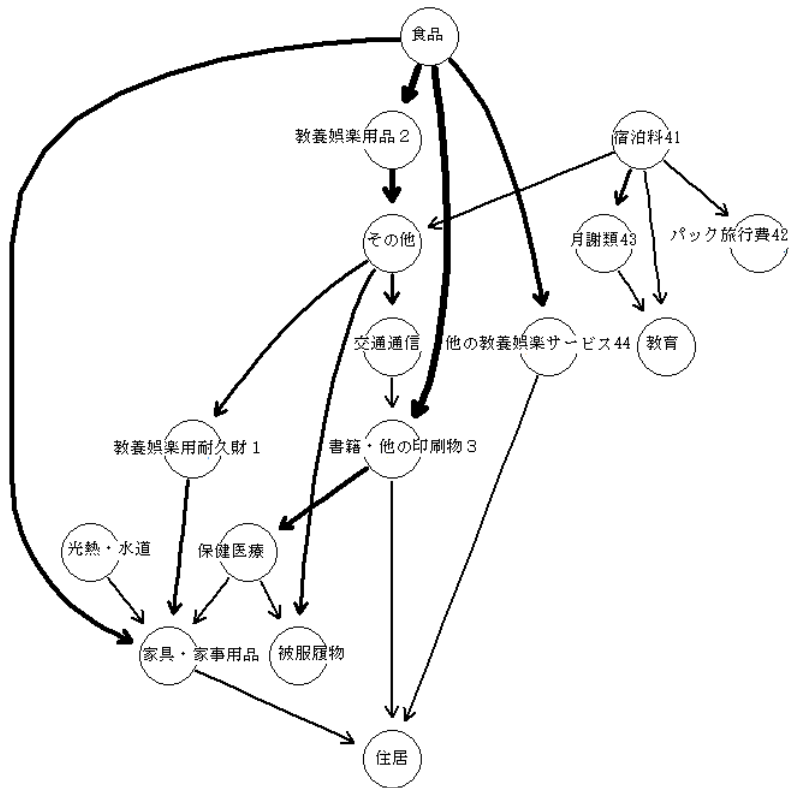


図6 ストレングス図：全世帯（農林漁家を除く）、iii）16費目分類

注

- *）本稿は科学研究費補助金基盤研究（B）「周期統計調査のマイクロデータによるコーホート分析－文化需要の実証分析－」（研究代表者：勝浦正樹）（課題番号：23330073）の研究成果の一部である。
- 1）たとえば、10 費目の、所得と価格を変数とする簡単なモデルを考える。このような簡単なモデルであっても、需要の加法性、同次性、対称性といった性質を用いパラメータに制約を掛けその数を減じたとしても、推定すべきパラメータは定数項 9 個、所得項 9 個、価格項 45 個の合計 63 個にのぼる。なお、これらの需要制約が本来検証無しに課してよい制約ばかりでないことも明らかである。
 - 2）これらの問いに答えることができれば、すなわち文化需要のデータを的確に説明するモデルを策定することができれば、それを用いて現状を把握し、今後の予測あるいは文化政策のインプリケーションを明らかにし、文化振興政策への提言を行うことができる。
 - 3）本節は、主に Pearl (2000)、鈴木 (2009) に基づく。
 - 4）連続型確率変数の場合、基本的に、多変量正規分布 $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ を仮定する（ $\boldsymbol{\mu}$ ：平均ベクトル、 $\boldsymbol{\Sigma}$ ：分散共分散行列）。
 - 5）本節は、主に Scutari (2010)、Højsgaard et al. (2012)、Wang et al. (2007) に基づく。
 - 6）ただし、学生の単身世帯、外国人世帯、その他（賄い付きの同居人がいる等で）捕捉の難しい世帯を除いている。
 - 7）これらの特徴は、期間中大きな変動はなかった。
 - 8）教養娯楽サービスのみが細目に対応している。

参考文献

- Højsgaard, S. D. Edwards, and S. Lauritzen (2012). *Graphical Models with R*. Springer.
- Katsuura, M. (2012). Lead-lag relationship between household cultural expenditures and business cycles, *Journal of Cultural Economics*, 36(1), 49-65.
- 勝浦正樹 (2012). 「社会生活基本調査のマイクロデータの再集計結果を用いた文化芸術活動の実証的研究：教育と所得水準を中心として」、浜田道夫・古隅弘樹編 (2012) 『文化経済学とコンピュータサイエンス』兵庫県立大学政策科学研究所、79-113.
- 永山貞則、勝浦正樹、衛藤英達 (2010). 『ワーク・ライフ・バランスと日本人の生活行動』日本統計協会.
- Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*, Cambridge University Press. (黒木学訳『統計的因果推論 モデル・推論・推測』共立出版.)
- Pourret, O., P. Naïm, and B. Marcot (2008). *Bayesian Networks: A Practical Guide to Applications*. Wiley.
- R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2012), ISBN 3-900051-07-0, <http://www.R-project.org/>.
- 西郷浩 (2012). 「社会生活基本調査からみた美術鑑賞行動の特徴」、浜田道夫・古隅弘樹編 (2012) 『文化経済学とコンピュータサイエンス』兵庫県立大学政策科学研究所、57-77.
- Scutari, M. (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, 35(3), <http://www.jstatsoft.org/>.
- Scutari, M. (2012). Package 'bnlearn'. <http://cran.r-project.org/>.
- 鈴木讓 (2009). 『ベイジアンネットワーク入門－確率的知識情報処理の基礎』培風館.
- Wang, M., Z. Chen, and S. Cloutier (2007). A hybrid Bayesian network learning method for constructing gene networks. *Computational Biology and Chemistry* 31, 361-372.

