

# 需要要因分解モデルによる予測と 多面的な顧客理解への活用

勝又壮太郎, 糸久 正人

## 1. はじめに

マーケティングにおける需要予測は、市場の構造を細分化することによって、高い精度を実現させてきた。需要を予測できる変数を組み込んだ回帰モデルから、異質な市場を仮定した分析（マーケット・セグメンテーション[13]）を可能にした混合分布モデルなど、多くの研究が蓄積されている。さらに、近年では、One to One マーケティング[9]に代表される個人単位の分析が注目され、それを実現する手法としてマルコフ連鎖モンテカルロ（Markov Chain Monte-Carlo; MCMC）法によって推定を行う階層ベイズモデル[10][11]が普及するようになった。階層ベイズモデルは高い予測精度を持つとともに非常に柔軟な拡張可能性を有し、多くの応用モデルが提案され[5]、マーケティング分析における標準的なツールとなりつつある。

しかし、階層ベイズモデルによって、個人単位のモデルを構築しても、複数の要因の結果から観測された値（意思決定）をそのまま利用するだけでは、顧客理解が進まず、適切なプロモーションを実施することは難しい。例えば顧客Aが製品Bを購入しなかったとき、顧客Aが製品Bを購入しなかったのか、購買したが他店で入手したのか、自社の店舗で観測される値は同じ「購買せず」だが、顧客の意思決定は大きく異なる。他店で入手したのであれば、自店に誘導するプロモーションが必要であり、そもそも購買をしなかったのであれば、購買意欲を喚起するプロモーションが必要となる。顧客Aが製品Bを自店舗で購入するという意思決定には、顧客Aが製品Bを購入するかどうかという意思決定の要因と自店舗に訪問するかどうかという意思決定の要因がある。顧客の行動が観測されたとき、意思決定の要因に分解して推定することが望ましい。個人単位の分析から、さらに細かい、個人の中の、意思決定の要因まで踏み込んだ階層ベイズモデルの構築が必要となるのである。

このような背景を踏まえて、本研究では、需要にいたる意思決定を要因に分解し、顧客の需要構造を推定する「分解モデル」を構築し、その有用性を検討することを目的とする。特に、本研究では次の3点を考慮して分析を行う。第1に、自店舗（自社）の保持しているデータを用いての推定である。他店での購買データが手に入ることはまれであり、ほとんどの企業は自らが持っているデータから顧客の傾向を推定せざるを得ない。入手可能なデータを最大限活用することのできるモデルを構築する。第2に、推定可能なモデルの構築である。通常、直接観測できない顧客の需要構造を推定することは困難である。そこで、意思決定の要因を分解する際、制約を与えて推定可能なモデルを定式化する。第3に、欠損補完のアルゴリズムである。観測できないデータ部分の欠損を扱う手法は難しいが、MCMC法によって確率的に欠損値を補完することで、適切な推定を行うことができる。

2節以降では、意思決定の分解とモデルの定式化を行い、推定方法を概説した後、実際の世帯ごとの食材利用データを分析対象として、推定と考察を行う。

## 2. モデル概説

本節では、まず、意思決定を要因に分解する分解モデルについて、その構築過程を解説する。その後、分解の有用性を検証するための対立モデルとして、要因に分解しない合成モデルを構築する。

### 2.1 分解モデル

#### 2.1.1 需要の分解

世帯  $h$  において時間  $t$  に食材  $j$  の需要 ( $d_{htj}$ ) が観測

かつまた そうたろう, いとひさ まさと  
東京大学 大学院経済学研究科  
〒113-0033 文京区本郷7-3-1  
受付 09.7.24 採択 09.11.10

されたとき、需要は次の2つの要因に分解することができる。

- 内食の意思決定 ( $y_{ht}$ )：世帯  $h$  が家で食事をするか外食するか
- 食材利用の意思決定 ( $z_{htj}$ )：世帯  $h$  が食材  $j$  を利用するかしないか

このとき、食材  $j$  が需要される確率  $\Pr(d_{htj}=1)$  は、家で食事をする事象  $y_{ht}=1$  と食材  $j$  を利用する事象  $z_{htj}=1$  がともに実現したと考えると、次のような同時確率で表すことができる。

$$\Pr(d_{htj}=1)=\Pr(y_{ht}=1, z_{htj}=1) \quad (1)$$

単純に食材を利用したかどうかで推定をするときは、要因が合成された状態  $d_{htj}$  が被説明変数として組み込まれることが一般的である。しかしながら、要因が異なるのであれば、別々に推定することが望ましい。ただし、別々に推定するためには、次の問題を解決しなければならない。

上記の同時確率を条件付き確率に分解すると、 $\Pr(d_{htj}=1)=\Pr(y_{ht}=1)\Pr(z_{htj}=1|y_{ht}=1)$  と表現することができる。ここで、内食確率  $\Pr(y_{ht}=1)$  および外食確率  $\Pr(y_{ht}=0)$  は観測可能であるが、食材の利用は常に観測可能というわけではない。世帯  $h$  が内食をしたときは観測できる、すなわち、 $\Pr(z_{htj}=1|y_{ht}=1)$  および  $\Pr(z_{htj}=0|y_{ht}=1)$  は観測できるが、世帯  $h$  が外食をしたときには、外食先でどの食材を食べたのかは観測できない。したがって、 $\Pr(z_{htj}=1|y_{ht}=0)$  および  $\Pr(z_{htj}=0|y_{ht}=0)$  が観測不可能となる。そこで、内食・外食の意思決定と食材利用の意思決定が独立である、という仮定を加える。これにより、需要確率を、内食確率と無条件食材利用確率との単純な積で表現することができる。

$$\Pr(d_{htj}=1)=\Pr(y_{ht}=1)\Pr(z_{htj}=1) \quad (2)$$

この仮定をおくことにより、観測不可能な値を補完することが可能となる。詳細は2.1.2節において解説する。

### 2.1.2 モデルの定式化

内食確率  $\Pr(y_{ht}=1)$  および、食材  $j \in \{1, \dots, J\}$  の利用確率  $\Pr(z_{htj}=1)$  は、0か1の利用履歴から推定される。すなわち、このモデルは、2項プロビットモデルとなる。 $y_{ht}$  および  $z_{ht}$  をまとめてモデル化を行うと、 $J+1$  変数のプロビットモデルとなる。ただし、 $z_{ht}$  は、外食した日 ( $y_{ht}=0$  が観測された日) は欠損となる。この欠損を補完するために、本研究では推定手法としてMCMC法を採用し、確率的に値を補完する。

次に、定式化のために潜在変数を導入する。これは、データ拡大法[12]を応用した文献[1]によって提案された方式である。まず、内食をするとき ( $y_{ht}=1$  のとき)、 $y_{ht}^* > 0$  となる潜在変数を定義する。また、食材  $j$  を利用するとき ( $z_{htj}=1$  のとき)、潜在変数  $z_{htj}^* > 0$  となる。 $z_{ht}=(z_{ht1}, \dots, z_{htj})'$  とおき、潜在変数には次のような関係を仮定する。

$$\begin{pmatrix} z_{ht}^* \\ y_{ht}^* \end{pmatrix} = B_h x_{ht} + v_{ht}, v_{ht} \sim \mathcal{N}(0, \Sigma) \quad (3)$$

$B_h$  は  $(J+1) \times K$  のパラメータであり、 $x_h$  は  $K$  次元の説明変数ベクトルである。 $\mathcal{N}(a, b)$  は、平均  $a$ 、分散  $b$  の正規分布を示す。このモデルでは、内食傾向や食材利用傾向が、世帯ごとに異なると仮定している、すなわち、個人異質性を考慮したモデルとなる。世帯の異質性は、パラメータ  $B_h$  で表現されている。

ここで、 $\Sigma$  には特殊な制約がかかる。まず、内食確率と食材利用確率が独立なので、 $y_{ht}$  と  $z_{ht}$  は独立となるため、 $\Sigma$  は次のような構造になる。

$$\Sigma = \begin{pmatrix} V & 0 \\ 0 & 1 \end{pmatrix} \quad (4)$$

$0$  は  $J \times 1$  の零行列である。多変量プロビットモデルの枠組みで定式化するので、多変量正規分布の分散項の一部  $V$  は、対角項が1で、非対角項が  $[-1, 1]$  の範囲を取り、なおかつ行列そのものが正定値であるという条件を満たす必要がある[3]。サンプリング方法をみると、文献[3][7]はメトロポリス・ヘイスティングス(M-H)法によって推定を行い、文献[4]は制約なしの乱数を発生させ、事後的に値を調整することで適切なサンプルを採取している。本研究では、文献[7]によって提案された、M-H法によるサンプリングを採用した。

潜在変数は、上述の条件を満たすようにサンプリングを行うと、次の分布からのサンプルとなる。まず、 $y_{ht}^*$  は、 $\mu_{ht} = x_{ht}' B_h$  とおくと、以下の方法でサンプルを採取する。

$$y_{ht}^* \sim \begin{cases} \mathcal{TN}_{(0, \infty)}(\mu_{ht, J+1}, 1) & \text{if } y_{ht}=1 \\ \mathcal{TN}_{(-\infty, 0]}(\mu_{ht, J+1}, 1) & \text{if } y_{ht}=0 \end{cases} \quad (5)$$

ここで、 $\mathcal{TN}_{(a, b)}(m, v)$  は、値域  $(a, b)$ 、平均  $m$ 、分散  $v$  の切断正規分布である。

また、食材利用に関する潜在変数は、 $y_{ht}=1$  のときは  $z_{htj}$  を観測することができる。まず、 $y_{ht}=1$  のときは、通常のプロビットモデルと同様の方法でサンプルを採取することができる。 $\mu_{ht} = B_h x_{ht}$  とおくと、潜在変数は次のようにサンプリングする。

$$z_{htj}^* \sim \begin{cases} \mathcal{TN}_{(0,\infty)}(m_{htj}, s_j) & \text{if } z_{htj}=1 \\ \mathcal{TN}_{(-\infty,0]}(m_{htj}, s_j) & \text{if } z_{htj}=0 \end{cases} \quad (6)$$

ここで、 $m_{htj}$  および  $s_j$  は次のような値となる。

$$m_{htj} = \mu_{htj} + \sum_{i,j} \Sigma_{-j,-j}^{-1} (z_{hi,-j}^* - \mu_{hi,-j}) \quad (7)$$

$$s_j = \Sigma_{jj} - \sum_{i,j} \Sigma_{-j,-j}^{-1} \Sigma_{-j,j} \quad (8)$$

$z_{hi,-j}^*$  および  $\mu_{hi,-j}$  はそれぞれ、 $z_{hi}^*$ 、 $\mu_{hi}$  から第  $j$  要素を除いたベクトルである。同様に、 $\Sigma_{-j,-j}$  は行列  $\Sigma$  の第  $j$  行を抜き出したベクトルから、さらに第  $j$  要素を除いたベクトルである。 $\Sigma_{-j,-j}$  は、行列  $\Sigma$  の第  $j$  行と第  $j$  列を除いた行列となる。

しかしながら、 $y_{ht}=0$  のときは、 $z_{ht}$  は欠損となり、食材利用に関して情報が得られない。ここで、 $z_{htj}^*|y_{ht}=1$  と  $z_{htj}^*|y_{ht}=0$  が同じ分布に従っている、すなわち、 $y_{ht}$  と  $z_{htj}$  が独立という仮定があれば、 $z_{htj}$  が観測されたときと同様の平均と分散の値を用いて、確率的に  $z_{htj}^*$  を補完することができる。内食確率と食材利用確率間の独立を仮定することで、 $y_{ht}=0$  のとき、上述の式(6)と同様の  $m_{htj}$  および  $s_j$  を用いて、 $z_{htj}^*$  は、次の分布からサンプルを得ることができる。

$$z_{htj}^* \sim \mathcal{N}(m_{htj}, s_j) \quad (9)$$

独立でなければ、場合分けを行い、 $z_{htj}^*|y_{ht}=1 \sim \mathcal{N}(m_{htj}^1, s_j^1)$  および  $z_{htj}^*|y_{ht}=0 \sim \mathcal{N}(m_{htj}^0, s_j^0)$  とおく必要がある。後者の平均  $m_{htj}^0$  と分散  $s_j^0$  を何らかの方法で推定する必要があるが、被説明変数となる  $z_{htj}$  が観測できないので、推定は困難である。

潜在変数を推定する式(3)に加えて、世帯の人数や収入などの説明変数から世帯  $h$  の個人属性を説明する項を以下のように定義する。

$$\text{vec}(B_h) = \Delta w_h + \eta_h, \eta_h \sim \mathcal{N}(0, \Psi) \quad (10)$$

$\text{vec}(\cdot)$  はベクトルオペレータである。サイズ  $m \times n$  の行列  $A$  があるとき、 $\text{vec}(A)$  は、行列  $A$  の列を縦に並べた  $mn$  次元ベクトルとなる。これは、式(3)の係数  $B_h$  が、 $D$  次元ベクトルのデモグラフィック変数  $w_h$  とサイズ  $(J+1)K \times D$  となる係数  $\Delta$  によって説明できると仮定するモデルである。この式(3)および式(10)を併せて、階層ベイズモデルとして定式化する。

$$\begin{cases} \begin{pmatrix} z_{ht}^* \\ y_{ht}^* \end{pmatrix} = B_h x_{ht} + v_{ht}, v_{ht} \sim \mathcal{N}(0, \Sigma) \\ \text{vec}(B_h) = \Delta w_h + \eta_h, \eta_h \sim \mathcal{N}(0, \Psi) \end{cases} \quad (11)$$

ここで、潜在変数  $z_{ht}^*$ 、 $y_{ht}^*$  やパラメータ  $B_h$  は、ギブスサンプリングによって採取が可能であり、文献[10]などでサンプリング方法が提示されている。また、 $\Delta$  と  $\Phi$  についても、それぞれ行列正規分布とウィシャート分布から採取が可能である。

## 2.2 合成モデル

対立モデルとして、需要  $d_{htj}$  を直接推定するモデルを構築する。このモデルでは、内食をしなかったために  $d_{htj}=0$  が観測されたケースと、内食したが食材  $j$  を利用しなかったために  $d_{htj}=0$  が観測されたケースは区別しない。

$$\begin{cases} d_{ht}^* = \Gamma_h x_{ht} + u_{ht}, u_{ht} \sim \mathcal{N}(0, \Theta) \\ \text{vec}(\Gamma_h) = \Lambda w_h + \xi_h, \xi_h \sim \mathcal{N}(0, \Phi) \end{cases} \quad (12)$$

このモデルも分解モデルと同様に、個人ごとの係数を推定する階層ベイズモデルとなる。 $\Gamma_h$  は  $J \times K$ 、 $\Lambda_h$  は  $JK \times D$  のパラメータであり、 $x_h$  および  $w_h$  は2.1.2節と同様の説明変数である。階層ベイズモデルは、非常に高い予測精度を持っていることが文献[5]などで確認されており、このモデルも高い予測精度を持つことが予想される。

このモデルは、通常の変量プロビットモデルと同様の構造であるため、潜在変数  $d_{htj}^*$  は次の分布からのサンプルとなる。

$$d_{htj}^* \sim \begin{cases} \mathcal{TN}_{(0,\infty)}(m_{htj}, s_j) & \text{if } d_{htj}=1 \\ \mathcal{TN}_{(-\infty,0]}(m_{htj}, s_j) & \text{if } d_{htj}=0 \end{cases} \quad (13)$$

$m_{htj}$  および  $s_j$  は、 $z_{htj}^* = d_{htj}^*$ 、 $\mu_{htj} = \Gamma_h x_{ht}$  と置き換えた式(7)、式(8)から得ることができる。

## 3. 実データの解析

2節で解説した2種類のモデルを用いて、実際データを分析する。分析に用いたデータは、経営科学系研究部会連合協議会主催の、平成20年度データ解析コンペティションにて提供された、世帯ごとの食材利用履歴である。収集期間は一年間(365日)、世帯数は合計194世帯である。

### 3.1 分析対象

本研究では、194世帯の夕食における内食と食材利用の傾向を推定する。分析期間としては、1月1日から3月31日までの90日間を学習期間としてモデルの推定を行い、4月1日から4月30日までの30日間を検証期間としてモデルの予測精度の検証に用いる。また、分析対象とする食品として、学習期間中に利用頻度が高かった上位10種類とする。具体的には、以下の食品となる。「コシヒカリ」と「その他白米」などの分類基準は、提供元のデータにしたがっている。

- 濃口醤油
- だしの素
- 長ねぎ
- 鶏卵
- にんじん
- コシヒカリ
- 緑茶
- 天然塩

- たまねぎ
- その他白米

### 3.2 説明変数

個人異質性に掛る変数  $x_{ht}$  および、世帯の属性  $w_h$  は、次のような変数を用いる。

$$x_{ht} = \begin{pmatrix} \text{切片} \\ \text{夫の休日 (D)} \end{pmatrix} \quad (14)$$

夫の休日ダミーは、その日が夫の休日なら 1 をとり、そうでなければ 0 をとるダミー変数である。

$$w_h = \begin{pmatrix} \text{切片} \\ \text{収入} \\ \text{専業主婦 (D)} \\ \log(\text{妻の年齢}) \end{pmatrix} \quad (15)$$

専業主婦はダミー変数である。分解モデルと合成モデルは、ともに同じ説明変数  $x_{ht}$ ,  $w_h$  を用いて推定を行う。

### 3.3 シミュレーション設定

初期値として、正規分布に従うパラメータは零行列または零ベクトル、ウィシャート分布に従うパラメータと多変量プロビットモデルの分散項  $\Sigma$ ,  $\Lambda$  は単位行列を用いた。また、それぞれのモデルは、5,000 回の稼働検査期間の後に 20,000 回を計算し、その 4 回に 1 回、合計 5,000 個をサンプルとして採取している。事前分布は付録に掲載している。

## 4. 分析結果と活用

### 4.1 モデルの当てはまり

まず、モデルそのものの当てはまりの良さを検討するために、周辺尤度を導出し、ベイズファクター (BF) による比較を行う。本研究では、周辺尤度の計算は文献[8]によるアルゴリズムを用いている。また、BF を計算するために、提案モデルに対して、階層化せず切片のみで食材利用確率 (分解モデルではこれに加えて内食確率) を推定した Null モデルから、周辺尤度を計算している。ただし、分解モデルと合成モデルは被説明変数の構造が異なるので、2つのモデル間の当てはまりを比較することはできない。本節では、各々のモデルについて、その Null モデルと比較して当てはまりの向上を検証する。

表 1 は対数周辺尤度と、対数 BF<sup>1</sup> である。対数 BF

<sup>1</sup> ベイズファクター (BF)  $B_{10}$  は、提案モデルの周辺尤度を  $M_1$ , Null モデルの周辺尤度を  $M_0$  とすると、

$$B_{10} = M_1 / M_0 \quad (16)$$

から求められるので、対数 BF は次の式から求める。

$$\log B_{10} = \log M_1 - \log M_0 \quad (17)$$

表 1 対数周辺尤度

	分解	合成
Null モデル	-552.3	-498.3
提案モデル	-443.6	-400.2
対数 BF	108.7	98.1

は、文献[6]の基準によると、5 を超えていれば Null モデルと差がないという仮説を棄却する証拠が特に強いとしているので、分解・合成双方のモデルともにモデルの当てはまりは十分向上しているといえる。

### 4.2 予測精度の比較

ここでは、得られた係数を用いて、検証期間の予測を行う。まず、それぞれのモデルにおける需要確率を計算する。仮定から、需要確率は内食確率と食材利用確率の積となるので、それぞれの確率 (周辺確率) を計算し、その積をとる<sup>2</sup>。計算にはパラメータ  $B_h$ ,  $\Gamma_h$  の事後平均  $\bar{B}_h$  および  $\bar{\Gamma}_h$  を用いた。分解モデルでは、 $\bar{\mu}_{ht} = \bar{B}_h x_{ht}$ , 合成モデルでは  $\bar{\mu}_{ht} = \bar{\Gamma}_h x_{ht}$  である。

$$\hat{\text{Pr}}(y_{ht}=1) = \int_{-\infty}^{\bar{\mu}_{ht, t+1}} \phi(z) dz \quad (18)$$

$$\hat{\text{Pr}}(z_{htj}=1) = \int_{-\infty}^{\bar{\mu}_{htj}} \phi(z) dz \quad (19)$$

$\phi(\cdot)$  は標準正規分布の密度関数である。得られた内食確率と食材利用確率から、分解モデルの予測需要確率  $\hat{d}_{htj}$  は式(2)で与えた定義通り、次のように計算する。

$$\hat{\text{Pr}}(d_{htj}=1) = \hat{\text{Pr}}(y_{ht}=1) \hat{\text{Pr}}(z_{htj}=1) \quad (20)$$

また、合成モデルにおける需要確率は、事後平均  $\bar{\Gamma}_h$  を用いて得られた  $\bar{\mu}_{ht} = \bar{\Gamma}_h x_{ht}$  から、次のように計算する。

$$\hat{\text{Pr}}(d_{htj}=1) = \int_{-\infty}^{\bar{\mu}_{htj}} \phi(z) dz \quad (21)$$

こうして得られた需要の予測値から、それぞれのモデルの予測精度を検証する。検証手法として、ヒットレートと ROC 曲線下面積の 2 つを用いる。検証期間は 4 月 1 日から 4 月 30 日までの 30 日間、分析対象人数は 194 人なので、 $30 \times 194 = 5,820$  件の予測機会がある。

ヒットレートは、閾値を 0.5 とおき、食材ごとに予測\実現テーブルの対角項の和を計算した。

表 2 の例では、予測において時間  $t$  に食材  $j$  を購入すると判定され ( $\hat{\text{Pr}}(d_{htj}=1) \geq 0.5$ )、かつ実際に購入した人数は  $a$  人である。同様に、予測において購入

<sup>2</sup> 食材利用確率同士は独立ではないので、「食材 A と食材 B を同時に利用する確率」を求めたいときには、多変量正規分布を積分して確率を計算する必要がある。

表2 予測\実現テーブル

予測\実現	購買	購買せず
購買	a	b
購買せず	c	d

表3 予測結果

食材	ヒットレート		ROC 曲線下面積	
	分解	合成	分解	合成
濃口醤油	0.693	0.693	0.649	0.649
長ねぎ	0.773	0.774	0.654	0.652
にんじん	0.753	0.751	0.629	0.627
緑茶	0.855	0.855	0.812	0.811
たまねぎ	0.731	0.731	0.617	0.617
だしの素	0.798	0.797	0.727	0.726
鶏卵	0.799	0.799	0.642	0.642
コシヒカリ	0.824	0.823	0.777	0.776
天然塩	0.826	0.825	0.757	0.757
その他白米	0.836	0.836	0.782	0.783
平均	0.789	0.788	0.705	0.704

されないと判定され、実際に購買されなかった人数は  $d$  人である。合計人数を  $H$  人とおくと、ヒットレートは  $(a+d)/H$  となる。

ROC 曲線下面積も同様に、食材ごとに検証を行っている。ROC 曲線下面積は、 $[0,1]$ の値をとり、0.5以上ならば予測モデルとして有用性があると判定することができる。詳細は文献[2]を参照のこと。

表3は、予測結果である。これを見ると、双方のモデルともにほぼ同等ではあるが、わずかに分解モデルの方が予測精度が高かった。また、ROC 曲線下面積を見ると、どちらのモデルも分析対象としたすべての食材の予測において0.5を超えているので、予測モデルとしての有用性はあるといえる。ただし、分解モデルは4.3節以降の活用において、多様な示唆を与えてくれる。

### 4.3 係数の考察

推定した上層の係数(事後平均)は表4の通りである。90%最高事後密度区間 (HPD) に0を含んでいなければ「\*」、95%HPDに0を含んでいなければ「\*\*」を添えている。

ここから、内食と食材利用の影響関係を考察することができる。下の合成モデルを見ると、夫の休日(夫休)には、一般的な世帯では多くの食材において利用確率が下がり、妻の年齢が高い世帯では多くの食材において利用確率が高まるという結果が得られている。しかしながら、上の分解モデルを見ると、変化の多くは内食確率に吸収されていることがわかる。すなわち、夫の休日には、一般的な世帯は外食し、妻の年齢が高い世帯は内食するという傾向があるために、観測され

表4 分解モデルの  $\Delta$  と合成モデルの  $\Lambda$

分解モデル ( $\Delta$ )	切片	収入	専業主婦	妻年齢
切片-濃口醤油	-1.65*	-0.02	0.07	0.44
切片-長ねぎ	-1.88**	-0.01	0.05	0.36*
切片-にんじん	0.58	0.03	0.04	-0.31*
切片-緑茶	-12.99**	-0.06	0.27	3.18**
切片-たまねぎ	-0.19	0.03	-0.07	-0.10
切片-だしの素	1.57	0.08	-0.02	-0.71**
切片-鶏卵	0.04	0.04	-0.04	-0.22
切片-コシヒカリ	-0.98	0.11	0.09	-0.21
切片-天然塩	-1.95	0.07	0.12	0.13
切片-その他白米	-2.49	-0.12	0.03	0.36
切片-内食	-0.40	0.05	0.10	0.37*
夫休-濃口醤油	-1.26**	-0.01	-0.07	0.33**
夫休-長ねぎ	0.47	0.05**	-0.02	-0.14
夫休-にんじん	-0.91*	-0.03	-0.04	0.24*
夫休-緑茶	1.23**	0.07**	-0.22**	-0.34**
夫休-たまねぎ	-0.92*	-0.03	-0.04	0.25*
夫休-だしの素	-0.07	-0.03	-0.02	0.01
夫休-鶏卵	-1.18**	-0.03	0.04	0.32**
夫休-コシヒカリ	-1.06*	0.01	-0.08	0.29*
夫休-天然塩	-0.87*	0.06**	-0.10	0.16
夫休-その他白米	0.00	0.08*	-0.18**	-0.03
夫休-内食	-2.35**	-0.02	-0.09	0.54**

合成モデル ( $\Lambda$ )	切片	収入	専業主婦	妻年齢
切片-濃口醤油	-1.97**	0.00	0.08	0.45*
切片-長ねぎ	-2.18**	0.00	0.07	0.39**
切片-にんじん	-0.02	0.04	0.05	-0.20
切片-緑茶	-11.74**	-0.03	0.19	2.77**
切片-たまねぎ	-0.74	0.04	-0.05	-0.01
切片-だしの素	0.98	0.09	0.01	-0.60*
切片-鶏卵	-0.42	0.05*	-0.02	-0.14
切片-コシヒカリ	-1.44	0.10	0.08	-0.11
切片-天然塩	-2.30*	0.08	0.15	0.18
切片-その他白米	-3.00	-0.10	0.11	0.42
夫休-濃口醤油	-2.17**	-0.01	-0.08	0.53**
夫休-長ねぎ	-0.73	0.05*	-0.03	0.15
夫休-にんじん	-1.74**	-0.03	-0.03	0.43**
夫休-緑茶	0.74	0.01	-0.14	-0.19
夫休-たまねぎ	-1.49**	-0.04*	-0.02	0.38**
夫休-だしの素	-1.38**	-0.03	-0.02	0.34**
夫休-鶏卵	-2.02**	-0.03	0.01	0.52**
夫休-コシヒカリ	-2.04**	-0.01	-0.04	0.54**
夫休-天然塩	-1.72**	0.03	-0.16**	0.43**
夫休-その他白米	-0.21	0.09**	-0.22**	0.03

た需要が変化するといえる。夫の休日における需要確率の変化は、内食確率の大きな変化が影響を与えているのであり、個々の食材の利用確率は、合成モデルで推定される程大きく変化しているわけではない。合成モデルではこの差を捉えることができていないのである。予測精度は双方のモデルともに同水準だったが、分解モデルの方がより詳細に顧客の行動を表現できているといえる。

### 4.4 プロモーション施策の差別化

分解モデルは要因を分けているため、需要確率を2次元的に把握することができる。検証期間の内食確率および食材利用確率の、個人ごとの平均を計算し、それをプロットしたものが図1である。

点線は内食確率×食材利用確率から計算される需要確率を、確率0.1ごとに引いたものである。最も左下に位置する点線が、需要確率=0.1の線となる。合成モデルでは、この需要確率のみが推定されるが、実際

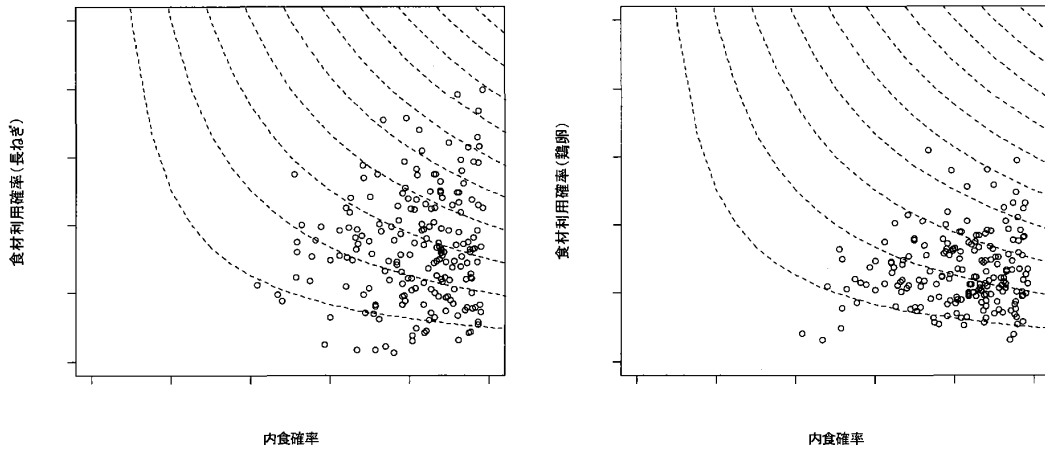


図1 内食確率と食材利用確率：長ねぎ（左），鶏卵（右）

に分解モデルによって内食確率と食材利用確率を計算すると、世帯ごとのばらつきがかなり大きいことが分かる。この図で左側にある世帯は「内食確率が低いために需要が低く」、図の下にある世帯は「食材利用確率が低いために需要が低い」といえる。この差異を見れば分かるように、同じ需要確率の世帯であっても異なる需要喚起施策が必要であるといえる。内食確率を上げることで需要を伸ばすことのできる世帯なら、調理済み惣菜など、家で簡単に食事を作れるような製品とともに食材を提供することで、食材の需要を喚起することができると考えられる。また、食材利用確率が低い世帯の需要を伸ばすなら、その食材を利用したレシピなどの提案をすることで需要を引き上げることができるだろう。世帯ごとの需要構造を正確に把握し、世帯の状態に合ったプロモーション施策を考えることが重要である。

## 5. 結論

本研究における分解モデルは、マーケティング分析において盛んに用いられている階層ベイズモデルを拡張し、MCMC法による欠損補完アルゴリズムを応用することで、需要を要因に分解して推定することを可能とした。分解モデルは、同時に推定を行った合成モデルと比較すると、より深い知見を得ることができる。すなわち、顧客の需要構造を理解することで、効果的なプロモーション施策が可能となる。このような意味において、個人単位の分析からさらに細分化された意思決定の要因単位での分析は、マーケティングにおいて非常に有用であるといえる。

分解モデルは、さまざまな応用が考えられる。例えば、今回の研究で分析対象としたデータは世帯の食材

利用に関するものだが、スーパーマーケットの来店・購買行動に応用することで、自店で購買したのか、自店では購買しなかったのかという顧客の囲い込みの程度や離反傾向を推定することも可能となる。すなわち、それぞれ「店舗へのロイヤル度」と「消費性向」と捉えることもでき、いまプロモーションをかけようと計画している顧客が、店舗ロイヤル度は高いが消費性向が低い顧客なのか、消費性向は高いが店舗ロイヤル度が低い顧客なのかを判別することができる。

今後のモデルの拡張としては、時間変化の要素を考慮した、長期の予測が可能なモデルの構築を考慮することができる。本研究では30日間の予測を行ったが、さらに先の将来を予測するモデルの構築のためには、時間変化の要素を組み込む必要がある。長期予測が可能なモデルを構築することができれば、店舗の製品調達だけでなく、製品の生産計画においても利用価値の高いモデルになると考えられる。

### A. 事前分布

各パラメータの事前分布を以下に示す。以下、 $H$ は分析世帯数、 $J$ は推定する食材の種類、 $K$ は説明変数  $x_{ht}$  の次元、 $D$ は説明変数  $w_h$  の次元、 $I_K$ はサイズ  $K$  の単位行列である。

#### A.1 分解モデルの事前分布

$$\text{vec}(\Delta) \sim \mathcal{N}_{(J+1)KD}(\delta_0^*, S_\delta) \quad (22)$$

$$\Psi^{-1} \sim \mathcal{W}(\nu_\Psi, \Omega_\Psi) \quad (23)$$

ここで、 $\delta_0^*$ は  $(J+1)K$ 次元の零ベクトル、 $S_\delta = 100I_{(J+1)K}$ 、 $\nu_\Psi = H$ 、 $\Omega_\Psi = H \times I_{(J+1)K}$ である。

$\Sigma$ は、本文中で示したように、 $V$ をサンプリングする。文献[7]によって提案された、次のような事前分布を用いる。

$$\pi(\text{vec}^*(V)) \propto \exp\{-0.5(\text{vec}^*(V) - v_0^*)'\}$$

$$\times G(\text{vec}^*(V) - v_0^*) \quad (24)$$

$\text{vec}^*(V)$  は行列  $V$  の対角項を除く上三角要素を分解して縦に並べ、ベクトル化する変則型のベクトルオペレータとする。また、本研究では、 $v_0^*$  は零ベクトル、 $G$  には単位行列を用いた。

## A.2 合成モデルの事前分布

$$\text{vec}(\Lambda) \sim \mathcal{N}_{JKD}(\lambda_0^*, S_\lambda) \quad (25)$$

$$\Phi^{-1} \sim \mathcal{W}(\nu_\Phi, \Omega_\Phi) \quad (26)$$

ここで、 $\lambda_0^*$  は  $JK$  次元の零ベクトル、 $S_\lambda = 100I_{JK}$ 、 $\nu_\Phi = H$ 、 $\Omega_\Phi = H \times I_{JK}$  である。

$\Theta$  は、 $V$  と同様、次のような事前分布を用いる。

$$\pi(\text{vec}^*(\Theta)) \propto \exp\{-0.5(\text{vec}^*(\Theta) - \theta_0^*)' \times F(\text{vec}^*(\Theta) - \theta_0^*)\} \quad (27)$$

$\theta_0^*$  は零ベクトル、 $F$  には単位行列を用いた。

**謝辞** 本研究に関して、阿部誠先生（東京大学）から大変有用なコメントをいただきました。また、経営科学系研究部会連合協議会および企業の方からは貴重なデータを提供いただきました。この場をお借りして御礼申し上げます。

## 参考文献

- [1] Albert, J. H. and Chib, S.: "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of American Statistical Association*, Vol. 88, pp. 669-679, (1993).
- [2] Blattberg, R. C., Kim, B. and Neslin, S. A.: *Database Marketing: Analyzing and Managing Customers*, Springer (2008).
- [3] Chib, S. and Greenberg, E.: "Analysis of Multivariate Probit Models," *Biometrika*, Vol. 85, pp. 347-361, (1998).
- [4] Edwards, Y. and Allenby, G.: "Multivariate Analysis of Multiple Response Data," *Journal of Marketing Research*, Vol. 41, pp. 321-334, (2003).
- [5] 勝又壮太郎: "音楽 CD 購買における個人嗜好の推定," *オペレーションズ・リサーチ*, Vol. 52, pp. 725-731, (2007).
- [6] Kass, R. E. and Raftery, A. E.: "Bayes Factors," *Journal of the American Statistical Association*, Vol. 90, pp. 773-795, (1995).
- [7] Manchanda, P., Ansari, A. and Gupta, S.: "The Shopping Basket: A Model for Multicategory Purchase Incidence Decisionsters," *Marketing Science*, Vol. 18, pp. 95-114, (1999).
- [8] Newton, M. A. and Raftery, A. E.: "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, Series B*, Vol. 64, pp. 3-48, (1994).
- [9] Peppers, D. and Rogers, M.: *The One to One Future: Building Relationships One Customer at a Time*, New York: Currency Doubleday (1993).
- [10] Rossi, P. E., Allenby, G. and McCulloch, R.: *Bayesian Statistics and Marketing*, Wiley (2005).
- [11] Rossi, P. E., McCulloch, R. and Allenby, G.: "The Value of Purchase History data in target marketing," *Marketing Science*, Vol. 15, pp. 321-340, (1996).
- [12] Tanner, M. A. and Wong, W. H.: "The Calculation of Posterior Distributions by Data Augmentation," *Journal of American Statistical Association*, Vol. 82, pp. 528-540, (1987).
- [13] Wedel, M. and Kamakura, W.: *Market Segmentation: Conceptual and Methodological Foundations: 2nd Ed.*, Springer (2000).