

階層ベイズ条件付きプロビットモデルによる 消費者選択行動の解析

-個票データとマルチソースの時系列データを統合して-

平泰浩 戸田敬之 松村優哉 山口愛 米津了輔

2015年12月2日(オープンゼミ)

星野崇宏研究会 データ解析パート

研究背景 1-消費者異質性-

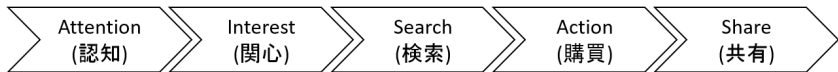


ビッグデータの蓄積

異質な消費者に対して行動を詳細に解析する必要性 **階層ベイズモデル**

研究背景 2-複数ソースの時系列データ-

消費者行動としての AISAS モデル



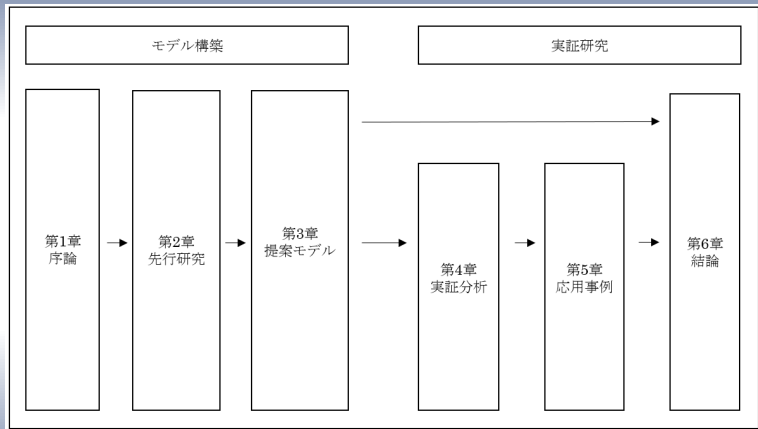
消費者発信型メディアを含めたマルチソースの時系列データ利用が重要

研究の目的

研究の目的

- 購買プロセスの変容を考慮した消費者一人ひとりの選択モデルを推定
- 結果から消費者異質性の中にある共通要因の解析
- 昨今の消費者選択行動を解明
- 提案したモデルを利用したマーケティングへの応用事例を紹介

研究の構成



Guadagni and Litle(1983)

- 消費者は前回選択行動に状態依存
- 離散選択モデルに消費者のブランドロイヤリティを組み込む

確率的効用

$$u_{ik} = v_{ik} + \varepsilon_{ik} \quad (1)$$

$$v_k^i = \sum_{j \in T_k} b_{jk} x_{jk}^i + \sum_{j \in T_C} b_{jk} x_{jk}^i \quad (2)$$

T_k は選択肢の持つ固有の特性、 T_C は価格などの共通のマーケティング変数

Guadagni and Litterle(1983)

インディケータ関数

$$y_k^i(n) = \begin{cases} 1 & \text{if } n \text{ 回目の消費者 } i \text{ の } k \text{ の選択} \\ 0 & \text{if otherwise} \end{cases} \quad (3)$$

$$x_k^i(n) = \alpha_b x_k^i(n-1) + (1 - \alpha_b) \times \begin{cases} 1 \\ 0 \end{cases} \quad (4)$$

Rossi, McCulloch and Allenby (1996)

消費者異質性を考慮した階層ベイズモデル

階層プロビットモデル

$$\mathbf{u}_{it} = \mathbf{X}_{it}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_{it} ; \boldsymbol{\varepsilon}_{it} \sim N(0, \Lambda), i = 1, \dots, I, t = 1, \dots, T \quad (5)$$

$$\boldsymbol{\beta}_i = \boldsymbol{\Delta}\mathbf{z}_i + \mathbf{v}_i ; \mathbf{v}_i \sim iid.N(0, V_{\beta}) \quad (6)$$

- \mathbf{X}_{it} … マーケティング変数の価格ベクトルなど
- $\boldsymbol{\beta}_i$ … 消費者の選択肢に対する感応度
- $\boldsymbol{\Delta}$ … 回帰係数行列
- \mathbf{v}_i … 観測されない異質性の分散を決めるパラメーター

Swait and Andrews(2003)

スキャナーパネルデータ (RP) とサーベイデータ (SP) を組み合わせたデータ強化法

スキャナーパネルデータへのランダム効用

$$u_i^{RP} = \beta_i^{RP} X_i^{RP} + \gamma Z_i + \varepsilon_i^{RP}, i \in C^{RP} \quad (7)$$

- X_i^{RP}, Z_i は属性ベクトル
- β^{RP}, γ はデータに適合したパラメーターのベクトル

Swait and Andrews(2003)

選択確率を表現する多項ロジットモデル

$$Pr_i^{RP} = \frac{\exp(\mu^{RP}(\beta_i^{RP} X_i^{RP} + \gamma Z_i))}{\sum_{i \in C^{RP}} \exp(\mu^{RP}(\beta_i^{RP} X_i^{RP} + \gamma Z_i))} \quad (8)$$

SP についても同様

ロジットモデルの問題点-I.I.A 特性-

選択枝が二つの場合

選択確率

$$Pr(y_{it} = A) = \frac{\exp(p_{Ait}\beta)}{\exp(p_{Ait}\beta) + \exp(p_{Bit}\beta)} \quad (9)$$

$$Pr(y_{it} = B) = \frac{\exp(p_{Bit}\beta)}{\exp(p_{Ait}\beta) + \exp(p_{Bit}\beta)} \quad (10)$$

オッズ比

$$\frac{Pr(y_{it} = A)}{Pr(y_{it} = B)} = \frac{\exp(p_{Ait}\beta)}{\exp(p_{Bit}\beta)} \quad (11)$$

ロジットモデルの問題点-I.I.A 特性-

選択肢が追加された場合

$$Pr(y_{it} = A) = \frac{\exp(p_{Ait}\beta)}{\exp(p_{Ait}\beta) + \exp(p_{Bit}\beta) + \exp(p_{Cit}\beta)} \quad (12)$$

$$Pr(y_{it} = B) = \frac{\exp(p_{Bit}\beta)}{\exp(p_{Ait}\beta) + \exp(p_{Bit}\beta) + \exp(p_{Cit}\beta)} \quad (13)$$

と変化するが、オッズ比は (11) のまま

論点

	Guadagni and Little	Rossi et al.	Swait and Andrews	本稿の提案モデル
ベースラインモデル	多項ロジットモデル	階層型多項 プロビットモデル	多項ロジットモデル	階層型条件付き プロビットモデル
推定方法	最尤法	MCMC法 (ベイズ推定)	最尤法	MCMC法 (ベイズ推定)
結論	ブランド ロイヤリティの考慮	消費者異質性の 考慮	複数データソースの 使用	左記のすべてを 考慮
問題点	I.I.A特性	変数選択を行わず	I.I.A特性	I.I.A特性の改善と 変数選択を行い 問題点を克服
ブランドロイヤリティの考慮	○	×	×	○
消費者異質性の考慮	×	○	×	○
データ強化	×	×	○	○

先行研究を補完するモデルを提案

モデルの仮定

- ① 消費者は過去の購買行動に状態依存し、ロイヤリティを持つブランドと持たないブランドとでは反応係数が異なる。
- ② 消費者一人ひとりの反応係数と切片項は異質性を持つが、異質性の中にデモグラフィック変数の共通要因を持つ。
- ③ 個票データと時系列データは時点とブランドが一致しているものは統合して利用可能である。

ブランドロイヤルティについて

消費者 i をロイヤリティのブランド $p(p \in A_{it})$ 別にセグメント化する

消費者のロイヤリティのブランド別のセグメント

$$i(p) = 1, \dots, N(p) \quad (14)$$

$$N = N(1) + \dots + N(H) \quad (15)$$

H は選択肢の数

効用関数を定義

マーケティング変数+時系列変数

$$U_{i(p)Ikt} = \alpha_{i(p)k}^0 + \sum_{m=1}^M \alpha_{i(p)I}^m x_{i(p)kt}^m + \sum_{l=1}^L \beta_{i(p)I}^l w_{kt}^l + e_{i(p)kt} \quad (16)$$

- 切片 $\alpha_{i(p)k}^0$ … 各消費者 $i(p)$ にとっての選択ブランド k 固有のブランド価値
- 変数 x_{ikt}^m … 各消費者が直面する $M(m = 1, \dots, M)$ 種類の固有のマーケティング変数
- 変数 w_{ht}^l … 時系列変化する個人レベルで紐づかない $L(l = 1, \dots, L)$ 種類の外部環境変数

ブランドロイヤルティ

インディケータ関数によるパラメーターの区別

$$I = \begin{cases} 1 & \text{if } k \neq p \\ 0 & \text{if } k = p \end{cases} \quad (17)$$

$$\alpha_{i(p)}^m = \alpha_{i(p)I}^m = \begin{cases} \alpha_{i(p)0}^m & \text{if } I = 0 \\ \alpha_{i(p)1}^m & \text{if } I = 1 \end{cases} \quad (18)$$

$$\beta_{i(p)}^l = \alpha_{i(p)I}^m = \begin{cases} \beta_{i(p)0}^l & \text{if } I = 0 \\ \beta_{i(p)1}^l & \text{if } I = 1 \end{cases} \quad (19)$$

パラメーター区別の例

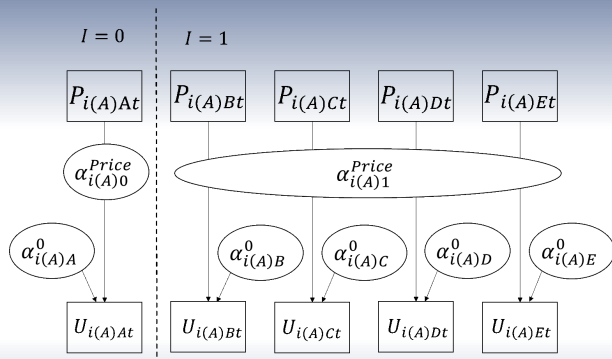


Figure: 各ブランドの価格変数への反応係数 $\alpha_{i(A)I}^{Price}$

反応係数

係数の階層構造

$$\alpha_{i(p)I}^m = \begin{cases} \mathbf{Z}_{i(p)}^{*\mathbf{T}} \Theta_{\alpha 0}^m + \eta_{\alpha_{i(p)0}^m} & m = 1, \dots, M \end{cases} \quad (20)$$

$$\beta_{i(p)I}^l = \begin{cases} \mathbf{Z}_{i(p)}^{*\mathbf{T}} \Theta_{\beta 0}^l + \eta_{\beta_{i(p)0}^l} & l = 1, \dots, L \end{cases} \quad (21)$$

- $\mathbf{Z}_{i(p)}^{*\mathbf{T}} = (1, z_{i(p)1}, \dots, z_{i(p)Q})$ はデモグラフィック変数行列 $Z_{i(p)}$ の先頭に定数項の為の 1 を置いた説明変数行列
- $\Theta_{\alpha 0}^m, \Theta_{\alpha 1}^m, \Theta_{\beta 0}^l, \Theta_{\beta 1}^l$ は回帰係数行列

ブランド固有の切片 (定数項)

切片の階層構造

$$\alpha_{i(p)k}^0 = \mathbf{Z}_{i(p)k}^T \Theta_{\alpha^0} + \eta_{\alpha_{i(p)k}^0} \quad (22)$$

$\mathbf{Z}_{i(p)k} = \left(0, 0, \dots, 1, 0, \dots, 0; \mathbf{Z}_{i(p)}^T \right)^T$ は、ブランド k 固有の定数項の為に k 番目は 1、 $h \neq k$ 番目は 0 をとる部分とデモグラフィック変数からなるブランド k に関する説明変数行列

ブランド選択確率

消費者 $i(p)$ が t 期にブランド k を選択する確率

$$\begin{aligned} & Pr(y_{i(p)t} = k | \boldsymbol{\mu}_{i(p)}, \Sigma) \\ = & Pr(\mathbf{u}_{i(p)t} \in R_{it}^{H-1} | \boldsymbol{\mu}_{i(p)}, \Sigma) \\ = & \int_{R_{it}^{H-1}} p(\mathbf{u}_{i(p)t} | \mathbf{X}_{i(p)t}, \boldsymbol{\mu}_{i(p)}, \Sigma) d\mathbf{u}_{i(p)t} \quad (23) \end{aligned}$$

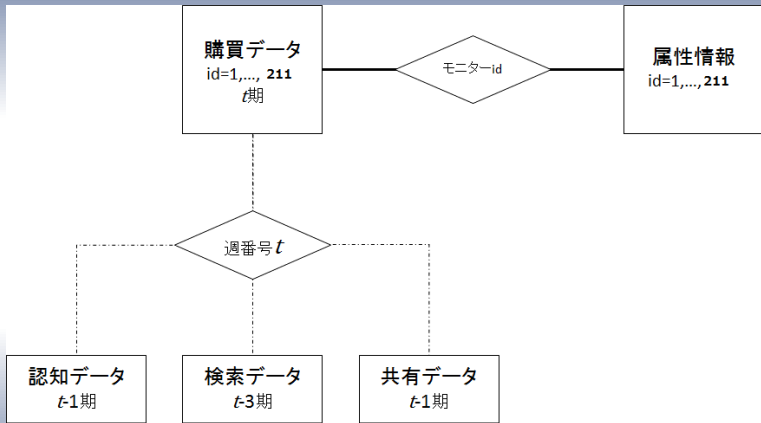
記号の説明

- $R_{i(p)t}^{H-1}$ … 消費者 $i(p)$ が t 期にブランド k を選択するための制約条件
- $u_{i(p)t}$ … 効用関数の式
- $X_{i(p)t}$ … 単位行列とマーケティング変数と時系列変数
- $\mu_{i(p)}$ … 切片と反応係数
- Σ … 誤差項 ($e_{i(p)kt}$) の従う分散

データについて

- 「第3回データビジネス創造コンテスト」で提供された、シャンプーの購買履歴データと AISAS モデルに応じた時系列データ
- ユニリーバ、資生堂、クラシエ、花王、P&G の5社
- 2014年10月1日～2015年3月31日
- 時点と購買メーカーが一致した時系列データと購買データを紐づけ
- 「2回以上連続で同一メーカーの商品を購入した消費者はそのメーカーにロイヤリティがある」と定義
- 対象5社にロイヤリティを持ち、期間中にスイッチした消費者211名を抽出

データの紐づけのイメージ



変数選択-個票データ-

- マーケティング変数は**購買価格**
- デモグラフィック変数は**年代**、**子供の有無**、**未既婚**、**世帯年収**、**世帯人数**、**性別**を抽出
- 多重共線性を回避するため**子供の有無**、**世帯人数**を削除

変数選択-時系列データ-

- 認知データとして対象 5 社の週ごとの**テレビ CM 放送総数**
- 共有データに関しては対象キーワード 5 社を含むツイートの**投稿総数・ネガティブツイート割合・ポジティブツイート割合**
- 多重共線性を回避するため**検索指数を削除**

変数選択-時系列データ-

スイッチング回数を目的変数にしたゼロ過剰ポアソン回帰 (ラグに関して)

SNS ラグ	AIC	CM ラグ	AIC
$t-1$	218.9*	$t-1$	235.5*
$t-2$	237.2	$t-2$	247
$t-3$	233.3	$t-3$	251.2
$t-4$	249.5	$t-4$	248.1
$t-5$	246.1	$t-5$	239.3

スイッチング回数を目的変数にしたゼロ過剰ポアソン回帰 (ツイートに関して)

SNS 変数	AIC
投稿数	223.1*
ネガティブ割合	256.3
ポジティブ割合	250.1

推定

- 211 人をランダムに二分割
- 推定にはマルコフ連鎖モンテカルロ法 (MCMC) を使用
- MCMC は 40,000 回 行い、最初の 30,000 回は burn in 期間として捨てる

モデル比較

以下の二つのモデルに対して適合度の比較

- ① 時系列情報のないマーケティング変数のみのモデル
- ② 条件付きプロビットモデルの効用関数の誤差項の分布の平均を 0 にしたモデル

結果

Table: 各モデルの適合度

	購買機会数	適合数	適合度
1	343	131	38.2 %
2	343	167	48.7 %
3 (提案モデル)	343	335	97.7 %*

回帰係数行列 Θ について

Table: 回帰係数行列の事後要約統計量

	年齢	性別	結婚ダミー	年収 500 万円以下ダミー	年収 500 ~ 1000 万円ダミー
価格係数 (ロイヤリティ有)	-0.00093	-0.00608	0.0347	0.0388	0.018
CM 係数 (ロイヤリティ有)	0.000569	-0.019	0.00339	-0.1117	-0.00436
SNS 係数 (ロイヤリティ有)	0.000112	-0.00323	-0.00081	-0.00374	0.000477
価格係数 (スイッチ対象)	0.000067	0.0332	-0.0970	0.00442	0.0425
CM 係数 (スイッチ対象)	-0.00105	0.0251	0.0257	0.0454	-0.00407
SNS 係数 (スイッチ対象)	-0.00004	0.00197	0.000553	0.00365	0.000728

分かること

- **ロイヤリティを持っていないメーカーに対しては、既婚者への値下げ効果が高い**
既婚者は購買時に店頭価格を比較して購入している
- **ロイヤリティを持っていないメーカーに対しては、既婚の女性と年収 500 万円以下の人に CM 効果が高い**
既婚の女性や比較的年収の低い消費者のテレビ CM の接触時間が長く、影響度が高い
- ソーシャルメディアに関しては大きな差が見られなかった

反応係数 α, β について

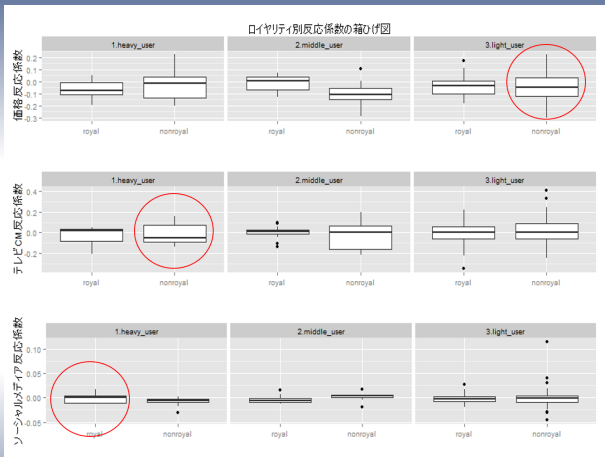


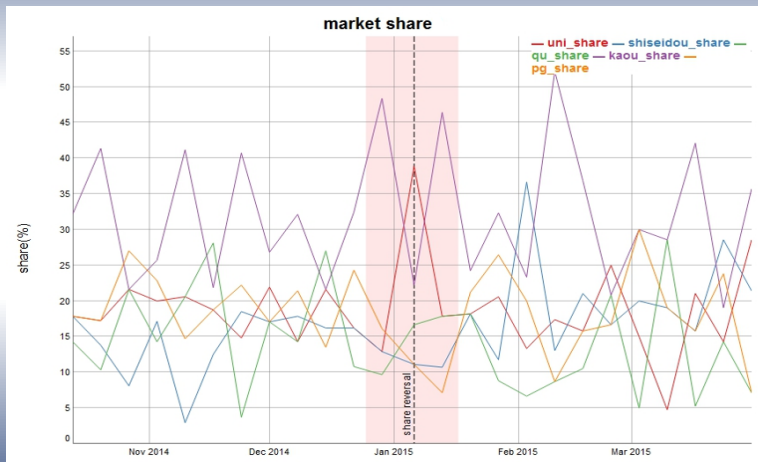
Figure: ロイヤリティ別反応係数の箱ひげ図

分かること

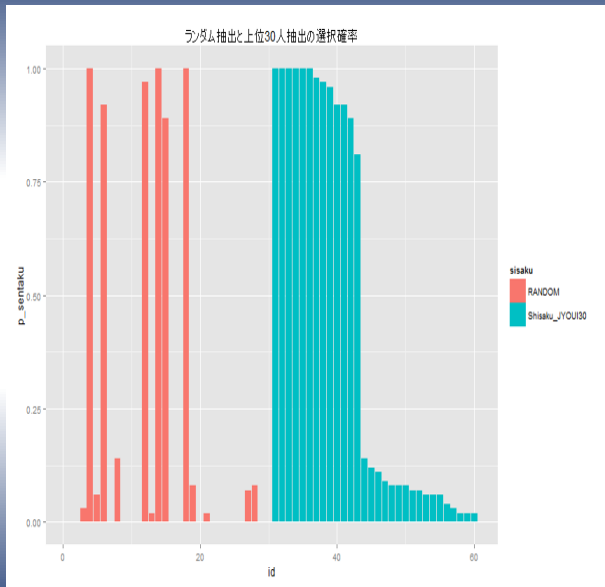
- **価格はライトユーザーとミドルユーザーに影響大**
価格がライトユーザーのスイッチング行動の要因になっている
- **CMはヘビーユーザーへの影響度が大きく、ライトユーザーに対してはほとんど影響がない**
ロイヤリティの高い消費者はCMを認知して購買に至る
- **SNSはヘビーユーザーのロイヤリティを持つメーカーに対する反応係数が高い**
ヘビーユーザーの購買プロセスはCMによる認知からSNS上での共有行動を行う消費者である

シチュエーション

2014年12月第2週～2015年2月第2週の間、市場シェアトップの花王がユリリーバに逆転された



花王の値下げ施策



花王の値下げ施策

既存のマスマーケティング

- ランダムに対象者を抽出
- 他社ロイヤルの内、花王の選択確率が0の消費者が20%いる
マスマーケティングのROIの低さ

マイクロマーケティング施策

- 対象期間中、他社ロイヤルの花王の選択確率上位30人を抽出
花王に乗り換え易い消費者にターゲットに絞った施策の有効性

結論

- ① 提案モデルの有効性が示された。
- ② 消費者一人ひとりの反応係数は、購買プロセスや属性によって異なる。
- ③ 消費者は選択肢にロイヤリティを持つかどうかによって反応係数が異なり、その大きさによっても反応係数が異なる。
- ④ 消費者異質性を考慮したマイクロマーケティングが有効な施策となり得る。