

# 状態空間モデルによる 購買間隔の規則性の推定

奥野 拓也, 中村 和幸

## 1. はじめに

近年, 大量のデータ (ビックデータ) が容易に蓄積可能となり, 蓄積したデータを活用したマーケティング活動が積極的に行われている. その中でも特に重要とされているのは顧客関係管理 (Customer Relationship Management: CRM) であり, これはマーケティング・サイエンスにおいても重要な分野の一つである.

CRM において最終購買日 (Recency), 来店頻度 (Frequency), 購入金額 (Monetary) を用いた RFM 分析は簡素であるが現場において広く使われる有用な手法であり, マーケティング・サイエンスにおいても RFM 分析の高度化に関するさまざまな研究が行われている [1-3].

その一方で近年 RFM の指標に加えて, 顧客の購買行動を測る新しい尺度としてクランピネス (Clumpiness) が Zhang et al. [4] によって提唱された. クランピネスは「イベント (来店・購買) の発生間隔が均等でない度合い」と定義され, 最小性 (イベント発生が等間隔であれば最小となる), 最大性 (イベントが集中して発生する場合は最大となる), 連続性 (イベントの発生間隔に応じて変動する) および収束性 (イベントの発生間隔が短くまたは長くなるにつれて, 増加または減少する) の四つの性質を有した指標である [21]. これら四つの性質を有したクランピネスは顧客ごとの購買間隔を用いて算出可能であり, RFM と同様に計算が簡単に行えるといったメリットがある. またクランピネスを測定

することで RFM では捉えられない顧客の購買行動を識別することが可能であるとし RFM にクランピネス (C) を加えた RFMC が今後の CRM において重要になると述べている. Zhang et al. [4] によって提唱されたクランピネスはマーケティング現場で利用されることを想定していることから簡単な計算によって算出される指標である. そのためデータの背景に存在する情報抽出が困難である. またクランピネスはイベントの発生間隔に着目した指標であり, これまでもイベント発生間隔に関する研究は古くから行われており, いづれの先行研究においてもイベント発生間隔に確率分布を仮定し消費者行動を分析している [5-9, 22, 23]. 例えば Gupta [5] は購買間隔に Erlang-2 分布を仮定するモデルを採用し, プロモーション効果が購買間隔へ与える影響を調べた. Moe and Fader [9] は購買間隔に指数分布を仮定し, その指数分布のパラメータにガンマ分布を仮定した階層ベイズモデルによって消費者の異質性をモデルに組み込んでモデル化し, 購買間隔の時間的変化を捉えることが重要であると述べている.

一方クランピネスに焦点を当てた研究として Platzer and Reutterer [1] がある. 彼らの研究では購買間隔にガンマ分布を仮定し, そのガンマ分布のパラメータにもガンマ分布を仮定する Pareto/GGG モデルを提案した. Pareto/GGG モデルは Zhang et al. [10] によって提案されたクランピネスの確率モデルと考えることができ, 消費者の異質性を取り込んだモデルである. Pareto/GGG モデルを用いて六つのデータ・セットで検証した結果, 個人の購買タイミングの規則性を推定することによって将来の行動を予測することが可能となると提案している. しかし Platzer and Reutterer [1] が提案したモデルはクランピネスの確率モデルであるといえるが, Moe and Fader [9] が指摘する購買間隔の時間変動については考慮されていない.

時間変動を考慮したモデル化として状態空間モデルを用いた研究は精力的に行われており, 集計データに

おくの たくや

明治大学大学院先端数理学研究科

〒 164-8525 東京都中野区中野 4-21-1

NTT テクノクロス (株) メディアイノベーション事業部

〒 231-0032 神奈川県横浜市中区不老町 2-9-1

okuno.takuya@po.ntt-tx.co.jp

なかむら かずゆき

明治大学総合数理学部

〒 164-8525 東京都中野区中野 4-21-1

受付 17.7.25 採択 17.11.5

対する適用例として青柳と佐藤 [11], 本橋と樋口 [12] による報告がある. また状態空間モデルを個人レベルに適用した例も報告されている [13, 14]. 特に個人レベルに適用した例はイベントの生起行動に着目して, チラシやクーポンのマーケティング施策がイベント生起へ与える影響を推定することを目的としている.

本研究は RFMC の C に着目して個人ごとのクランピネスの時間変動を推定し, 顧客の購買特性 (購買パターン) を明らかにすることが目的である. 具体的には顧客の購買履歴から購買発生率と購買不規則性の時間変化を推定する. ここで購買発生率とは購買の発生頻度を表し, 購買不規則性とは購買発生間隔のばらつき具合 (クランピネスに対応する) を表す. モデル化は購買の発生を点過程として捉え, 購買発生間隔の分布にガンマ分布を仮定し, ガンマ分布のパラメータに事前分布を導入することで状態空間モデルの枠組みで行う.

本稿の残りの部分は以下のように構成される. 2節では, 本研究で提案するモデルの説明と定式化を行う. 3節では, 提案したモデルを実際のデータに適用し, その結果を考察する. 最後に, 4節で本研究のまとめをする.

## 2. モデル

本節では顧客の購買行動をモデル化するために購買間隔をモデル化し推定手法を説明する.

### 2.1 購買間隔のモデル化

購買発生 の 離散時間時系列データ  $\{t_1^{(j)}, t_2^{(j)}, \dots, t_{n_j}^{(j)}\}$  が与えられたもと購買の発生率と購買の不規則性を推定する問題を考える. ここで  $j$  が顧客,  $n_j$  が顧客  $j$  の購買数および  $t_i^{(j)} (i = 1, \dots, n_j)$  が顧客  $j$  の  $i$  回目の購買時刻である. はじめに推定のため購買間隔  $T_i^{(j)} = t_{i+1}^{(j)} - t_i^{(j)} (i = 1, 2, \dots, n_j - 1)$  の生成分布を導入する. 購買間隔の生成分布はこれまでの先行研究と同様にある確率分布から生成されると仮定しモデル化する. 本研究では購買間隔は生成分布  $p(T_i^{(j)} | \lambda^{(j)}(t_i), \kappa^{(j)}(t_i))$  から生成されると仮定する. ここで  $\lambda^{(j)}(t_i), \kappa^{(j)}(t_i)$  はそれぞれ顧客  $j$  の  $i$  番目の購買間隔  $T_i^{(j)}$  におけるパラメータであり, 購買発生率と購買不規則性と定義する. つまり購買発生間隔は発生率  $\{\lambda^{(j)}(t)\}$  と, 購買不規則性  $\{\kappa^{(j)}(t)\}$  をもつ確率過程から生成されるとする. つぎに購買発生間隔の生成分布  $p(T_i^{(j)} | \lambda^{(j)}(t_i), \kappa^{(j)}(t_i))$  を決定する. 生成分布は Platzer and Reutterer [1] と同様にガンマ分布を採用し,

$$\begin{aligned} p(T_i^{(j)} | \lambda^{(j)}(t_i), \kappa^{(j)}(t_i)) &= \frac{1}{\Gamma(\kappa^{(j)}(t_i))} \left( \lambda^{(j)}(t_i) \right)^{\kappa^{(j)}(t_i)} \\ &\times \left( \kappa^{(j)}(t_i) \right)^{\kappa^{(j)}(t_i)} \left( T_i^{(j)} \right)^{\kappa^{(j)}(t_i) - 1} \\ &\times \exp \left( -\lambda^{(j)}(t_i) \kappa^{(j)}(t_i) T_i^{(j)} \right) \end{aligned} \quad (1)$$

と定義する. ここで  $\Gamma(\kappa^{(j)}(t_i))$  はガンマ関数である. 式 (1) の  $\kappa^{(j)}(t_i)$  が大きいときに規則的な購買を表し,  $\kappa^{(j)}(t_i)$  が小さいときに不規則的な購買を表す. 特に  $\kappa^{(j)}(t_i) = 1$  のとき, ポアソン分布となりランダム購買となる (図 1). 図 1 に示すように,  $\kappa$  が大きいと発生間隔が規則的になり, 小さくなるほど不規則的に発生する.

購買間隔の生成分布をモデル化したので購買の発生確率を考える. 顧客  $j$  の時刻  $t_i^{(j)}$  に購買が発生する確率は, 各時刻ごとに独立であると仮定し以下のように与えられるとする,

$$\begin{aligned} p(\{t_i^{(j)}\} | \{\lambda^{(j)}(t)\}, \{\kappa^{(j)}(t)\}) &= \prod_{i=1}^{n_j-1} p(T_i^{(j)} | \lambda^{(j)}(t_i^{(j)}), \kappa^{(j)}(t_i^{(j)})). \end{aligned} \quad (2)$$

### 2.2 推定方法

前節にて購買間隔をモデル化したので,  $\lambda^{(j)}(t), \kappa^{(j)}(t)$  の時間変動を推定することを考える. すなわち式 (2) の条件付き確率をベイズの定理を用いて反転させる.

$$\begin{aligned} p(\{\lambda^{(j)}(t)\}, \{\kappa^{(j)}(t)\} | \{t_i^{(j)}\}; \gamma_\lambda^{(j)} \gamma_\kappa^{(j)}) &= \frac{1}{p(\{t_i^{(j)}\}; \gamma_\lambda^{(j)} \gamma_\kappa^{(j)})} \\ &\times p(\{t_i^{(j)}\} | \{\lambda^{(j)}(t)\}, \{\kappa^{(j)}(t)\}) \\ &\times p(\{\lambda^{(j)}(t)\}; \gamma_\lambda^{(j)}) p(\{\kappa^{(j)}(t)\}; \gamma_\kappa^{(j)}). \end{aligned} \quad (3)$$

式 (3) から  $\lambda^{(j)}(t), \kappa^{(j)}(t)$  を推定するためには購買発生率と購買規則性に事前分布をモデル化する必要がある. そこで  $\lambda^{(j)}(t), \kappa^{(j)}(t)$  の事前分布としてガウス過程を導入する. これは  $\lambda^{(j)}(t), \kappa^{(j)}(t)$  が直前の値と大きく変動しないことを仮定しており, 大きな変動に対してオーバーフィッティングを防ぐことと計算可能性から選定した:

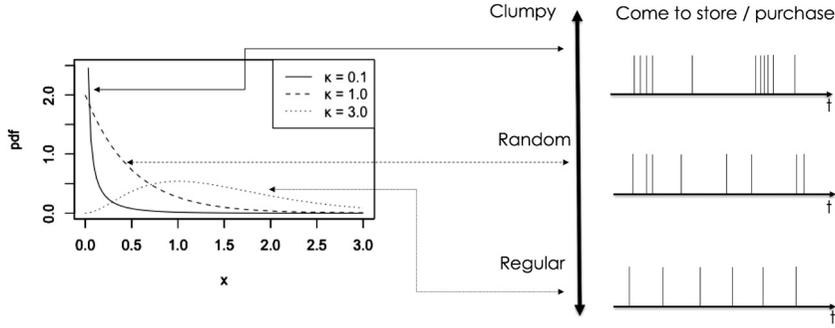


図1 ガンマ分布とガンマ過程から発生させたイベントの発生例 (左:  $\kappa = 0.1, 1.0, 3.0$  を変化させたときのガンマ分布. 右: ガンマ過程から発生させたイベント発生の概略図)

$$p\left(\{\lambda^{(j)}(t)\}; \gamma_\lambda^{(j)}\right) = \frac{1}{Z\left(\gamma_\lambda^{(j)}\right)} \exp\left[-\frac{1}{2\left(\gamma_\lambda^{(j)}\right)^2} \int_0^{T^{(j)}} \left(\frac{d\lambda^{(j)}(t)}{dt}\right)^2 dt\right], \quad (4)$$

$$p\left(\{\kappa^{(j)}(t)\}; \gamma_\kappa^{(j)}\right) = \frac{1}{Z\left(\gamma_\kappa^{(j)}\right)} \exp\left[-\frac{1}{2\left(\gamma_\kappa^{(j)}\right)^2} \int_0^{T^{(j)}} \left(\frac{d\kappa^{(j)}(t)}{dt}\right)^2 dt\right]. \quad (5)$$

ここで  $\gamma_\lambda^{(j)}, \gamma_\kappa^{(j)}$  はハイパーパラメータで,  $Z(\gamma_\lambda^{(j)}), Z(\gamma_\kappa^{(j)})$  は規格化定数を表す [15]. ハイパーパラメータの決定には式 (6) の周辺尤度を最大にすることで最適な値を EM アルゴリズムによって求める.

$$\begin{aligned} p\left(\{t_i^{(j)}\} | \gamma_\lambda^{(j)}, \gamma_\kappa^{(j)}\right) &= \int \int p\left(\{t_i^{(j)}\} | \{\lambda^{(j)}(t)\}, \{\kappa^{(j)}(t)\}\right) \\ &\quad \times p\left(\{\lambda^{(j)}(t)\}; \gamma_\lambda^{(j)}\right) p\left(\{\kappa^{(j)}(t)\}; \gamma_\kappa^{(j)}\right) \\ &\quad \times d\{\lambda^{(j)}(t)\} d\{\kappa^{(j)}(t)\} \end{aligned} \quad (6)$$

ハイパーパラメータ決定後に事後分布である式 (7) を最大化するような  $\lambda^{(j)}(t), \kappa^{(j)}(t)$  を推定する.

$$\begin{aligned} p\left(\{\lambda^{(j)}(t)\}, \{\kappa^{(j)}(t)\} | \{t_i^{(j)}\}; \gamma_\lambda^{(j)}, \gamma_\kappa^{(j)}\right) &\propto p\left(\{t_i^{(j)}\} | \{\lambda^{(j)}(t)\}, \{\kappa^{(j)}(t)\}\right) \\ &\quad \times p\left(\{\lambda^{(j)}(t)\}; \gamma_\lambda^{(j)}\right) p\left(\{\kappa^{(j)}(t)\}; \gamma_\kappa^{(j)}\right) \end{aligned} \quad (7)$$

### 2.3 状態空間モデルによる表現

前節までに示した購買間隔のモデルは非線形ガウス型状態空間モデルによって表現することができ, 時変回帰係数の推定は状態空間モデルにおける状態を推定する問題として定式化できる [16, 20]. 状態空間モデルは観測モデルとシステムモデルから定式化され, データの観測される定式化を記述するモデルを観測モデル, 時変回帰係数の時間発展の定式化を記述するモデルをシステムモデルと呼ぶ. 本モデルでは, 観測モデルを式 (1) とし, 式 (4), 式 (5) からシステムモデルは

$$\begin{aligned} p\left(\lambda_{i+1}^{(j)} | \lambda_i^{(j)}; \gamma_\lambda^{(j)}\right) &= \frac{1}{\sqrt{2\pi\left(\gamma_\lambda^{(j)}\right)^2 T_i^{(j)}}} \\ &\quad \times \exp\left[-\frac{\left(\lambda_{i+1}^{(j)} - \lambda_i^{(j)}\right)^2}{2\left(\gamma_\lambda^{(j)}\right)^2 T_i^{(j)}}\right] \\ p\left(\kappa_{i+1}^{(j)} | \kappa_i^{(j)}; \gamma_\kappa^{(j)}\right) &= \frac{1}{\sqrt{2\pi\left(\gamma_\kappa^{(j)}\right)^2 T_i^{(j)}}} \\ &\quad \times \exp\left[-\frac{\left(\kappa_{i+1}^{(j)} - \kappa_i^{(j)}\right)^2}{2\left(\gamma_\kappa^{(j)}\right)^2 T_i^{(j)}}\right]. \end{aligned}$$

もしくは

$$\begin{aligned} p\left(\theta_{i+1}^{(j)} | \theta_i^{(j)}; \gamma^{(j)}\right) &= \frac{1}{\sqrt{2\pi|\mathbf{Q}_i^{(j)}|}} \\ &\quad \times \exp\left[-\frac{1}{2}\left(\theta_{i+1}^{(j)} - \theta_i^{(j)}\right)^\top \left(\mathbf{Q}_i^{(j)}\right)^{-1} \left(\theta_{i+1}^{(j)} - \theta_i^{(j)}\right)\right] \\ \mathbf{Q}_i^{(j)} &= \begin{pmatrix} \left(\gamma_\lambda^{(j)}\right)^2 T_i^{(j)} & 0 \\ 0 & \left(\gamma_\kappa^{(j)}\right)^2 T_i^{(j)} \end{pmatrix} \end{aligned}$$

と表され

$$\theta_{i+1}^{(j)} = \theta_i^{(j)} + w^{(j)},$$

と記述することができる [16, 17]. ただし  $\theta_i^{(j)} = (\lambda_i^{(j)}, \kappa_i^{(j)})^\top$ ,  $\gamma^{(j)} = (\gamma_\lambda^{(j)}, \gamma_\kappa^{(j)})^\top$  および  $w_i^{(j)}$  は平均 0, 分散共分散行列  $Q^{(j)}$  のガウス分布に従う.

状態空間モデルの状態推定方法は観測モデルが非ガウス型であるため推定には非線形カルマンフィルタを用いる.

### 3. 実データへの適用

推定に利用したデータ概要について 3.1 節で述べ, そのあとに 2 節にて示したモデルを用いて推定した結果について示す.

#### 3.1 データ概要

本研究では経営科学系研究部会連合協議会主催, 平成 28 年度データ解析コンペティションで提供されたデータを利用した. データ期間は 2015 年 4 月 1 日から 2016 年 3 月 31 日の 1 年間であり, ファッション系 EC サイト 1 店舗における顧客の購買履歴を使用した. 分析対象は図 2 に示すように購買数の上位 5 カテゴリ, すなわちトップス・シューズ・パンツ・アンダーウェアおよびジャケット/アウターである. なお, 全購買数の 8 割近くを占めているためトップ 5 のカテゴリのみを対象とした. また各カテゴリの全顧客を対象にせず各カテゴリ購買数の上位 10% の顧客を対象とした. 各カテゴリの分析対象者数は表 1 に示すとおりである. 分析対象を上位 10% に絞った理由は,  $\lambda_i^{(j)}, \kappa_i^{(j)}$  を個人ごと・カテゴリごとに推定するため全顧客を分析対象とすると相応のマシンリソースが必要となるためである. また購買金額ではなく購買数の上位 10% の顧客を選定しているのは本研究の目的が RFMC の C (クランピネス) に焦点を当てているため購買頻度が高い顧客を対象とした. このように, 分析対象を絞ったのは計算時間の都合上の問題であり, 推定に必須の制約ではない.

#### 3.2 推定結果

本節では 2 節にて構築したモデルを 3.1 節のデータ (トップス) を利用して推定した結果の一例を図 3 に示す. 横軸はデータ提供開始期間 (2015 年 4 月 1 日) からの日数を表しており, 上・中段は購買発生率  $\lambda_i^{(j)}$ , 購買規則性  $\kappa_i^{(j)}$  であり実線が推定値の平均値, 破線が分散を表している. なお図 3 では中盤で購買頻度が低下し購買発生率が大きく低下し, 後半に上昇するも全体的に減少傾向であるため購買頻度が低下している. 中図では全体的に減少傾向であるため購買間隔が広がっ

表 1 カテゴリごとの分析対象者数

カテゴリ	全購買数	分析対象者
トップス	1,631,630	653
シューズ	702,405	617
パンツ	564,993	423
アンダーウェア	350,731	181
ジャケット/アウター	230,984	487

表 2 ハイパーパラメータ  $\gamma$  のカテゴリごとの推定値の平均値と分散. いづれの値も小さく安定した推定値が得られている.

カテゴリ	$\bar{\gamma}_\lambda$	$\bar{\gamma}_\kappa$
トップス	0.0047	0.0318
シューズ	0.0016	0.0199
パンツ	0.0025	0.0242
アンダーウェア	0.0013	0.0170
ジャケット/アウター	0.0032	0.0259

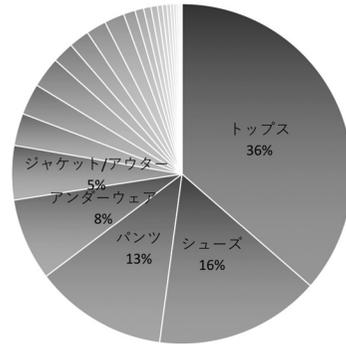


図 2 購買カテゴリの割合

てきていることがわかる. また, 下段は推定に用いた入力データであり, スパイクが立っている位置が購買発生日を表している. 図 3 から安定した推定値が得られていることが確認できる. 全顧客についても安定した推定ができており, 推定したハイパーパラメータ  $\gamma^{(j)}$  の平均値  $\bar{\gamma} = \sum_{j=1}^{J_c} \gamma^{(j)} / J_c$  は表 2 に示すとおりであり十分小さな値が得られている. ここで  $J_c$  は各カテゴリの分析対象者数であり具体的な数値は表 1 の右端に示す. また下段は推定に用いた入力データで, スパイクが立っているところに購買があったことを表す.

購買発生率  $\lambda_i^{(j)}$  は購買の起りやすさを表す指標であり上昇傾向であれば購買意欲が高く再購買が期待され, 下降傾向であれば再購買の意欲が低下していると理解できる. ただし購買発生率は確率ではないため注意が必要である. 図 3 の上段に示すとおり, 本顧客の購買発生率は前半と後半部分が高くなり, 中盤が落ち込んでいる. 図中の下段の購買状況からも中盤付近

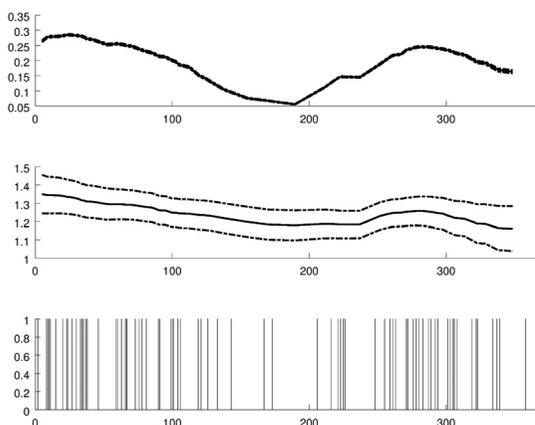


図3 推定結果

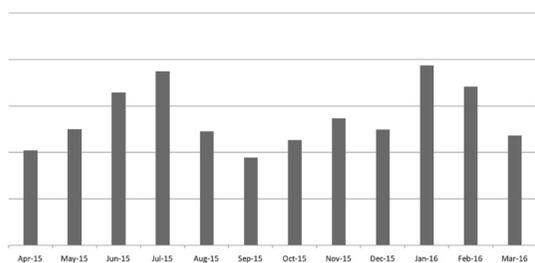


図4 トップカテゴリの月別販売個数

では購買頻度が減少している様子が見て取れる。また図4に推定に利用したカテゴリデータの月別販売数を示す。縦軸は月ごとの購買数（縦軸はサイトの規模等が特定されてしまうため割合）を表す。図4からわかるとおり図3の顧客はカテゴリの売上と連動した形で購買発生率が高くなっている。購買数が高くなる月をみると7月と1月がピークとなっており、ファッション系ECサイトにおいてセールが実施されたのではないかと推測できる。このような検証は状態空間モデルを利用し動的な変動を推定することでデータに含まれていない知見が獲得可能となる一例である。

購買規則性  $\kappa_i^{(j)}$  は購買の規則性を表す指標であり、高い値を示すと購買間隔がある程度一定となる購買パターン（定期購買）を表し、低い値ではいわゆる「マイブーム」などの短い期間で固まった購買パターンを表す（図1）。図3の中段に示す購買規則性の推定値は前半緩やかに減少し、後半一時的に上昇するがその後すぐに減少傾向である。また全体的に減少傾向であることから購買頻度が徐々に低下している様子が見て取れる。一人ひとりの結果をみることで同様の考察が可能となるが、本誌では紙面の都合上、詳細は顧客1名の結果を示すに留める。次にカテゴリごと・顧客ごと

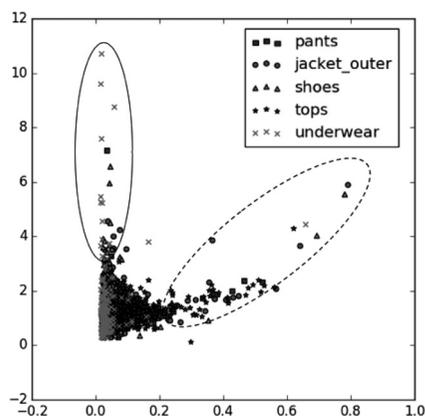


図5 全カテゴリ・全顧客の推定値  $\bar{\theta}^{(j)} = (\bar{\lambda}^{(j)}, \bar{\kappa}^{(j)})^T$  の散布図

の推定値を図5に示す。図中の横軸は式(8)に示す顧客  $j$  の購買発生率の推定値  $\lambda_i^{(j)}$  の時間軸方向の平均値  $\bar{\lambda}^{(j)}$ 、縦軸は同様に式(9)に示す購買規則性の推定値  $\kappa_i^{(j)}$  の時間軸方向の平均値  $\bar{\kappa}^{(j)}$  を表す。

$$\bar{\lambda}^{(j)} = \frac{\sum_{i=1}^{n_j} \lambda_i^{(j)}}{n_j} \quad (8)$$

$$\bar{\kappa}^{(j)} = \frac{\sum_{i=1}^{n_j} \kappa_i^{(j)}}{n_j} \quad (9)$$

図中の実線で囲まれた部分は購買規則性が高く定期的な購買（本論文では  $\kappa \geq 3$  を定期購買顧客と定義する）が見込まれる顧客群である。利用頻度が高いアンダーウェアが多く存在し定期的な購買発生が見込まれることも納得がいく結果である。ただしこれらの顧客は購買発生率が低いため直近購買や購買頻度を評価する方法では優良顧客として扱われないことがある。しかし年間の売上のベースラインの確保という意味では重要な顧客である。また点線で囲んだ顧客は、購買頻度が高いかつ購買が発生しやすい（直近に購買が発生しやすい）顧客である。つまり購買発生率  $\bar{\lambda}_i^{(j)}$  が大きい値の顧客は次回購買が期待される顧客であり、そのカテゴリにおける優良な顧客であると考えられる。そのため継続的にケアをして顧客の離脱を防ぐことが重要であるといえる。図6~10は図5をカテゴリごとに分解した図である。図6はトップスの推定値であり、購買回数が他カテゴリよりも圧倒的に多いため購買の発生率の高い顧客が多く存在する。図7はシューズの推定値であり、トップスと比較して定期購買層が多いことが確認できる。シューズは季節ごとに大きく変化するため定期的な購買が多いと考えられる。図8はパンツの推定値であり、パンツは年間を通して利用できるアイテムであることから購買回数が多い顧客が少な

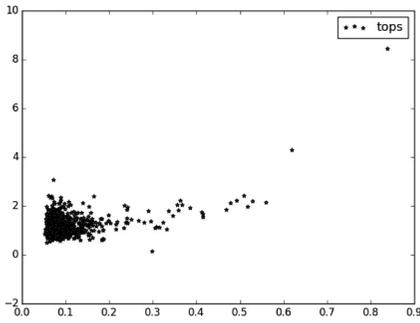


図 6 トップスにおける推定値  $\bar{\theta}^{(j)}$  の散布図

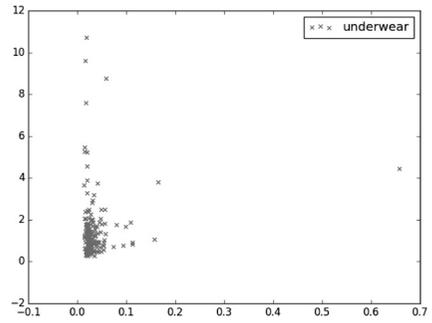


図 9 アンダーウェアにおける推定値  $\bar{\theta}^{(j)}$  の散布図

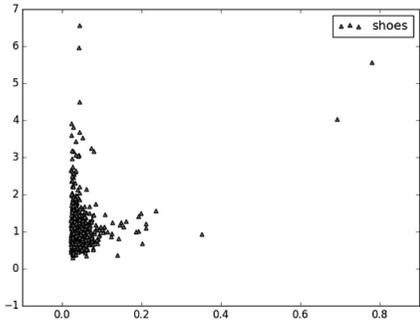


図 7 シューズにおける推定値  $\bar{\theta}^{(j)}$  の散布図

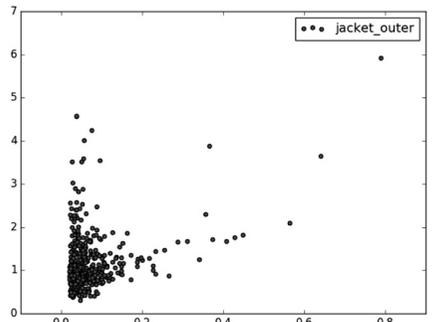


図 10 ジャケット/アウターにおける推定値  $\bar{\theta}^{(j)}$  の散布図

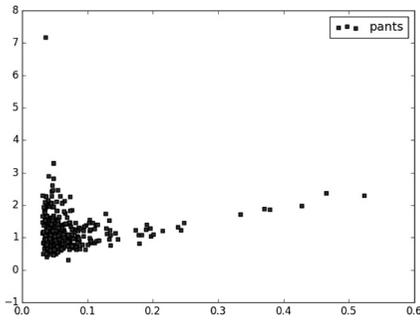


図 8 パンツにおける推定値  $\bar{\theta}^{(j)}$  の散布図

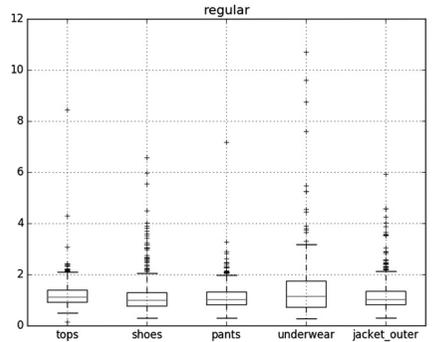


図 11 カテゴリごとの購買規則性の推定値.

くない。そのため、 $\bar{\lambda}^{(j)}$  が大きな顧客が多いと考えられる。図 9 はアンダーウェアの推定値であり、シューズと同様に定期的な購買顧客が多い。他カテゴリに比べて同じ商品の利用回数が多いと想像されるため定期的な購買が発生していると考えられる。図 10 はジャケット/アウターの推定値であり、寒い時期に購買が多いと予想したが、定期的な購買が発生している顧客の存在が確認できる。また全体から見れば販売数は少ないが高頻度で購入する顧客と定期的な購買が発生する顧客が入り交じったカテゴリであると図から読み取れる。

図 11 はカテゴリごとの購買規則性を箱ひげ図で示

したものである。図からカテゴリ間での中央値に大きな違いはない。これは多くの顧客がランダム購買していることを示している反面、各カテゴリにおいて定期的な購買が発生している顧客の存在も確認できる。

### 3.3 一期先予測による評価

本節では本手法と平均購買間隔による手法による一期先予測において性能を比較し評価する。ここで平均購買間隔による手法とは顧客  $j$  の第  $t_i^{(j)}$  日までの購買履歴から平均購買間隔を算出し、次回の購買日  $\hat{t}_{i+1}^{(j)} = t_i^{(j)} + \sum_{k=1}^i T_k^{(j)} / i$  を予測する。ゆえに顧客ごとに  $\{\hat{t}_3^{(j)}, \dots, \hat{t}_{n_j}^{(j)}\}$  の購買日を予測する。同様に本手

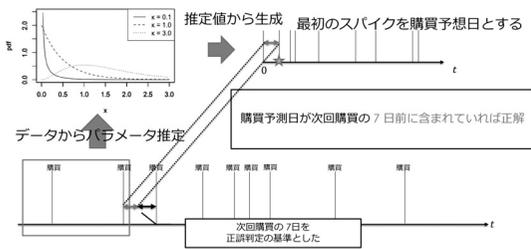


図 12 評価方法の概要

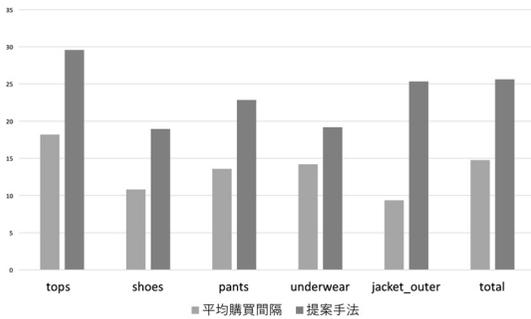


図 13 平均購買間隔による方法と本手法の正解率 (左: 平均購買間隔, 右: 本手法)

表 3 本手法と平均購買間隔による一期先予測の正解率

M	平均購買間隔	本手法
4	10.3	18.7
5	11.9	21.4
6	13.4	23.3
7	14.7	25.6

法による一期先予測も顧客  $j$  の  $i$  回目の購買日  $t_i^{(j)}$  までのデータを利用した推定値から次回の購買日  $\tilde{t}_{i+1}^{(j)}$  を予測する (図 12)。

各手法の一期先予測の評価指標として予測値  $\tilde{t}_{i+1}^{(j)}$  と実際の購買日  $t_{i+1}^{(j)}$  の差分が  $M$  日以内であるかを評価した。予測誤差が  $M$  日以内である割合を正解率とし、各カテゴリ・各個人ごとに算出し、一期先の予測精度として比較する。したがってカテゴリごとの予測の正解率  $RC$  は、

$$RC = \frac{\sum_{j=1}^{J_c} RC^{(j)}}{J_c},$$

$$RC^{(j)} = \frac{\sum_{i=3}^{n_j} I(\tilde{t}_i^{(j)}, t_i^{(j)})}{n_j},$$

ここで  $RC^{(j)}$  はカテゴリにおける個人  $j$  の正解率を表し、 $n_j$  は顧客  $j$  の総購買数、 $J_c$  が全顧客数、 $\tilde{t}_i^{(j)}$  が顧

客  $j$  の一期先予測購買日および  $t_i^{(j)}$  が顧客  $j$  の第  $i$  回目の実測の購買日を表す。また  $I(\tilde{t}_i^{(j)}, t_i^{(j)})$  を

$$I(\tilde{t}_i^{(j)}, t_i^{(j)}) = \begin{cases} 1 & |\tilde{t}_i^{(j)} - t_i^{(j)}| \leq M \\ 0 & |\tilde{t}_i^{(j)} - t_i^{(j)}| > M \end{cases}$$

と定義する。本研究では  $M$  は実務で利用することを考え  $M = 7$  とした。これは実務においてマーケティングが顧客へアクションが取れる主な手段は週 1 回程度送付されるメルマガが多いことから妥当であると考えられる [18]。

図 13 に本手法と比較するために実務で利用されている平均購買間隔による一期先予測と比較した結果を示す。平均購買間隔による手法における  $RC$  はトップス 18.2%、シューズ 10.8%、パンツ 13.6%、アンダーウェア 14.1%、ジャケット/アウター 9.4%、全体 14.7% となった。また本手法における  $RC$  はトップス 29.5%、シューズ 18.9%、パンツ 22.9%、アンダーウェア 19.2%、ジャケット/アウター 25.3%、全体 25.6% である。すべてのカテゴリにおいて本手法が上回っており、全体としても 10% 以上向上していることが図から読み取れる。また  $M \leq 7$  の場合における一期先の予測精度を表 3 に示す。いずれの  $M$  の値においても本手法が約 10% 程度上回っていることが表から確認できる。

#### 4. おわりに

本研究では近年提唱された RFMC の C に焦点を当て、その動的な変動を捉えるために購買間隔を状態空間モデルでの枠組みで定式化し、EC サイトの 1 年分の実データを用いて検証した。その結果、各カテゴリには売上のベースとなるような定期購買者の存在が示唆された。また本手法によって検出可能な顧客は従来の RFM 分析では見落とされる顧客であるが重要な顧客であるといえる。また 3.3 節の予測を用いることで次回購買が予測可能になり、比較手法よりもより精度の高いタイミングで施策を実施することが可能となる。

今後の課題として購買間隔生成にマーケティング変数を取り入れることが考えられる。これは Zhang et al. [4] が述べるように RFMC を統合的に扱う必要がある。そのため本モデルに RFM を取り込んだモデル化が必要である。また本研究では計算時間の都合上、全ての顧客を分析対象としなかった。そのため実務で利用できる範囲に計算時間を抑えることも今後の課題である。先行研究では同様の点過程モデルで購買間隔ではなく購買発生確率を焦点にしモデル化し購買予測

しているアプローチも提案されている [19]。これらの比較や組み合わせた分析方法も今後の課題となる。

**謝辞** 本研究にあたり、明治大学田野倉葉子博士からは、多くの有用かつ建設的なコメントいただきました。最後になりましたが、継続的にデータを提供いただいている経営科学系研究部会連合協議会の皆様とデータを提供いただきました提供者の方々にお礼申し上げます。

#### 参考文献

- [1] M. Platzter and T. Reutterer, “Ticking away the moments: Timing regularity helps to better predict customer activity,” *Marketing Science*, **35**, pp. 779–799, 2015.
- [2] 阿部誠, “RFM 指標と顧客生涯価値—階層ベイズモデルを使った非契約型顧客関係管理における消費者行動の分析—,” *日本統計学会誌*, **41**, pp. 51–81, 2011.
- [3] M. Abe, “Counting your customers one by one: A hierarchical Bayes extension to the Pareto/NBD model,” *Marketing Science*, **28**, pp. 541–553, 2009.
- [4] Y. Zhang, E. T. Bradlow and D. S. Small, “New measures of clumpiness for incidence data,” *Journal of Applied Statistics*, **40**, pp. 2533–2548, 2013.
- [5] S. Gupta, “Impact of sales promotions on when, what, and how much to buy,” *Journal of Marketing Research*, **25**, pp. 342–355, 1988.
- [6] A. S. C. Ehrenberg, “The pattern of consumer purchases,” *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **8**, pp. 26–41, 1959.
- [7] D. C. Jain and N. J. Vilcassim, “Investigating household purchase timing decisions: A conditional hazard function approach,” *Marketing Science*, **10**, pp. 1–23, 1991.
- [8] Y.-H. Park and P. S. Fader, “Modeling browsing behavior at multiple websites,” *Marketing Science*, **23**, pp. 280–303, 2004.
- [9] W. W. Moe and P. S. Fader, “Dynamic conversion behavior at e-commerce sites,” *Management Science*, **50**, pp. 326–335, 2004.
- [10] Y. Zhang, E. T. Bradlow and D. S. Small, “Predicting customer value using clumpiness: From RFM to RFMC,” *Marketing Science*, **34**, pp. 195–208, 2015.
- [11] 青柳憲治, 佐藤忠彦, “3 階層多変量状態空間モデリングによる動的市場反応形成メカニズムの解明,” *日本オペレーションズ・リサーチ学会和文論文誌*, **58**, pp. 70–100, 2015.
- [12] 本橋永至, 樋口知之, “市場構造の変化を考慮したブランド選択モデルによる購買履歴データの解析,” *マーケティング・サイエンス*, **21**, pp. 37–59, 2013.
- [13] 佐藤忠彦, 樋口知之, “動的個人モデルによる購買生起行動の解析,” *日本統計学会論文誌 シリーズ J*, **38**, pp. 1–19, 2008.
- [14] 奥野拓也, 中村和幸, “個人別セールスプロモーション効果の推定,” *情報処理学会論文誌*, **9(3)**, pp. 61–74, 2016.
- [15] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.
- [16] T. Shimokawa and S. Shinomoto, “Estimating instantaneous irregularity of neuronal firing,” *Neural Computation*, **21**, pp. 1931–1951, 2009.
- [17] S. Koyama, T. Shimokawa and S. Shinomoto, “Phase transitions in the estimation of event rate: A path integral analysis,” *Journal of Physics A: Mathematical and Theoretical*, **40(20)**, pp. F383–F390, 2007.
- [18] メルラボ, メールマガジンに関する意識調査, <https://mailmarketinglab.jp/survey-about-mail-magazine-2014/> (2017 年 7 月 1 日閲覧)
- [19] H. Kim, N. Takaya and H. Sawada, “Tracking temporal dynamics of purchase decisions via hierarchical time-rescaling model,” In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 3–7, 2014.
- [20] T. Shimokawa, S. Koyama and S. Shinomoto, “A characterization of the timerescaled gamma process as a model for spike trains,” *Journal of Computational Neuroscience*, **29**, pp. 183–191, 2010.
- [21] 中山雄司, “顧客関係管理研究の新動向—来店/購買間隔の不均一性を測るクランビネス指標—,” *甲南経営研究*, **57**, pp. 161–181, 2016.
- [22] 山口景子, “頻度の時間変化を考慮した階層ベイズモデルによるウェブサイト訪問行動の分析,” *マーケティング・サイエンス*, **22**, pp. 13–29, 2014.
- [23] W. W. Moe and P. S. Fader, “Capturing evolving visit behavior in clickstream data,” *Journal of Interactive Marketing*, **18**, pp. 5–19, 2004.