論文・事例研究

チェック・インから購買までの時間を利用したマーケティング・セグメンテーション
―潜在混合分布モデルを利用した交差検証法による段階的ベイズ分析―

長尾　圭一郎，豊田　秀樹，秋山　隆

1. はじめに

消費者行動モデルは企業や周りの環境による刺激に対する購買者の反応で図式化され，消費者の購買行動は文化的，社会的，個人的，心理的な要因から影響を受ける [1]．この中で個人的な要因に着目すると「年齢」や「ライフスタイル」などが挙げられる．一方，企業からの刺激は「製品」や「価格」などに加え，「プロモーション」も大きな要素となる．IT 技術の向上により，近年ではスマートフォンアプリを用いたプロモーションが注目されている．

“MUJI passport”は（株）良品計画が 2013 年から無料配布しているスマートフォンアプリである [2]．これをダウンロードした顧客は，店舗在庫検索や商品購入によるマイルの加算といったサービスが受けられる．アプリ内では，顧客が商品の評価を投稿できるページ“my MUJI”も利用可能であり，これによって顧客は，店舗において，商品を比較・評価しながら購買を決定することが可能となった．しかし，MUJI passport が既存のサービスと一線を画すのは“チェック・イン”システムの存在によるところが大きい．

スマートフォンに搭載されている GPS 機能を用いて，ユーザは半径 600 m 以内の無印良品の店舗にチェック・インすることができる．チェック・インは 1 店舗につき，1 日 1 回と制限されており，チェック・インを行った顧客には 1 回につき 10 マイルが進呈される．

チェック・インは店舗購買を主とするユーザの消費者行動を把握するうえで重要な役割を果たす．たとえば，チェック・インから購買までの時間を算出することで，そのユーザが朝の通勤・通学の際に日常的にチェック・インを行い，帰りがけに購買している顧客であるのか，通勤の際にチェック・インを行い，昼休みに購買を行う顧客であるのかといった判別を行うことができる．また，通常の顧客（以下，通常ユーザ）と顧客の最高ステージに到達した“ダイヤモンド・ユーザ[1]”の消費者行動の相違に着目し，マーケティング・セグメンテーションを行えば，購買に対して相対的に効率の高い結果が期待できる．

1.1 データ説明

本研究では，経営科学系研究部会連合協議会主催の平成 26 年度データ解析コンペティションで提供された，2013 年 5 月 15 日から 2014 年 6 月 30 日までの無印良品 441 店舗における受注データと MUJI passport によるチェック・インのデータを利用した．観測変数として，顧客のチェック・インから購買までの時間を受注データの購買時間から MUJI passport データのチェック・イン時間を差し引くことで算出した．このとき，単位は分となるように変数を操作した．

次に，データクレンジングとして，チェック・インから購買までの時間は，両方が同じ日に行われたものに限定した．また，夜中にアプリを操作しているときにチェック・インすることも可能であるが，消費者行動

ながお　けいいちろう
早稲田大学大学院文学研究科
〒 162–8644　東京都新宿区戸山 1–24–1
kei1row@fuji.waseda.jp
とよだ　ひでき，あきやま　たかし
早稲田大学文学学術院
〒 162–8644　東京都新宿区戸山 1–24–1
toyoda@waseda.jp
akiyamat@aoni.waseda.jp
受付 15.7.20　採択 15.11.7

[1] （株）良品計画ではマイル数に応じて「シルバー」から「ダイヤモンド」へと 5 段階にわたって顧客のステージが上がる制度を採用している [2]．本研究では，最上位ステージに位置するダイヤモンド・ユーザを優良顧客とみなす．

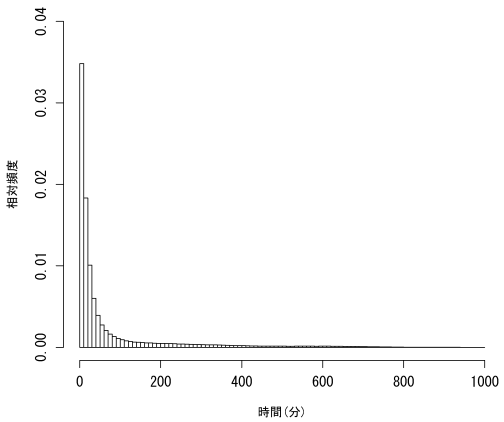

図 1　通常ユーザのチェック・インから購買までの時間の相対度数分布（単位：分）

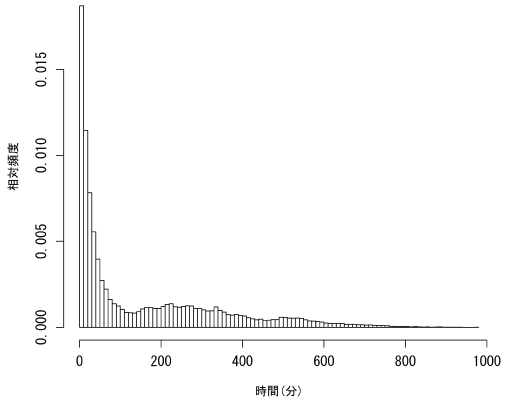

図 2　ダイヤモンド・ユーザのチェック・インから購買までの時間の相対度数分布（単位：分）

の連続性を重視して，データは早朝からのものが妥当であると考え，6 時から 24 時までのチェック・インに限定した．購入してからチェック・インしたトランザクションについても妥当性がないと判断し，除外した．

トランザクション数に関する，通常ユーザのチェック・インから購買までの時間の相対度数分布を図 1 に，ダイヤモンド・ユーザのチェック・インから購買までの時間の相対度数分布を図 2 に示す．

1.2　相対度数分布の比較

図 1 で示された通常ユーザのチェック・インから購買までの時間の相対度数分布の形状は単調減少関数である裾の重い単一分布の存在を示唆する．対して，図 2 で示されたダイヤモンド・ユーザのチェック・インから購買までの時間の相対度数分布では，150 分付近までは単調に減少しているが，そこから増加・減少を繰り返し，240 分付近と 520 分付近にピークをもつような 2 峰の分布が確認できる．ここから，ダイヤモンド・ユーザの分布は通常ユーザの分布とその他二つの単峰分布の混合分布であり，一つひとつの分布が独自のセグメントを構成していることが示唆される．すなわち，ダイヤモンド・ユーザには三つの潜在的なクラスが存在し，通常ユーザの分布とダイヤモンド・ユーザに顕著な二つの分布の潜在混合分布モデルで表現できることが考えられる．

1.3　本研究の流れ

以上を受け，本研究では，通常ユーザとダイヤモンド・ユーザのチェック・インから購買までの時間の相対度数分布の相違に着目し，通常ユーザにおけるどのような特徴をもった顧客にプロモーションをかければ，購買に対して相対的に効率の高い結果が得られるのか検証する．

本研究は四つの分析から構成される．分析 1 では，時間の分布に適用される単調減少関数である四つの理論分布を候補に取り上げ，通常ユーザのチェック・インから購買までの時間の相対度数分布に最も妥当な理論分布を特定する．分析 2 では，分析 1 で特定した分布を混合分布の要素分布として利用し，ダイヤモンド・ユーザに顕著なセグメントを構成する二つの分布を，潜在混合分布モデルを用いて特定する．分析 3 では，分析 2 で明らかにした潜在混合分布モデルの混合比率に線形予測子を組み込むことで，セグメントごとの特徴の違いを明らかにする．最後に，分析 4 では，分析 3 で得られたモデル構造を利用し，実際にダイヤモンド・ユーザに顕著なセグメントの特徴をもつ通常ユーザを掘り起こすことで，セグメント別の購買に対する相対的な効率について検証する．

本研究では，交差検証法を利用し，段階的にモデルの構造を明らかにすることで頑健な結果を得ることを試みる．このため，前の分析で得られた結果を次の分析に反映することができるベイズ分析を分析 2 と分析 3 で採用した．交差検証法は異なるデータに対する得られたモデルの当てはまりのよさを確認する方法であり，ベイズ分析ではデータの 2 度使いが許されないことから，分析 1 から分析 4 までのデータは期間で分け，それぞれ独立となるものを用いた[2]．

2.　分析 1

2.1　目的

分析 1 では，通常ユーザのチェック・インから購買までの時間の相対度数分布に最も妥当な理論分布を特定する．

2.1.1 候補となる理論分布

図 1 の相対度数分布は裾の重い単調減少関数の形状を示している．また，その変量が時間であることから，これらの特徴によく適用される理論分布として，以下の四つの分布を候補として取り上げた．

一つ目は指数分布であり，$X_1, \ldots, X_i, \ldots, X_N$ における X_i を通常ユーザの i 番目のトランザクションにおけるチェック・インから購買までの時間としたとき，確率密度関数は

$$f(x) = \lambda e^{-\lambda x}, \quad 0 \leq x < \infty, \quad \lambda > 0 \tag{1}$$

と表される．指数分布の最尤推定量は解析的に求まり，

$$\hat{\lambda} = \overline{X} \tag{2}$$

であることが知られている．ここで $\overline{X} = 1/N \sum_{i=1}^{N} X_i$ は通常ユーザのチェック・インから購買までの時間の平均を表す．

指数分布はポアソン過程における待ち時間の分布として得られる [3]．チェック・インから購買までを 1 回観察のポアソン過程とみなし，この待ち時間の分布として指数分布を利用することは妥当であると考えた．また，図 1 の曲線は忘却曲線に似た形状を示している．忘却曲線にも指数関数モデルが適用されており [4]，図 1 の曲線が顧客の店舗に対する記憶保持時間の分布を表していると仮定し，指数分布を適用する動機とした．

二つ目は対数正規分布であり，確率密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^* x} \exp\left[-\frac{(\log x - \mu^*)^2}{2\sigma^{*2}}\right], \ 0 < x < \infty \tag{3}$$

と表され，最尤推定量は $Y_i = \log X_i$ とおくとき

$$\hat{\mu}^* = \overline{X}, \quad \hat{\sigma}^{*2} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \overline{Y})^2 \tag{4}$$

と解析的に求まる．

対数正規分布は「顧客がシステムに到着してから去るまでの時間」を表す「サービス時間」の分布として利用される [5]．この定義は，病院外来の待ち時間，路上駐車の駐車時間，情報通信の待ち時間などを包括する．顧客がチェック・インを行うことで，システムにアクセスを行い，会社での勤務，商品の選別などの時間を経て，購買によってシステムから去ることはサービス時間の定義に適うと考えた．

三つ目は第 1 種パレート分布であり，確率密度関数は

$$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}}, \quad x \geq \alpha \tag{5}$$

であり，最尤推定量は

$$\hat{\alpha} = \min X_i, \quad \hat{\beta} = N\left[\sum_{i=1}^{N} \log\left(\frac{X_i}{\hat{\alpha}}\right)\right]^{-1} \tag{6}$$

である．

第 1 種のパレート分布は全体の数値の大部分は，全体を構成する要素の一部が生み出すというパレートの法則をモデル化した分布である．パレート分布もまた「サービス時間」の分布として利用される [6]．サービス時間の分布としての適合度を対数正規分布とパレート分布で比較することで，よりチェック・インから購買までの時間をモデル化する分布を特定することが可能であると考えた．

四つ目はべき関数分布であり，確率密度関数は

$$f(x) = \left(\frac{\gamma}{\beta}\right)\left(\frac{x}{\beta}\right)^{\gamma-1}, \ 0 \leq x \leq \beta, \ \beta > 0 \tag{7}$$

と表される．最尤推定量については

$$\hat{\gamma} = \left[\frac{1}{N}\sum_{i=1}^{N} \log X_i\right]^{-1} \tag{8}$$

と求まるが，β については解析的に求まらないことが知られている [3]．

べき関数族の分布は指数関数族の分布に比べて，裾の重い形状となることが知られている [3]．べき関数型であるべき関数分布を候補に加えることで，分布の特定に幅をもたせることができると考え，適用の動機とした．また，べき関数分布に従う確率変数の負の対数をとったものが指数分布に，確率変数の分数をとったものが第 1 種のパレート分布に従うことが知られている [3]．チェック・インから購買までの時間がこのように可塑的なべき関数分布で説明される可能性は十分考えられる．

---

[2] 本研究では，新しい独立なデータを用いて，モデルを深化させたときに，妥当な結果（分析 2 における確率足し上げ図（図 4）や分析 4 におけるセグメント別の購買に対する相対的な効率の相違）が得られていることを確認することでモデルの精度を評価した．この方法は学習データとテストデータの関係性で述べられる一般的な交差検証法とは異なったアプローチであるが，広義には交差検証法的であると考え，「交差検証法」という語句を利用した．また，ベイズ分析において，データの 2 度使いが禁じ手とされている理由の一つにベイズ更新によって，推定の精度を示す事後標準偏差が確実に小さくなることが挙げられる．本研究は一般的なベイズ更新とは異なったアプローチをとっているが，事後標準偏差が小さくなる危険性を鑑みて，新しい独立なデータを利用し続けることで，学術的な公平性を保った．

2.1.2 推定方法

2014 年 2 月までの通常ユーザのデータを用い，以上四つの理論分布について，指数分布，対数正規分布，第 1 種パレート分布はデータから母数の最尤推定値を計算し，それを当てはめた分布の形状と適合度指標である AIC (Akaike's information criterion) を確認した．最尤推定量が解析的に求まらないべき関数分布に関しては，後述する HMC (Hamiltonian Monte Carlo) 法を用いて，一様分布を事前分布として母数をベイズ推定し，ここでは EAP (expected a posteriori) 推定値ではなく，MAP (maximum a posteriori) 推定値を代入した分布の形状と AIC を確認した．

MAP 推定値は一様分布を事前分布としたときに，最尤推定値に一致することが知られている．これを利用し，本研究では，解析的に求められないべき関数分布の最尤推定値を MAP 推定値から数値的に求めた．なお，MAP を推定する際には，母数の標本をソーティングし，階級幅を狭めていったときに最も多くの標本が含まれる階級を求めた．

2.1.3 ハミルトニアンモンテカルロサンプリング

べき関数分布の MAP 推定値を求めるために，本研究ではハミルトニアンモンテカルロ (Hamiltonian Monte Carlo; HMC [7]) 法を用いる．HMC 法はハイブリッドモンテカルロ法とも呼ばれる．

HMC 法では，母数を個別にサンプリングするのではなく，すべての母数について一度にサンプリングを行うため，従来のマルコフ連鎖モンテカルロ法におけるサンプリング方法よりも収束が速いという利点がある．加えて，HMC 法を用いることによって分析者は統計的なモデル構成に専念できるというメリットがある．従来のギブスサンプリングでは，変数ごとの全条件付き事後分布を導出することが必要であった．これにはしばしば高度な数理統計学的知識が要求されるため，理論的関心が方法論の数理にあるわけではないデータ分析者のギブスサンプリングを用いた階層ベイズ法の利用を遠ざける一因となった．HMC 法では対数尤度の導関数が得られていればよく，比較的数理的障壁は低い．さらに実際の計算では，モデルにおける必要な導関数はソフトウェアによって与えられるため，分析者はモデル構成に傾注することが可能となる．

分析 1 におけるべき関数分布の母数の推定には HMC 法を実装した Stan [8] と統計解析環境 R を利用した．計算における具体的な設定は連鎖の構成を 1 とし，事後分布から 11,000 回サンプリングを行い，最初の 1,000 回を破棄して，残りの 10,000 個の標本を用いた．事前分布には無情報を仮定し，定義域を十分広げた一様分布を用いた．なお，計算の効率化を図り，データには全 613,645 トランザクションから 10,000 トランザクションをランダムサンプリングしたものを利用した．

表 1　母数の最尤推定値

理論分布	母数の推定値	
指数分布	$\hat{\lambda} = 71.042$	
対数正規分布	$\hat{\mu}^* = 3.003$	$\hat{\sigma}^{*2} = 1.607$
第 1 種パレート分布	$\hat{\alpha} = 1$	$\hat{\beta} = 0.333$

表 2　べき関数分布の母数に関するベイズ推定結果

母数	MAP	post.sd	95%下側	95%上側	$\hat{R}$
β	886.001	0.326	886.006	887.179	1.000
γ	0.267	0.003	0.261	0.272	1.000

表 3　4 種類の理論分布における AIC

理論分布	AIC
指数分布	6629358
対数正規分布	6168086
第 1 種パレート分布	6428113
べき関数分布	6685800

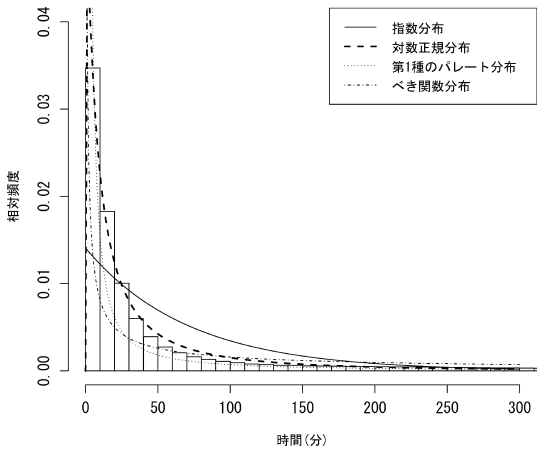

図 3　相対度数分布の上に推定値を用いて描画した 4 種類の理論分布（実線：指数分布，太い破線：対数正規分布，点線：第 1 種パレート分布，1 点破線：べき関数分布）

2.1.4 結果

上述した 4 種類の理論分布について，母数の最尤推定値を表 1 に，べき関数分布の母数の MAP 推定値 (MAP) と事後標準偏差 (post.sd)，95%確信区間 (95%下側 −95%上側)，収束判定指標 $\hat{R}$ をそれぞれ表 2 に示す．また，これらの推定値を用いて描画した 4 種類の理論分布を図 3 にまとめて示し，推定値を用いて算出した情報量規準 AIC を表 3 に示す．

2.2 考察

HMC 法の収束判定のために，Gelman and Rubin [9] が提案した $\hat{R}$ を用いた．$\hat{R}$ は 1 に近ければ収束，そうでなければ非収束であると解釈する．具体的な目安として，$\hat{R}$ が 1.2 ないし，1.1 よりも小さければ収束したと判断する基準が Gelman [10] で提唱されている．表 2 で示された母数に関して，$\hat{R}$ は 1.1 以内に収まっており，不変分布への収束が示唆される．

図 3 を目視で確認したところ，対数正規分布，第 1 種のパレート分布，べき関数分布，指数分布の順に当てはまりがよい印象を受ける．次いで，AIC を確認すると対数正規分布，第 1 種パレート分布，指数分布，べき関数分布の順に当てはまりのよさが示唆された．

結論として，分析 1 では，通常ユーザのチェック・インから購買までの時間の相対度数分布に最も当てはまりのよい理論分布として対数正規分布を得た．すなわち，通常ユーザのチェック・インから購買までの時間の分布は，顧客がシステムに到着してから去るまでの時間であるサービス時間の分布を表していることが示唆される．

3. 分析 2

3.1 目的

図 2 より，ダイヤモンド・ユーザのチェック・インから購買までの時間の相対度数分布は通常ユーザの分布よりも複雑な形状をしており，分析 1 のような単一分布でのモデル化は難しいことが予想される．また，図 1 と図 2 の相対度数分布の比較から，ダイヤモンド・ユーザの相対度数分布の形状は通常ユーザの分布とその他二つのダイヤモンド・ユーザに顕著な単峰分布との混合分布であることが示唆される．この混合分布がライフサイクルの観点から以下に仮定される三つの異なる消費者行動をとるユーザの行動結果が一つにまとめられたものである可能性が考えられることから，分析 2 ではダイヤモンド・ユーザに異なる消費者行動をとる三つの潜在クラスを仮定し，潜在混合分布モデルを用いて，ダイヤモンド・ユーザにおけるモデル構造を明らかにする．

消費者行動 1：店舗に入ってからチェックインを行い，そのまま購買を行う．

消費者行動 2：出勤時にチェック・インを行い，昼休みに購買を行う．

消費者行動 3：出勤時にチェック・インを行い，退勤後に購買を行う．

分析 1 で得られた通常ユーザの分布は主に消費者行動 1 を示すが（セグメント 1），ダイヤモンド・ユーザにおいては，消費者行動 2 と消費者行動 3 を示す割合が通常ユーザよりも多いと考え，それぞれ，ダイヤモンド・ユーザに顕著なセグメント 2 とセグメント 3 を構成すると仮定した．なお，これら三つの消費者行動の仮定が正しかったかについては後の分析 3 において，三つのセグメントの特徴を明らかにすることで検証される．

3.2 方法

3.2.1 潜在混合分布モデル（1 変量モデル）

ダイヤモンド・ユーザの各トランザクションにおいて，チェック・インから購買までの時間 x が得られたとき，データの確率密度分布 $p(x)$ は C 個の理論分布 $f_1(x|\boldsymbol{\theta}_1), f_2(x|\boldsymbol{\theta}_2), \ldots, f_C(x|\boldsymbol{\theta}_C)$ の重み付き線形結合

$$p(x) = \sum_{c=1}^{C} \pi_c f_c(x|\boldsymbol{\theta}_c) \tag{9}$$

によってモデル化することができる．ここで，π_c は各分布の混合パラメータ（混合比率）であり，$\boldsymbol{\theta}_c$ は各母集団を表す分布の母数ベクトルである．混合パラメータは以下の条件を満たす．

$$\sum_{c=1}^{C} \pi_c = 1, \quad 0 \leq \pi_c \leq 1 \tag{10}$$

3.2.2 候補となる分布の組み合わせ

本研究では，前述の 3 種類のセグメントの仮定から，クラス数 $C = 3$ で分析を行った．分析 1 の結果を利用し，通常ユーザの消費者行動を代表する分布（セグメント 1）には対数正規分布 $LN(\mu_1^*, \sigma_1^{*2})$ を仮定し，母数は再度推定することとした．

ダイヤモンド・ユーザに顕著なセグメントであるセグメント 2 とセグメント 3 を構成する分布に関しては以下の 3 種類のモデルを考えた．

モデル 1：ガンマ分布 $\Gamma(\alpha_2, \beta_2)$ ＋ ガンマ分布 $\Gamma(\alpha_3, \beta_3)$

モデル 2：正規分布 $N(\mu_2, \sigma_2^2)$＋ 正規分布 $N(\mu_3, \sigma_3^2)$

モデル 3：対数正規分布 $LN(\mu_2^*, \sigma_2^{*2})$ ＋ 対数正規分布 $LN(\mu_3^*, \sigma_3^{*2})$

モデル 1 のガンマ分布の組み合わせに関しては，基準変数が時間であることから，定義域が正であり，指数分布と同じく待ち時間の分布として利用されるガンマ分布を適用した．前述の消費者行動の仮定から，ダイヤモンド・ユーザに顕著なセグメントを構成する分布は y 軸から離れたところにモードをもつことが想定される．指数分布では実現不可能なこのような特徴を

ガンマ分布は表現することができる.

モデル 2 の正規分布の組み合わせに関しては，混合分布モデルに最も頻繁に仮定される分布であることを鑑みた．また，ダイヤモンド・ユーザに顕著なセグメントは習慣的な特定の消費者行動を示すと仮定し，正規分布で習慣性に散らばりを含めたモデルを構成できると考えた.

モデル 3 の対数正規分布の組み合わせに関しては，分析 1 によって示唆された，チェック・インから購買までの時間は，サービス時間を表す対数正規分布によって上手く説明されるという結論をダイヤモンド・ユーザに顕著なセグメントに応用した.

3.2.3 事前分布

本研究では，前述した三つの消費者行動を事前情報として用いることにより，ベイズ分析での潜在混合分布モデルにおいて問題となるラベルスイッチングに対処した.

具体的には，三つのモデルごとに母数の関数で表現されるモードが，消費者行動 2 を仮定した分布においては出勤時から昼休みまでが 4 時間程度と考えられることから，$x = 240$ 付近であるように，消費者行動 3 を仮定した分布においては出勤時から退勤時までが 9 時間程度と考えられることから，$x = 520$ 付近であるように事前分布を設定した．ただし，事前情報が強くなることは好ましくないので，分散は比較的大きい値を設定した．たとえば，モデル 2 においては，以下のとおりである.

$$\mu_2 \sim N(240, 100^2), \quad \sigma_2^2 \sim \Gamma^{-1}(19, 1800)$$
$$\mu_3 \sim N(520, 100^2), \quad \sigma_3^2 \sim \Gamma^{-1}(19, 1800)$$

ここで，正規分布の分散には以下の確率密度をもつ逆ガンマ分布を事前分布として用いた.

$$\Gamma^{-1}(\sigma^2|\alpha, \beta) = \frac{\beta}{\Gamma(\alpha)}\sigma^{2-(\alpha+1)}e^{-\beta/\sigma^2}, \quad \sigma^2 > 0$$

逆ガンマ分布のモードと分散は解析的に以下で求められる.

$$\text{mode} = \frac{\beta}{\alpha+1}, \quad \text{Var} = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$$

すなわち，モデル 2 においては，チェック・インから購買までの時間の定義域を考慮し，σ^2 のモードが 90，分散が 580 となるような事前分布を仮定した.

3.2.4 推定方法

2013 年 5 月から 9 月までのダイヤモンド・ユーザのデータを用い，上述した三つのモデルの母数を HMC 法でベイズ推定した．それと同時に，ベイズ分析におけるモデル比較の指標である WAIC (Watanabe-Akaike information criterion) [11] を計算した.

分析 1 のべき関数分布と同様，母数の推定には HMC 法を実装した Stan と統計解析環境 R を利用した．事後分布から 5,000 回サンプリングを行い，最初の 1,000 回を破棄して，残りの 4,000 個の標本を用いた.

表 4 モデル 1 の母数に関する推定結果

母数	EAP	post.sd	95%下側	95%上側	$\hat{R}$
π_1	0.587	0.005	0.577	0.598	1.000
π_2	0.302	0.016	0.269	0.333	1.000
π_3	0.111	0.015	0.101	0.141	1.000
μ_1^*	3.035	0.020	2.998	3.031	1.000
σ_1^*	1.413	0.013	1.386	1.439	1.000
α_2	6.978	0.491	6.071	7.988	1.000
β_2	0.018	0.016	0.017	0.019	1.000
α_3	15.607	0.969	13.754	17.589	1.001
β_3	0.069	0.005	0.060	0.079	1.001

表 5 モデル 2 の母数に関する推定結果

母数	EAP	post.sd	95%下側	95%上側	$\hat{R}$
π_1	0.574	0.005	0.563	0.584	1.000
π_2	0.252	0.011	0.229	0.273	1.000
π_3	0.174	0.010	0.155	0.196	1.002
μ_1^*	2.999	0.021	2.960	3.038	1.000
σ_1^*	1.405	0.013	1.379	1.431	1.002
μ_2	249.531	2.980	243.810	255.412	1.000
σ_2	80.803	2.185	76.651	85.046	1.001
μ_3	474.929	9.417	455.660	493.147	1.002
σ_3	130.449	4.656	121.271	139.195	1.003

3.3 結果

三つのモデルの母数の EAP 推定値 (EAP) と事後標準偏差 (post.sd)，95% 確信区間（95%下側−95%上側），収束判定指標 $\hat{R}$ をそれぞれ表 4, 5, 6 に示す.

表 6 より，モデル 3 は収束判定指標 $\hat{R}$ が 1.1 に収まっておらず，事後標準偏差も 0.000 と不可解な値を示していることから，マルコフ連鎖は非収束であることが示唆される.

ダイヤモンド・ユーザのチェック・インから購買までの時間のヒストグラムの上に，母数の推定値を用いて，推定に成功したモデル 1 とモデル 2 の分布を描いた．このとき二つのモデルで図 4 のような確率足し上げ図を同じように得た[3]．この 2 種類のモデルにおけ

[3] モデル 1 の図は割愛する.

表 6　モデル 3 の母数に関する推定結果

母数	EAP	post.sd	95%下側	95%上側	$\hat{R}$
π_1	0.595	0.000	0.595	0.595	1.365
π_2	0.001	0.000	0.001	0.001	2.988
π_3	0.404	0.000	0.404	0.404	1.462
μ_1^*	3.064	0.001	3.063	3.065	1.343
σ_1^*	1.421	0.001	1.420	1.422	1.422
μ_2^*	5.352	0.000	5.352	5.352	1.000
σ_2^*	0.000	0.000	0.000	0.000	1.273
μ_3^*	5.753	0.001	5.751	5.755	1.616
σ_3^*	0.429	0.000	0.428	0.429	1.943

表 7　モデル 1，モデル 2 における WAIC

	WAIC
モデル 1	1610407
モデル 2	1038202

る WAIC を表 7 に示す．

3.4　考察

モデル 1 とモデル 2 の確率足し上げ図を目視で確認したところ，明確にどちらが当てはまりがよいか判断することができなかった．そこで WAIC を用いて，モデル比較を行った．表 7 より，モデル 2 のほうが当てはまりのよいことが示唆された．

結論として，分析 2 ではダイヤモンド・ユーザのチェック・インから購買までの時間の相対度数分布に最も当てはまりのよい混合分布モデルとして，モデル 2（対数正規分布と二つの正規分布）が得られた．ダイヤモンド・ユーザに顕著な二つのセグメントはそれぞれ $N(249.5, 80.0^2)$ と $N(474.9, 130.4^2)$ によって構成され，セグメント 2・3 においては，習慣的な消費者行動が散らばりをもって正規モデル化されることが示唆される．

セグメント 1（消費者行動 1）に所属するダイヤモンド・ユーザの比率は 57% 程度と最も大きく，セグメント 2（消費者行動 2）は 25% 程度，セグメント 3（消費者行動 3）は 17% 程度であった．このように，ダイヤモンド・ユーザにおいては，ライフサイクルに依存した消費者行動（消費者行動 2 と 3）が仮定されるセグメントは全体の 40% 程度を占めており，これはチェック・インが購買を促す潜在的な動機づけの役割を果たしていることを示唆する．また，チェック・インから購買までの時間が長くなるにつれ，セグメントの比率は減少していることから，チェック・インした記憶が保持されやすい期間ほど，購買に結びつくと解釈できる．

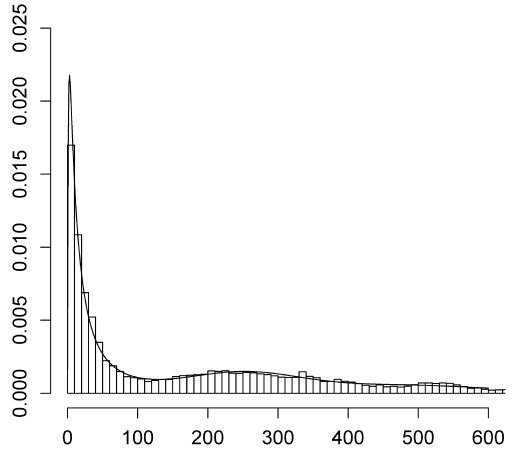

図 4　相対度数分布の上に推定値を用いて描いたモデル 2（確率足し上げ）

4.　分析 3

4.1　目的

潜在混合分布モデルの最終的な目的は，得られたデータが所属する母集団の特徴を明らかにすることである．よって，分析 3 では，ダイヤモンド・ユーザに顕著なセグメントの特徴を明らかにする方法を論じ，ユーザのライフスタイルを加味した分析を行う．

4.2　方法

4.2.1　線形予測子を利用した混合ロジットモデル

分析 2 から得られたモデル 2 を採用し，潜在混合分布モデルの混合比率 π_c に線形予測子を組み込む．具体的には，混合比率を以下のように指定する．

$$\pi_1 = \frac{1}{1+\exp(b_{02}+b_{12}x_1+\cdots)+\exp(b_{03}+b_{13}x_1+\cdots)} \tag{11}$$

$$\pi_2 = \frac{\exp(b_{02}+b_{12}x_1+\cdots)}{1+\exp(b_{02}+b_{12}x_1+\cdots)+\exp(b_{03}+b_{13}x_1+\cdots)} \tag{12}$$

$$\pi_3 = \frac{\exp(b_{03}+b_{13}x_1+\cdots)}{1+\exp(b_{02}+b_{12}x_1+\cdots)+\exp(b_{03}+b_{13}x_1+\cdots)} \tag{13}$$

ここで，セグメント 1 はベースラインとなっており，すべての係数の値は 0 である．予測変数には，6 時から 24 時までの 18 時間における 3 時間ごとの「チェック・インした時間帯（x_1〜x_6）」，「性別（x_7：男性=1，女性=0）」，「クレジットカード（MUJI カード）の有無 (x_8)」をそれぞれ 0/1 のカテゴリカル変数として用いた．

表 8　線形予測子の係数の推定値

母数	EAP	post.sd	95%下側	95%上側	$\hat{R}$
b_{02}	−0.911	0.481	−1.815	0.026	1.004
b_{12}	3.137	0.586	1.986	4.299	1.001
b_{22}	−0.450	0.487	−1.389	0.475	1.005
b_{32}	−0.208	0.484	−1.149	0.701	1.004
b_{42}	−0.925	0.485	−1.871	−0.015	1.004
b_{52}	−13.140	3.971	−19.647	−6.180	1.000
b_{62}	−12.300	4.748	−19.706	−3.897	1.000
b_{72}	−0.811	0.093	−0.998	−0.640	1.000
b_{82}	0.892	0.056	0.785	1.002	1.002
b_{03}	−1.184	0.511	−2.173	−0.163	1.002
b_{13}	5.014	0.600	3.859	6.197	1.003
b_{23}	0.019	0.515	−1.018	1.015	1.002
b_{33}	−1.070	0.516	−2.107	−0.079	1.002
b_{43}	−13.639	3.807	−19.711	−6.942	1.004
b_{53}	−13.464	3.811	−19.608	−6.609	1.001
b_{63}	−11.883	4.782	−19.597	−3.661	1.001
b_{73}	−0.188	0.100	−0.386	0.001	1.002
b_{83}	0.727	0.076	0.581	0.875	1.000

4.2.2　推定方法

2013 年 10 月から 2014 年 2 月までのダイヤモンド・ユーザのデータを用い，混合分布の母数は分析 2 で特定した値で固定し，切片と回帰係数のみを HMC 推定することとした．2013 年 5 月から 9 月までのダイヤモンド・ユーザのデータを用い，上述したモデルの母数を HMC 法でベイズ推定した．母数の推定には HMC 法を実装した Stan と統計解析環境 R を利用した．事後分布から 20,000 回サンプリングを行い，最初の 10,000 回を破棄して，残りの 10,000 個の標本を用いた．

4.3　結果

線形予測子の係数の推定値を表 8 に示す．収束判定指標である $\hat{R}$ はすべて 1.1 以内に収まっており，不変分布への収束が示唆される．

4.4　考察

表 8 より，通常のメンバー・ユーザを代表するセグメント 1 と比較して，ダイヤモンド・ユーザに顕著であるセグメント 2・3 は x_1 に対してそれぞれ $b_{12} = 3.137, b_{13} = 5.014$ と高い係数の値であることがわかる．ここから，ダイヤモンド・ユーザに顕著なセグメントに所属するユーザは朝 6 時から 9 時の間にチェック・インする傾向があることが示唆される．すなわち，朝の通勤・通学時間などに計画的にチェック・インしていると考えられる．x_5, x_6 に対する係数については，$b_{52} = -13.140, b_{62} = -12.300$ のように高い負の値を示していることから，夕方を超えてからチェック・インするユーザはセグメント 2・3 には少ない傾向であることがわかる．これらは前述の消費者行動の仮定に合致する．

また，x_8 に対しても，$b_{82} = 0.892$, $b_{83} = 0.727$ とやや高い値であることから，セグメント 1 と比較して，セグメント 2・3 に属するユーザは MUJI カードを持っている傾向が高いことが示唆される．

5.　分析 4

5.1　目的

分析 3 で明らかになったモデル構造を利用し，実際にダイヤモンド・ユーザに顕著なセグメントの特徴をもつ通常ユーザを掘り起こすことで，セグメント別の購買に対する相対的な効率について検証する．

5.2　方法

分析 2 で推定された混合分布の母数の値と分析 3 で推定された切片と回帰係数を用いて，2014 年 3 月から 6 月までの通常ユーザのチェック・インから購買までの時間の一つひとつのトランザクションの所属確率を事後対数尤度を用いて算出した．この中から，5 回以上のトランザクションをもつユーザについて，最も事後対数尤度が高いセグメントを所属セグメントとして特定した．

最後に，通常顧客の消費者行動を表すセグメント 1 に属するユーザと優良顧客に顕著なセグメント 2・3 に属するユーザの 4 カ月の総購買額を比較し，セグメント 2・3 の購買に対する相対的な効率について検証した．

5.3　結果

セグメント 1 に所属するユーザとセグメント 2・3 に属するユーザの 4 カ月の総購買額を密度曲線で描いたものを図 5 に示す．

5.4　考察

図 5 より，実線で描かれたセグメント 1 に属する通常ユーザより，破線で描かれたセグメント 2・3 に属する通常ユーザのほうが購買に対する相対効果が高いことが見て取れる．セグメント 2・3 に属する通常ユーザは計 1,042 人であり，セグメント 1 に属するユーザの約 1/10 の人数であった．しかし，同じ人数で比較したときの総購買額はセグメント 2・3 のほうが約 700 万円多かった．よって，限られたプロモーション費用を有効利用するのであれば，セグメント 2・3 に属するユーザにアプローチすることが推奨される．

6.　おわりに

本研究では，チェック・インから購買までの時間を利用し，通常のメンバー・ユーザのセグメントを構成す

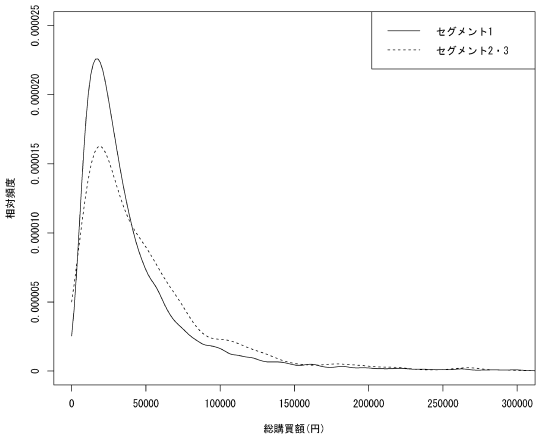

図 5 セグメント 1 とセグメント 2・3 に属するユーザの 4 カ月の総購買額（実線：セグメント 1，破線：セグメント 2・3）

る分布として対数正規分布を，ダイヤモンド・ユーザに顕著なセグメントを構成する分布として二つの正規分布を混合分布モデルを用いて特定した．加えて，混合分布モデルの混合比率に線形予測子を組み込むことによって，ダイヤモンド・ユーザに顕著なセグメントが通勤の時間帯に計画的にチェック・インする傾向があることを明らかにすることができた．交差検証法においては，ダイヤモンド・ユーザに顕著なセグメントに所属する顧客は購買額において相対的に効率の高い結果を示しており，一連の分析の妥当性が示された．

参考文献

[1] P. Kotler and K. L. Keller, *Marketing Management*, 12th edition, Prentice Hall Inc., 2006.（恩蔵直人，月谷真紀，『コトラー & ケラーのマーケティング・マネジメント（第 12 版）』，ピアソン・エデュケーション，2008.）

[2]（株）良品計画，「MUJI passport」，http://www.muji.net/passport/（2015 年 5 月 15 日閲覧）

[3] 蓑谷千凰彦，『統計分布ハンドブック』，朝倉書店，2004.

[4] H. Ebbinghaus, *Über das Gedächtnis*, Duncker & Humbolt, 1885.

[5] S. Chakrabortya, K. Muthuramanb and M. Lawleyc, “Sequential clinical scheduling with patient no-shows and general service time distributions,” *IIE Transactions*, **42**, pp. 354–366, 2010.

[6] C. M. Harris, “The Pareto distibution as a queue service discipline,” *Operations Research*, **16**, pp. 307–313, 1968.

[7] S. Duane, A. D. Kennedy, J. B. Pendleton and D. Roweth, “Hybrid Monte Carlo,” *Physics Letters B*, **195**, pp. 216–222, 1990.

[8] Stan Development Team, *Stan Modeling Language User’s Guide and Reference Manual*, Version 2.5.0, 2015.

[9] A. Gelman and D. B. Rubin, “Inference from iterative simulation using multiple sequences (with discussion),” *Statistical Science*, **7**, pp. 457–511, 1992.

[10] A. Gelman, “Inference and monitoring convergence,” *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.), Chapman & Hall, pp. 131–143, 1996.

[11] S. Watanabe,“Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory,” *Journal of Machine Learning Research*, **11**, pp. 3571–3591, 2010.