

〔論文〕

# マーケティングにおける購買間隔モデルの展望

—ベイズモデリングによる実証分析—

猪狩良介

## 1 はじめに

近年、マーケティング・データ環境が非常に豊かになってきており、店頭における商品の購買履歴や、パソコンを利用したインターネット上でのサイト閲覧・購買履歴、スマートフォンを利用したアプリ利用履歴や位置情報など、多種多様なデータが利用可能になってきている。企業はこれらのデータを有効活用して、マーケティング意思決定や消費者との関係強化などをはかっている。特に購買履歴データに消費者IDと属性を付与したID付POSデータは、いつ、誰が、何を、いくらで、何個購入したのかという経時的な購買履歴を記録しており、これを分析することでどのような属性の消費者がどの商品をどれだけの頻度で購入するのかなどを把握することができる。ID付POSデータは、チラシの配布や値引きプロモーション等の店舗側の情報を入手できるという利点があるが、特定の店舗における購買履歴を収集しているため競合店舗における購買履歴は一般的には記録されていない。そのため、競合店舗の利用などを分析するにはスキャン・パネルデータが利用される。スキャン・パネルデータは、店舗ではなく消費者の購買行動の把握に焦点を当てているため、自社店舗だけではなく競合店舗における購買履歴も記録されているのが特徴である。このようなID付POSデータやスキャン・パネルデータを利用して、消費者の購買行動の解明を目的とした研究がマーケティング・サイエンスでは蓄積されてきている。

本研究では消費者の購買時期のモデル化に焦

点を当てる。購買時期とは、消費者が前回商品を購買してからどのくらい時間が経過してから再度商品を購買するのかわかり、購買時期の分析は購買間隔モデル (Interpurchase-Timing Model) や購買タイミングモデル (Purchase-Timing Model) などと呼ばれている。購買間隔モデルはマーケティング・サイエンス分野では重要なテーマの一つであり、販促プロモーションにより他社への離反などを防いだり、自社店舗への来店時期を早めることができる、などをデータから明らかにすることができる。本研究では、購買間隔データの特性について整理するとともに、購買間隔データを扱う統計モデルについて紹介する。購買間隔モデルでは、時間の長さを従属変数とした解析を行うため、生存時間解析が一般的に利用される。生存時間解析の利点として、時間データに特徴的な打ち切り構造を考慮できる、価格やプロモーションといったマーケティング変数の効果を含めることができる、といった点があげられる。また、本研究は購買間隔モデルの系譜と発展についてのレビューを行うとともに体系的にまとめることも目的としている。なお、顧客関係性管理においても、消費者の購買頻度を扱うことがあり、購買間隔モデルと非常に近い分野ではあるが、本研究では顧客関係性管理における諸モデルは割愛する。また、実証分析では近年のマーケティング分析でよく利用されるベイズ統計学を用いた様々な購買間隔モデルを実際のマーケティングデータに適用し、考察を行う。

## 2 購買間隔データと統計モデル

### 2.1 購買間隔の特性

購買間隔モデルでは、前回に購買した時点から次に購買するまでの経過時間 (Duration Time) を分析対象 (従属変数) とし、多くは経過日数を扱う。図1に購買間隔のイメージを示す。矢印の範囲の観測期間は、例えば購買履歴データを記録している期間であり、それ以外の期間に製品をいつ購入したかをデータから判断することはできない。例えば、顧客1は観測開始前のどのタイミングで購買をしているかは

わからない。一方で顧客3は、観測終了時点で商品を購入していないが、その後すぐに購買している可能性もあるし、逆にその後は購買していない可能性もある。このような観測期間外で購買が発生している場合を打ち切りデータ (Censored Data) といい、顧客1のような状況を左打ち切り (Left Censored)、顧客3のような状況を右打ち切り (Right Censored) と呼ぶ。

次に、購買間隔データの特性を示す。図2は Igari and Hoshino (2017) で用いられた tissue/toilet paper 商品の購買間隔の分布である。購

図1 購買間隔のイメージ

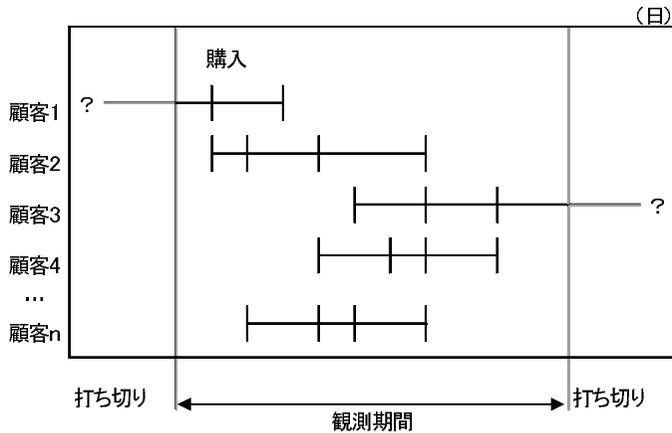
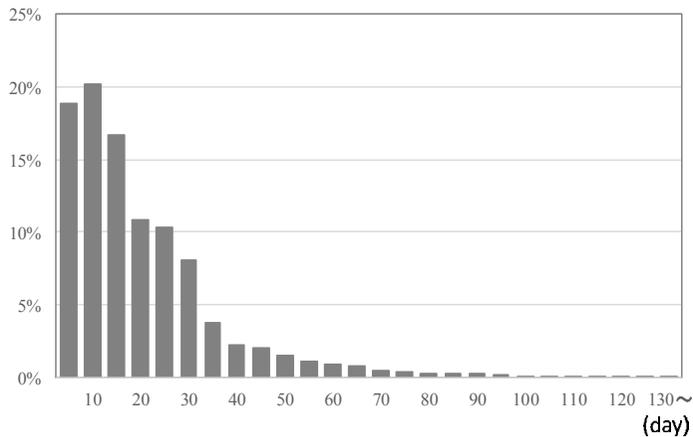


図2 Tissue/Toilet Paper 商品の購買間隔の分布 (Igari and Hoshino, 2017 を転載)



買間隔は次に購買が発生するまでの継続時間であることから、非負である（0より大きいプラスの値をとる）。また、商品カテゴリの利用頻度に関係して比較的0の近くに購買が集中し、時間の経過が長くなるにつれて、購買は少なくなることが多い。図から、tissue/toilet paper 商品カテゴリは前回購買から10日前後の購買発生が多く、その後は時間の経過により購買発生回数は少なくなっており、右に歪んだ分布となっている。また、購買履歴データは観測期間が決まっており、期間外の購買発生はデータとして記録されていない。そのため、購買間隔データは打ち切りが生じてしまい、打ち切りを適切に扱う必要がある。

このような量的データを分析する際には、回帰分析の利用がまず考えられるが、購買間隔データに回帰分析を利用することには問題がある。まず、図2のように一般的に購買間隔のデータは正規分布に従わないことから、正規性を仮定している回帰分析の利用はふさわしくない。また、回帰分析では打ち切りの構造を考慮することはできない。次に、打ち切りを考慮したTobit回帰分析の利用が考えられるが、Tobit回帰分析も正規性を仮定している点では回帰分析と同じである。その他には、打ち切られたデータを除いた解析も考えられるが、その場合は選択バイアス（e.g., Heckman, 1977）等の問題が生じることが知られている。また、ある一定時間（例えば1ヶ月）に購買が発生したかどうかを従属変数とした2項選択モデル（ロジスティック回帰分析等）も考えられるが、その場合は継続時間の長さを予測できない、といった問題が生じる。そのため、購買間隔などの継続時間を従属変数とする場合は、生存時間解析が一般的に利用される。

## 2.2 生存時間解析

生存時間解析（Survival Analysis）は、イベントが発生するまでの時間を従属変数とした統計解析手法であり、生物統計、医学統計、経済学やマーケティングなど、さまざまな分野で応用されている。イベントとは興味のある事象であり、医学等では患者の死亡を、マーケティン

グでは消費者の購買を、それぞれイベント発生と定義することが多い。生存時間解析については、Klein and Moeschberger（2003）や Hosmer et al.（2011）などに詳しい。また、マーケティング分野において生存時間解析を広く紹介した文献としては、Chintagunta and Dong（2006）や里村（2015）、佐藤（2015）がある。

生存時間解析では、ハザード関数と生存関数の2つを考える。ハザード関数とは、時点 $t$ までイベントが発生していないという条件の下で次の瞬間にイベントが発生する度合いを表す。一方で、生存関数は時点 $t$ までイベントが発生しない確率を表す。本研究では生存時間解析で最も代表的な方法である比例ハザードモデル（Proportional Hazard Model）を考える。

ハザード関数  $h(t)$  は、

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{p(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = h_0(t)\lambda, \quad (1)$$

と表す。ここで、 $h_0(t)$  は時間の経過によるイベントの発生のしやすさの基準を表すベースラインハザード関数である。また、一般的には  $x$  を共変量、 $\beta$  をその係数として、

$$\lambda = \exp\left(b + \sum_{k=1}^K x_k \beta_k\right), \quad (2)$$

とする。なお、係数  $\beta$  は、正の場合に購買時期を早め、負の場合に購買時期を遅くすると解釈する。そのため生存時間解析を利用した購買間隔モデルでは、価格は負の係数に、販促プロモーション等は正の係数になることが期待される。また、 $b$  は定数項に相当するが、モデル（仮定する確率分布）によっては  $\exp$  の中には含めず、ベースラインハザード関数に含めることもある。ベースラインハザード関数  $h_0(t)$  には確率分布を仮定したパラメトリックモデルと、特定の確率分布を仮定しないセミパラメトリックモデルが利用可能である。セミパラメトリックモデルでは、Cox 比例ハザードモデル（Cox, 1972）が一般的である。また、パラメトリックモデルでは、以下の確率分布がよく利用される。

- 指数分布： $h(t) = \lambda$
- ワイブル分布： $h(t) = at^{\alpha-1}\lambda$

• 対数ロジスティック分布： $h(t) = \frac{at^{\alpha-1}\lambda}{1+\lambda t^{\alpha}}$

• 指数べき分布： $h(t) = at^{\alpha-1}\lambda^{\alpha} \exp((\lambda t)^{\alpha})$

1つめの指数分布のハザード関数は時間  $t$  には依存しない。つまり、ある時点におけるイベントの発生のしやすさは、すでにどれだけの時間が経過したかには影響されない<sup>1)</sup> モデルであり、分析対象がこれに適しているか注意する必要がある。一方で、ワイブル分布や対数ロジスティック分布、指数べき分布は、形状パラメータ  $\alpha$  により、時間の経過によるイベントの発生のしやすさを表現することができる。

生存関数  $S(t)$  は、

$$S(t) = \exp\left\{-\int_0^t h(u) du\right\}, \quad (3)$$

と表現される。

ここで、ハザード関数と生存関数を掛け合わせて確率密度関数  $f(t)$  を計算する。

$$f(t) = h(t)S(t). \quad (4)$$

消費者  $i (i=1, \dots, n)$  の購買間隔  $t_i$  に対する比例ハザードモデルの尤度関数は、一般的に右打ち切りを考慮して

$$L = \prod_{i=1}^n h(t_i)^{\delta_i} S(t_i), \quad (5)$$

と表現される。ここで、 $\delta$  は観測期間内にイベントが発生する場合に 1 を、観測期間内にイベントが発生しない場合には 0 をとる打ち切りインディケータである。母数の推定は、パラメトリックモデルでは最尤法を、セミパラメトリックモデルの場合は部分尤度法を一般的に利用する。

### 2.3 消費者異質性の導入とベイズ推定法

近年のマーケティングでは、消費者異質性が非常に重要となっており、消費者 1 人 1 人の母数を推定することが多い (e.g., Rossi *et al.* 1996; 阿部・近藤, 2005; 照井, 2008)。通常、このような連続的な異質性を捉えるためには、変量効果モデルやマルチレベルモデル、一般化線形混合モデル等が利用される。生存時間解析では、特に個人  $i$  固有の母数 (定数項に相当) を含むモデルを Frailty モデルと呼ぶ。また、

特にマーケティングでは、共変量の係数にも異質性を認めるモデルも利用されている。

Frailty モデルでは、

$$h(t_i) = h_0(t_i)\omega_i\lambda_i, \quad \lambda_i = \exp\left(\sum_{k=1}^K x_{ik}\beta_k\right), \quad (6)$$

として、 $\omega_i (>0)$  をハザード関数に含めることで異質性を考慮可能である。なお、 $\omega_i$  は連続的な確率変数であり、ガンマ分布を仮定することが多い (e.g., Ibrahim *et al.*, 2005)。

また、 $\omega_i$  を  $\exp$  の中に含めることで、

$$h(t_i) = h_0(t_i)\lambda_i^*, \quad \lambda_i^* = \exp\left(b_i + \sum_{k=1}^K x_{ik}\beta_k\right), \quad (7)$$

と表現することもできる。ここで、 $-\infty < b_i < \infty$  となることから、 $b_i$  には正規分布を仮定することで扱いが容易になる (Ibrahim *et al.*, 2005)。

また、共変量の係数  $\beta_k$  にも異質性を導入し、

$$\lambda_i^* = \exp\left(b_i + \sum_{k=1}^K x_{ik}\beta_{ik}\right), \quad (8)$$

とするモデルもある。ここで、 $(b_i, \beta_{i1}, \dots, \beta_{iK})' = \beta_i$  として、一般的には多変量正規分布を仮定する。

また、 $b_i$  や  $\beta_i$  などの異質性をモデルに含める際はベイズ統計学を利用することが多い。ベイズ統計学では、事前分布と尤度関数から事後分布を構成し、一般的にはマルコフ連鎖モンテカルロ (Markov chain Monte Carlo; MCMC) 法を利用して母数の推論を行う。MCMC 法を利用することで、数値最適化を行わずに母数を推定することが可能となり、複雑なモデルにも応用しやすい。また、異質性を含めたモデルでは、尤度を計算する際に異質性 (潜在変数) を積分消去する必要があるが、MCMC 法ではデータ拡大法 (Tanner and Wei, 1987; Albert and Chib, 1993) を利用して、潜在変数も通常の母数と同様にサンプリングが可能であるという利点がある。ベイズ統計学については Gelman *et al.* (2013) が、生存時間解析のベイズ推定については Ibrahim *et al.* (2005) が詳しい。本研究では、消費者異質性を導入したベイズ統計学による購買間隔モデルに関する実証分析を行う。

### 3 購買間隔モデル

#### 3.1 購買間隔モデルの系譜

##### (1) 初期のアプローチ

最も初期の購買間隔の分析は、ある期間内に何回購入するかという購買頻度に対する確率分布のフィッティングである。Ehrenberg (1959) は購買頻度のデータに対して負の2項分布 (Negative Binomial Distribution; NBD) モデルを利用し、その後 NBD モデルは広く用いられるようになった。しかし、NBD モデルは購買頻度にポアソン分布を仮定しており、これを購買間隔モデルに適用すると指数分布を仮定していることになる。前述のように、指数分布ではある時点におけるイベントの発生のしやすさは、経過時間には影響されない (購買直後でも、1週間後でも変わらない) モデルであり、日用品の購買間隔に応用するには問題がある (Chatfield and Goodhard, 1973)。そこで、Herniter (1971) や Chatfield and Goodhard (1973) は購買間隔データに対して Erlang 分布<sup>2)</sup> を適用した。また、Gupta (1988) は Erlang 分布にマーケティング変数の効果を盛り込んだ購買間隔モデルを提案した。Gupta (1988) は、粉コーヒーを分析対象とし、チラシ、ディスプレイ、値引き、家庭内在庫等をマーケティング変数として用いた。結果として、チラシやディスプレイ、値引きは購買時期を早めるが、家庭内在庫は購買時期を遅くすることが確認された。

一方で Neslin *et al.* (1985) は、購買間隔を従属変数とした回帰分析を提案した。具体的には、前回購買からの経過時間を従属変数とし、クーポン利用、広告 (製造側と小売側の2つ)、価格、家庭内在庫等を独立変数とし、トイレットペーパーとコーヒー商品の購買履歴データを用いて分析を行った。彼らの仮説では、クーポンや広告等のプロモーションは購買時期を早め、家庭内在庫は購買時期を遅くするというものであり、一部のターゲットに絞った解析ではその仮説通りの結果となったが、必ずしも仮説通りではないことが確認された。彼らの研究では、購買間隔の長さを線形回帰分析で扱っているが、決定係数がいずれも 0.1 未満と非常に低

いことから、モデルの説明力に問題があり、その結果として仮説に反した結果が多く得られている可能性がある。前述のように、時間のデータを従属変数として扱う場合、打ち切りを適切に扱わないと推定結果にはバイアスが生じることが知られている。後に、Helsen and Schmittlein (1993) によって、打ち切りのある場合は回帰分析ではなく生存時間解析を利用すべきであると指摘されている。

##### (2) 生存時間解析による購買間隔モデルの一般化

生存時間解析を利用した購買間隔モデルとしては Jain and Vilcassim (1991) および Helsen and Schmittlein (1993) が有名であり、Seetharaman and Chintagunta (2003) は包括的な実証分析を行っている。

Jain and Vilcassim (1991) は、共変数として販促プロモーションといったマーケティング変数を利用し、観測されない異質性を含めた購買間隔モデルを提案した。Jain and Vilcassim (1991) のモデルは、生存時間解析を利用した今日の購買間隔モデルを一般化した研究とも考えられる。実証分析としては、スキャン・パネルデータを利用して、インスタントコーヒーと粉コーヒーの購買を分析した。マーケティング変数の効果としては、価格や前回の購買数量は購買時期を遅くすることが、またチラシは購買時期を概ね早くすることが確認された。しかし、クーポンについては2つのカテゴリで異なる結果が得られた。なお、指数分布や Erlang 分布、ワイブル分布では実際の購買間隔を表現するのは難しいことを検証しており、それらを包含する柔軟な Box-Cox 関数をベースラインハザード関数に採用している。

一方で Helsen and Schmittlein (1993) も同様に、生存時間解析を利用した購買間隔モデルを提案した。彼らの研究は、購買間隔モデルに生存時間解析とそれ以外の方法を用いた場合の整理と比較をしており、その中で生存時間解析の利点をまとめている点で、非常に重要である。具体的には、購買間隔モデルを捉える方法として、(1) 継続時間を従属変数とした回帰分

析、(2) 購買間隔がある長さ以上かそうでないかを従属変数とする2項選択モデル、(3) 生存時間解析(比例ハザードモデル)の3つを議論している。しかし彼らは、(1)の回帰分析では、打ち切りが存在する場合には独立変数の効果にバイアスがかかり、また独立変数が時間で変化する場合に回帰分析では対応できないと指摘している。一方で(2)の2項選択モデルでは期間の適切な区切りが不明な点や、個人の期待購買間隔を予測できないといった問題があるとしている。そこで、(3)生存時間解析を利用することで、打ち切り問題への対処や、時間によって変動する共変量を利用することができ、購買時期を正確に予測できると主張している。また、クラッカー商品に関するデータを利用した実証分析では、価格は購買時期を遅くするが、値引きは購買時期を早めることが確認された。

その後、マーケティング・サイエンスでは生存時間解析を利用した購買間隔モデルが主流になっていくことになり、様々な蓄積がされる。特に Seetharaman and Chintagunta (2003) は、様々な生存時間解析についての包括的な実証分析を行った。彼らは、(1)連続時間の生存時間解析と離散時間モデル(離散比例ハザードモデル)<sup>3)</sup>の比較、(2)様々な確率分布を利用したパラメトリックモデルの比較、(3)パラメトリックモデルとセミパラメトリックモデルの比較、(4)単一イベントモデルと競合リスクモデル<sup>4)</sup>の比較、(5)観測されない消費者異質性、の5つについて、洗濯洗剤およびペーパータオルの2つの商品カテゴリについて検証を行った。結果から、観測データを利用した場合は連続時間モデルよりも離散時間モデルの方が適していることが明らかになった。また、パラメトリックモデルでは、指数べき分布を仮定したモデルが、指数分布・Erlang分布・ワイブル分布・対数ロジスティック分布を仮定したモデルよりも適していることを示した。また、セミパラメトリックモデルを利用しても、パラメトリックモデルと比較して有用な知見は得られなかった。さらに(4)では、競合リスクモデルを利用して、前回の購買機会で購入した場合と

非購買の場合では、マーケティング変数の効果が異なるという新しい知見も得られた。彼らの研究は、生存時間解析を利用した購買間隔モデルの包括的な検証を行っている点で、非常に重要である。一方で、ベイズ推定法の導入と、マーケティング・サイエンスで主流となっている階層ベイズによる異質性のアプローチが今後の課題として挙げられている。

### (3) 消費者の異質性と繰り返し購買

購買間隔の長さは、消費者個人や世帯による異質性が存在すると考えられる。例えば、人数が多い世帯は間隔が短くなると想定され、単身世帯は長くなることが想定される。そのため、これまでのマーケティング・サイエンスの流れに沿い、購買間隔モデルにおいても異質性が組み込まれてきた。特に生存時間解析では、個人によって異なる効果を含めた Frailty モデルが標準的に利用されており、消費者(または世帯)毎の連続的な異質性を考慮するために利用されている。また、連続的な異質性の背後にも、連続的な確率分布を利用する場合と、離散的な分布を考慮する場合がある。Gupta (1991) は Erlang 分布を利用したモデルの背後に連続的な異質性を含めたモデルを提案している。また、Jain and Vilcassim (1991) は、個人によって異なる定数項が離散的な場合と連続的な場合の双方を比較しており、異質性の分布は商品カテゴリによって異なるため、離散的なモデルの方が良いと主張している。また、Allenby *et al.* (1999) は、一般化ガンマ分布を用いた生存時間解析に連続的な異質性を導入し、動的な変化を含めたモデルを提案している。近年のマーケティング・サイエンスでは階層ベイズモデルを利用して消費者の異質性を考慮したモデルが標準的に利用されており、Allenby *et al.* (1999) は購買間隔モデルにおいてこの形を導入した研究でもある。

また、潜在クラスモデルを利用してセグメントレベルの異質性を組み込んだモデルも提案されている。Seetharaman and Chintagunta (2003) は、前述のように様々なモデルを比較する中で、ベースラインハザード関数とマーケ

ティング変数の効果がセグメントによって異なるモデルを提案し、モデルの説明力が向上することを確認した。また、Vakratsas and Bass (2002) や猪狩・星野 (2016) は、購買間隔モデルの背後にセグメント別の異質性を仮定し、さらに消費者（世帯）属性を利用して各消費者（世帯）がどのセグメントに所属するのかを多項ロジットモデルにより予測する方法を提案している。セグメントへの所属をモデル化することで、潜在顧客の属性がわかれば、その人がどのセグメントに所属するかを予測することができる、という利点がある。

次に、日用品などの繰り返し購買される場合を考える。医学で患者の生存時間に焦点を当てるとその対象者の観測は終了する。しかし、マーケティングでは ID 付 POS データやスキャン・パネルデータを利用することから、その消費者の繰り返しの購買を捉えることができる。そのため、本研究で紹介している購買間隔モデルの文献のほとんどは、繰り返し購買を扱っている。しかし、これらは同一の消費者における複数の購買間隔は独立として扱うケースが多く、一度購買が発生したら継続時間は 0 から再スタートすることになる。そこで一般的には、消費者異質性を仮定し、Fraitly モデルによって繰り返しデータからその消費者独自の母数を推定したり、過去の購買間隔を共変量として利用することで考慮する。一方で、確率過程を利用して、過去の購買間隔を明示的に考慮する方法もある。Bijiwaard *et al.* (2006) は計数過程 (Counting Process) を利用した繰り返し購買間隔モデルを提案した。彼らのモデルは、購買間隔の各データを独立に扱うのではなく過去の購買履歴を情報として盛り込み、また観測開始からの経過時間やカレンダータイムを考慮できる点に特徴がある。

#### (4) 時間によって変動するマーケティング変数の扱い

購買間隔モデルでは、価格やクーポン、チラシといった様々なマーケティング変数が共変量として用いられてきた。性別や年齢といったデ

モグラフィック属性とは異なり、価格やクーポンなどの変数は日々変動する。このような変数は生存時間解析では時間依存共変量 (time-varying covariates) などと呼ばれているが、生存時間解析において時間依存共変量の扱いは難しい。なぜならば、生存時間解析では尤度を構成するために必要な生存関数の計算に、継続時間の長さの積分計算を含んでおり、連続的な時間依存共変量を利用することは通常難しいためである。そのため、購買間隔モデルでは、共変量であるマーケティング変数は、購買発生時点の価格等の情報を用いることが多い (e.g., Jain and Vilcassim, 1991; Franses and Paap, 2001)。

しかし、消費者は実際には価格の高い時期は購買を行わず、値引きされているときに購買すると考えるのが自然である。このような状況で有効なのは、離散比例ハザードモデル (Discrete-Time Proportional Hazard Model) または Grouped Hazard モデルと呼ばれる方法である。離散比例ハザードモデル (Grouped Hazard モデル) では、時間を週などの離散時間の単位に分割し、連続時間を離散時間の和に変換することで、生存関数の積分計算を和計算に変換する。

- 一般的な生存関数：

$$S(t) = \exp \left\{ - \int_0^t h(u) du \right\}$$

- 離散比例ハザードモデルの生存関数：

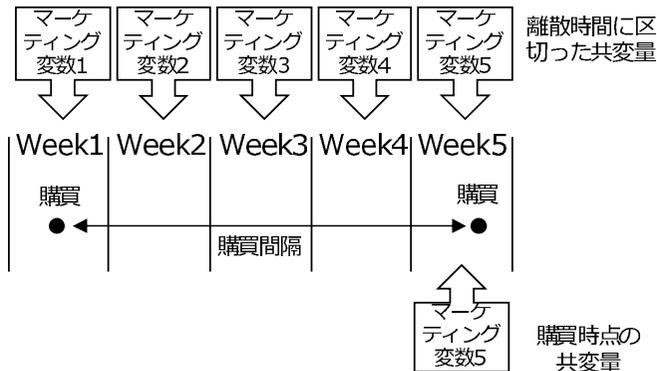
$$S(t) = \exp \left\{ - \sum_{u=1}^t \lambda(u) \int_{u-1}^u h_0(w) dw \right\}$$

ここで、 $\lambda(u) = \exp \left( b + \sum_{k=1}^K x_{uk} \beta_k \right)$

である。

例えば Seetharaman and Chintagunta (2003) は、離散比例ハザードモデルにより連続時間を週別に分割し、週別の価格等を共変量として利用しており、観測データに対しては離散比例ハザードモデルの方が精度が良いことを示している。また、Seetharaman (2004) は加法リスクモデル<sup>5)</sup>を利用した離散時間型のハザードモデルを提案している。Gupta (1991) や Fok *et al.* (2012) は、購買間隔を週別に分割し、時

図3 時間依存共変量のイメージ (Fok et al. 2012 を参考に筆者作成)



間によって変動するマーケティング変数を導入している（イメージを図3に示す）。特に、Fok *et al.* (2012) は、生存関数に過去のラグを含めることで、プロモーションの短期効果と長期効果を分離するモデルを提案している。一方で、ID付POSデータやスキャンパネルデータでは、購買が発生していない場合の価格等のマーケティング変数は得られていないことも多く、実際の分析では利用できないこともある。

### 3.2 購買間隔モデルの拡張

#### (1) 複数カテゴリ・ブランド選択を考慮した購買間隔モデル

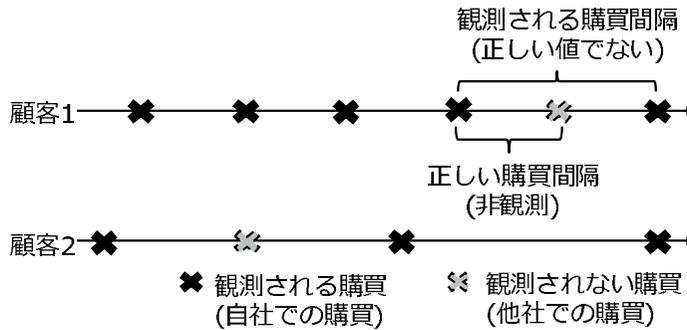
これまでに紹介した購買間隔モデルでは、分析対象を商品カテゴリとして単一の商品カテゴリの購買間隔を分析しているものが多い。しかし、メーカーや小売は複数の商品を提供していることが多く、一つの商品カテゴリだけではなく、複数の商品カテゴリで管理する必要がある。例えば、パスタ麺とパスタソースなどは、購買時期が双方に影響しあっており、双方を同時に管理する方が望ましい。このような複数商品の購買行動はブランド選択モデルでは広く研究されている（e.g., Seetharaman *et al.*, 2005）が、購買間隔モデルへの応用は多くない。

Chintagunta and Haldar (1998) は2つの関連する商品カテゴリの購買間隔モデルを提案した。彼らは、Farlie-Gumbel Morgenstern 族と

呼ばれる接合分布を利用して2つのカテゴリ間の関係性を考慮する生存時間解析モデルを提案した。彼らは、パスタとパスタソース、液体洗剤と粉洗剤、(缶詰) スープとヨーグルトという商品カテゴリのデータを利用して実証分析を行った。結果から、ポジティブに関係しあっている商品カテゴリ間では、片方の商品カテゴリの値引きが、当該カテゴリともう一方の商品カテゴリの購買確率をそれぞれ高めることが確認された。一方で、Park *et al.* (2014) は、もし消費者が複数のカテゴリを購買をしているのであれば、それは次の購買間隔に影響するとして、ハザード関数に過去の購買時点における商品カテゴリの選択を組み込んだモデルを提案している。結果から、多くの商品カテゴリでは、過去に購入していると間隔が長くなることが確認されたが、購買間隔への影響がない商品カテゴリも見られた。また、Fok and Paap (2009) は、購買間隔モデルは商品カテゴリ単位で扱うことが多いが、価格等のマーケティング変数の多くはブランド単位で得られていることから、ブランドレベルのマーケティング変数を利用しながら、カテゴリレベルの購買間隔を扱うモデルを提案している。具体的には、購買間隔とブランド選択の同時モデルを構築し、購買が発生しない場合の潜在的な消費者の選好を扱うモデルを提案している。

一方で、ブランド選択構造を組み込んだ購買

図4 不完全な購買間隔データのイメージ (Igari and Hoshino, 2018 を参考に作成)



間隔モデルも提案されている。Vilcassim and Jain (1991) および Gönül and Srinivasan (1993)、Chintagunta (1998) は、マルチステートモデル<sup>6)</sup>を利用して、ブランド選択を生存時間解析で表現している。これにより、ブランド間の競争構造を把握したり、消費者がバラエティシーキング<sup>7)</sup>かどうかなどを把握することができる。同様に、Wedel *et al.* (1995) は、時間を週別に区切った離散時間モデルにセグメント別の異質性を組み込み、ブランドスイッチングをマルチステートモデルによって表現するモデルを提案した。また、マルチステートモデルは Leszczyc *et al.* (2000) により、複数の競合店舗間の来店間隔の分析にも応用されている。また、猪狩・星野 (2016) はマルチステートモデルと類似している競合リスクモデル<sup>8)</sup>を利用して、複数のチャンネル (EC 購買と店頭購買) における購買間隔モデルを提案している。

**(2) 他社購買を考慮した購買間隔モデル**

店舗レベルの購買記録や、自社のデータベース上に記録されている購買履歴は、自社ブランドや自店舗における購買間隔のみを記録しており、競合他社や競合店舗における購買履歴は記録されていないことが多い。しかし、商品の購買間隔を考える場合、自社データベース上で記録されている購買記録の間にも競合店舗で商品を購入していれば、それは欠測イベントを含む不完全な購買間隔データを利用していることになり、そこから推定された分析結果は誤って

いる可能性が高い (図4 を参照)。しかし、自社データベース上では、どの時点に他社購買がおきているかを示す変数<sup>9)</sup>が記録されていないため、統計的に扱いが非常に難しい。

そこで、Chen and Steckel (2012) および Igari and Hoshino (2018) は、不完全な購買間隔データから本来の購買間隔モデルの母数を推定する方法を提案している。Chen and Steckel (2012) は、指数分布を仮定したハザードモデルをベースに、観測されている購買間隔が複数の購買間隔の和からなると考え、Erlang 分布で表現するモデルを提案している。一方で Igari and Hoshino (2018) は、観測されている購買間隔を、1つの購買間隔からなる観測値と2つ以上の購買間隔の和からなる観測値の2つに分離し、1つの購買間隔からなる観測値のデータから本来求める母数を推定するモデルを提案している。しかし、いずれもどの時点に他社購買がおきているかを示す変数が観測されておらず、観測データのみからは母数の推定は難しいことから、統計的データ融合を活用して、マクロレベルの情報をモデルに組み込むことで、適切にモデルを推定する方法を提案している。Igari and Hoshino (2018) の実証分析の結果では、他社購買を無視して購買間隔を分析すると、価格などのマーケティング変数の効果を過小評価することを示している。

**(3) インターネットの訪問間隔と購買間隔**

近年では、Web アクセスデータを利用した

インターネット上の消費者行動を捉える研究も多くされている。そこで、購買間隔モデルを応用し、インターネット上におけるサイト訪問時期やEC上での購買のタイミングを捉える研究もされている。Telang *et al.* (2004) は、これまでの購買間隔モデルの知見を、Webサイトの訪問間隔に応用した。Telang *et al.* (2004) は、自宅でのインターネット利用データを利用して、主要な検索エンジンへの訪問間隔に対して、離散的な異質性を考慮した比例ハザードモデルを適用した。また、ワイブル分布、指数べき分布、幾何分布を仮定したモデルの分析結果を比較したところ、指数べき分布を仮定したモデルが最も適しているという結果となった。これは、Seetharaman and Chintagunta (2003) による購買間隔モデルの結果と同じであり、インターネットの訪問間隔にもこれまでの購買間隔モデルの知見が応用できることを示している。

一方で、Moe and Fader (2004) や Park and Fader (2004)、山口 (2014) はインターネットサイトの訪問間隔のモデルに指数分布を仮定したモデルを提案した。指数分布は、時間の経過による購買のしやすさが変化しないため、店頭での購買にはふさわしくないとされており (Chatfield and Goodhard, 1973)、ワイブル分布や対数ロジスティック分布、指数べき分布など、時間の経過による購買のしやすさの変化を柔軟に表現できる確率分布が採用されてきた。しかし、インターネットでは、閲覧終了直後にも再度閲覧をすることは往々にして生じるため、店頭での購買間隔には適さないとされる指数分布の利用も、インターネットには当てはまると考えることができる点が興味深い。また、星野 (2013) や新美・星野 (2015) は、ワイブル分布を利用し、消費者の異質性を組み込んだインターネットサイトの訪問間隔モデルを提案している。

また、Manchanda *et al.* (2006) は、インターネット上における購買間隔モデルを提案している。特に彼らの研究は、インターネット上のバナー広告の露出が購買間隔に与える影響に焦点を当てており、Wedel *et al.* (1995) と同様の連続時間を離散時間に区切ったモデルを採

用して、時間依存共変量としてバナー広告への接触を盛り込んでいる。結果から、広告接触によりインターネット上の購買間隔が短くなることが確認された。

## 4 ベイズ統計学による購買間隔モデルの実証分析

### 4.1 モデル

実証分析では、ベイズ統計学を利用したさまざまな購買間隔モデルの比較を行う。Seetharaman and Chintagunta (2003) は購買間隔モデルについて体系的な分析を行っているが、ベイズ統計学を利用したアプローチと、消費者異質性への対処 (特に階層ベイズモデル) については議論していない。そこで、ベイズ統計学による消費者異質性を考慮した購買間隔モデルの実証分析を行う。ここでは、消費者 (世帯)  $i(i=1, \dots, n)$  の  $j(j=1, \dots, J_i)$  回目の購買における購買間隔を  $t_{ij}$  として、繰り返しのある購買間隔モデルを考える。

確率分布としては、[1] 指数分布、[2] Erlang-2 分布、[3] ワイブル分布、[4] 対数ロジスティック分布、[5] 指数べき分布の5種類を比較する。なお、Seetharaman and Chintagunta (2003) の結果を参考に、ノンパラメトリックなベースラインハザード関数を仮定したセミパラメトリックモデルは含めない。

$$[1] \text{ 指数分布 : } h(t_{ij}) = \lambda_{ij}$$

$$[2] \text{ Erlang-2 分布 : } h(t_{ij}) = \frac{\lambda_{ij}^2 t_{ij}}{1 + \lambda_{ij} t_{ij}}$$

$$[3] \text{ ワイブル分布 : } h(t_{ij}) = \alpha t_{ij}^{\alpha-1} \lambda_{ij}$$

$$[4] \text{ 対数ロジスティック分布 :}$$

$$h(t_{ij}) = \frac{\alpha t_{ij}^{\alpha-1} \lambda_{ij}}{1 + \lambda_{ij} t_{ij}^{\alpha}}$$

$$[5] \text{ 指数べき分布 :}$$

$$h(t_{ij}) = \alpha t_{ij}^{\alpha-1} \lambda_{ij}^{\alpha} \exp((\lambda_{ij} t_{ij})^{\alpha})$$

次に、異質性については、(1) 異質性を仮定しないモデル、(2) Frailty モデル、(3) 階層ベイズモデルの3種類を比較する。

$$(1) \lambda_{ij} = \exp\left(b + \sum_{k=1}^K x_{ijk} \beta_k + \sum_{l=1}^L w_{il} \gamma_l\right)$$

$$(2) \lambda_{ij} = \exp\left(b_i + \sum_{k=1}^K x_{ijk} \beta_k + \sum_{l=1}^L w_{il} \gamma_l\right),$$

$$b_i \sim N(\bar{b}, V_b)$$

$$(3) \lambda_{ij} = \exp\left(b_i + \sum_{k=1}^K x_{ijk} \beta_{ik} + \sum_{l=1}^L w_{il} \gamma_l\right),$$

$$(b_i, \beta_{i1}, \dots, \beta_{iK})' = \beta_i \sim MVN(\bar{\beta}, V_\beta)$$

ここで、 $x_{ij}$  は消費者  $i$  の購買機会  $j$  における購買時点の情報を用いたマーケティング変数であり、 $w_i$  は時点共通の消費者  $i$  の属性である。

(1) のモデルは、生存時間解析を利用した一般的な購買間隔モデルである。また、(2) は定数項  $b_i$  が消費者により異なり、それが正規分布に従う Frailty モデルである。一方で、(3) は、定数項に加えて、マーケティング変数の効果  $\beta_{ik}$  にも異質性を仮定し、それが多変量正規分布に従う階層ベイズモデルであり、マーケティング変数の効果が消費者によってどのように異なるのかを捉えることができるモデルである。

(2) および (3) のモデルの尤度は、消費者によって異なる母数  $b_i$  と  $\beta_i$  を積分消去して計算される。

$$L = \prod_{i=1}^n \int p(b_i | \bar{b}, V_b) \prod_{j=1}^{j_i} h(t_{ij} | \alpha, \beta, \gamma, b_i)^{\delta_{ij}}$$

$$\times S(t_{ij} | \alpha, \beta, \gamma, b_i) db_i, \tag{9}$$

$$L = \prod_{i=1}^n \int p(\beta_i | \bar{\beta}, V_\beta) \prod_{j=1}^{j_i} h(t_{ij} | \alpha, \gamma, \beta_i)^{\delta_{ij}}$$

$$\times S(t_{ij} | \alpha, \gamma, \beta_i) d\beta_i. \tag{10}$$

ここで、 $\beta = (\beta_1, \dots, \beta_K)'$ 、 $\gamma = (\gamma_1, \dots, \gamma_L)'$  である。また、 $\alpha$  は確率分布 [3] ~ [5] のみに含まれる。

なお、母数の推定はいずれも MCMC 法を利用する。事前分布としてはそれぞれ適切な確率分布を設定し、フラットな事前分布となるようにハイパーパラメータを設定した。いずれの確率分布を採用したモデルにおいても、母数  $\alpha$ 、 $b$ 、 $\beta$ 、 $\gamma$  については事後分布がよく知られた確率分布の形をしていないため、Metropolis-Hastings (MH) アルゴリズムを用いる。また、異質性モデルの (2) (3) では、潜在変数  $b_i$  および  $\beta_i$  のサンプリングには MH アルゴリズム

によるデータ拡大法を利用する。また、 $\bar{b}$ 、 $V_b$ 、 $\bar{\beta}$ 、 $V_\beta$  は事後分布がよく知られた確率分布の形をしているため、Gibbs Sampling を利用する<sup>10)</sup>。各モデルにおいて Geweke (1992) の方法によって MCMC の収束判定を実施し、収束後の 20,000 回の結果を母数の推論に利用する。

## 4.2 分析結果

実データ解析では、Franses and Paap (2001) で用いられた A.C.Nielsen 社のスキャン・パネルデータを利用する。また、「液体洗濯洗剤」商品カテゴリを分析対象とし、購買回数が 4 回以上の消費者を対象とした 209 世帯・計 2095 購買のデータを利用する<sup>11)</sup>。マーケティング変数  $x$  として、「過去の購買機会における購買量」「価格 (対数価格の前回との差分)」「チラシ」を用いる。また、消費者属性  $w$  としては「家族人数」を用いる。推定結果を表 1 に示す。表 1 は、ベイズ推定の事後平均および事後標準偏差 (SD) を表す。事後平均は、最尤推定値や最小二乗推定量に相当する。また、\* はベイズ信用区間 (Credible Interval) により 5% 水準で有意な結果であることを示している。

まずは、(1) 異質性なしモデルの結果を解釈する。「過去の購買量」はおおむねマイナス有意の結果が得られており、前回の購買機会が多く購買するほど、次の購買時期が遅くなるという、自然な結果が得られている。また、「価格」についてもおおむねマイナスの結果であり、価格が高くなると、購買時期が遅くなるという結果となっている。一方で、「チラシ」は Erlang-2 モデル、ワイブルモデル、指数べきモデルにおいてマイナス有意の結果となっている。セール時期などが決まっている場合、「チラシ」を見ることで購買を先延ばしにし、セール時期等に合わせて店舗に訪問する可能性があり、それによりマイナスの結果になっている可能性があると推察される。また、消費者属性の「家族人数」は全てのモデルでプラス有意となっており、家族人数が増えると消費量が増えるため、購買間隔が短くなると推察される。購買間隔モデルの結果を利用することで、消費者

表 1 推定結果

	指数		Erlang-2		ワイブル		対数ロジスティック		指数べき	
	事後平均	事後SD	事後平均	事後SD	事後平均	事後SD	事後平均	事後SD	事後平均	事後SD
<b>(1)異質性なしモデル</b>										
$\alpha$	—	—	—	—	1.124	0.015 *	1.923	0.037 *	0.712	0.011 *
$\beta$ 定数項	-3.958	0.070 *	-3.262	0.049 *	-4.507	0.096 *	-6.756	0.183 *	-4.639	0.053 *
過去の購買量	-0.003	0.001 *	-0.003	0.000 *	-0.003	0.001 *	-0.011	0.001 *	-0.002	0.001
価格(対数差)	-0.067	0.040	-0.070	0.028 *	-0.080	0.039 *	-0.099	0.069	-0.058	0.030
チラシ	-0.095	0.056	-0.096	0.039 *	-0.120	0.058 *	0.001	0.099	-0.133	0.048 *
$\gamma$ 家族人数	0.110	0.015 *	0.110	0.010 *	0.119	0.015 *	0.289	0.029 *	0.074	0.012 *
<b>(2)Frailtyモデル</b>										
$\alpha$	—	—	—	—	1.457	0.017 *	2.207	0.029 *	1.020	0.039 *
$\beta$ 過去の購買量	0.000	0.001	-0.003	0.001 *	-0.003	0.001 *	-0.011	0.001 *	0.000	0.001 *
価格(対数差)	-0.003	0.001 *	-0.050	0.030	-0.076	0.044	-0.107	0.070	-0.001	0.001
チラシ	-0.050	0.094	0.019	0.045	0.004	0.063	0.132	0.109	-0.007	0.082
$\gamma$ 家族人数	2.386	0.107 *	0.112	0.018 *	0.186	0.028 *	0.116	0.038 *	2.069	0.104 *
$b$ 異質性平均: 定数項	-6.360	0.050 *	-3.466	0.052 *	-6.350	0.078 *	-7.720	0.095 *	-6.655	0.056 *
異質性分散: 定数項	0.259	0.049 *	0.211	0.027 *	0.493	0.064 *	0.955	0.110 *	0.298	0.056 *
<b>(3)階層ベイズモデル</b>										
$\alpha$	—	—	—	—	1.687	0.026 *	2.651	0.049 *	1.169	0.026 *
$\gamma$ 家族人数	0.151	0.055 *	0.085	0.025 *	0.221	0.059 *	0.605	0.084 *	0.068	0.011 *
$\beta$ 異質性平均: 定数項	-4.023	0.178 *	-3.193	0.118 *	-7.070	0.243	-10.600	0.374 *	-4.395	0.102 *
異質性平均: 過去の購買量	-0.008	0.005 *	-0.006	0.005 *	-0.009	0.005	-0.017	0.006 *	-0.005	0.005
異質性平均: 価格(対数差)	-0.058	0.056	-0.046	0.052	-0.048	0.094	-0.077	0.128	-0.041	0.055
異質性平均: チラシ	0.024	0.076	0.053	0.062	0.087	0.120	0.220	0.166	0.065	0.062
異質性分散: 定数項	0.294	0.126 *	0.579	0.132 *	2.208	0.449 *	3.817	1.178 *	0.773	0.165 *
異質性分散: 過去の購買量	0.005	0.001 *	0.005	0.001 *	0.005	0.001 *	0.006	0.001 *	0.005	0.001 *
異質性分散: 価格(対数差)	0.095	0.029 *	0.141	0.040 *	0.544	0.141 *	0.848	0.238 *	0.227	0.050 *
異質性分散: チラシ	0.134	0.046 *	0.257	0.082 *	0.770	0.162 *	1.723	0.573 *	0.277	0.067 *

※(3)階層ベイズモデルの異質性の分散については、対角成分のみを示す

属性やマーケティング変数の投入方法によって、次の購買時期を予測することができる。例えば、家族人数が3人で、前回の購買量が1000円の世帯が次に購買する期待購買時期などを予測することができる。

次に、(2) Frailty モデルの結果を解釈する。「過去の購買量」は、指数分布モデルと指数べきモデルを除いて、マイナス有意の結果が得られている。これは異質性なしモデルとほぼ同じであるが、指数べきモデルについては符号が反転している。一方で、「価格」および「チラシ」の効果が、ほぼ有意ではなくなっている。また、「家族人数」は、異質性なしモデルと同様にプラス有意な結果が得られているが、係数の大きさが異質性なしモデルと比較して大きくなっている。さらに、ワイブルモデル、対数ロジスティックモデル、指数べきモデルでは、形

状パラメータ  $\alpha$  の値が異質性なしモデルと比較して大きくなっており、代わりに定数項が小さくなっていることがわかる。異質性なしモデルでは一見有意に得られている結果でも、消費者(世帯)別の異質性を考慮すると、必ずしもマーケティング変数の効果は有意ではないことがわかる。

最後に、(3) 階層ベイズモデルの結果を解釈する。階層ベイズモデルでは、異質性の分散  $V_{\beta}$  は、項目間の相関(共分散)も推定しているが、表には対角成分の結果のみを記載している。「家族人数」の係数は、いずれもプラス有意の値となっており、家族人数が増えるほど、購買間隔が短くなるという結果が得られている。一方で、係数の絶対値は異質性なしモデルと近い値をとっており、Frailty モデルよりも小さな値となっている。また、異質性平均の

表2 モデル比較 (異質性なしモデルの結果)

	指数	Erlang-2	ワイブル	対数 ロジスティック	指数べき
対数周辺尤度	-10229	-10278	-10203	<b>-10096</b>	-10415
DIC	20457	20556	20405	<b>20192</b>	20830
BIC	20485	20585	20439	<b>20226</b>	20864

定数項も、異質性なしモデルと近い値になっているが、ワイブル分布モデルと対数ロジスティックモデルは、異質性なしモデルと Frailty モデルよりも小さい値となっていることがわかる。代わりに、ワイブルモデルと対数ロジスティックモデルでは、形状パラメータ  $\alpha$  の値が異質性なしモデルや Frailty モデルと比較して大きくなっている。また、異質性平均の「過去の購買量」については、おおむねマイナス有意の値をとっているが、「価格」は有意な結果が得られなかった。一方で、「チラシ」の異質性平均は、有意な結果は得られていないが、いずれもプラスの値となっており、異質性なしモデルの結果とは符合が逆転している。これらの結果からもわかるように、消費者異質性を考慮する場合としない場合では、係数の符号や有意かどうかの結果が変わることがわかる。

次に、どの確率分布を利用した購買間隔モデルが適しているのかを検証する。(1) 異質性を仮定しないモデルについて、対数周辺尤度、偏差情報量規準 (Deviance Information Criterion; DIC; Spiegelhalter *et al.*, 2002)、ベイズ情報量規準 (Bayesian Information Criterion; BIC; Schwarz, 1978) を用いて検証を行う<sup>12)</sup>。対数周辺尤度の計算には、Newton and Raftery (1994) および Gelfand and Dey (1994) による調和平均による計算方法を用いた。表2に、モデル比較結果を示す。結果から、3つの指標全てにおいて、対数ロジスティック分布を仮定したモデルが最も適しているという結果となり、次にワイブル分布を仮定したモデルが適しているという結果となった。一方で Erlang 分布や指数べき分布を仮定したモデルは、指数分布を仮定したモデルよりも悪い結果となり、Seetharaman and Chintagunta (2003) とは異

なった結果が得られた。分析者は、購買間隔モデルを利用する際には、利用する商品特性などを考慮して、注意深くモデルを選ぶ必要があると言える。

## 5 まとめと今後の課題

本研究では、マーケティング・サイエンスにおける購買間隔モデルに関するレビューを実施した。はじめに、購買間隔データの特性について整理するとともに、購買間隔データを扱う統計モデルについて紹介した。特に、購買間隔データは、時間の長さを従属変数とした解析を行うため、回帰分析を利用することには問題があることを示し、生存時間解析を紹介した。次に、購買間隔モデルの系譜として、確率分布モデルや回帰モデルによる初期のアプローチから、生存時間解析を利用した購買間隔モデルの一般化と、それを拡張したトピックスについて紹介した。購買間隔モデルはマーケティングにおいて重要なテーマの一つであり、クーポンや値引きプロモーションにより他社への離反などを防いだり、自社店舗への来店時期を早めるなどを、データから明らかにできるという利点がある。なお、今回は購買間隔に焦点を当てているため、新製品投入から採用までの時間 (e.g., Sinha and Chandrashekar, 1992) や、耐久財の買い替え時期の予測 (e.g., Grewal *et al.* 2004)、美容院の来店間隔の分析 (e.g., 小西, 2006) などは含めていないが、これらの課題にも購買間隔モデルと同様のアプローチが利用されており、マーケティングにおいて生存時間解析は有用であることがわかる。

また、Seetharaman and Chintagunta (2003) の比較分析をベイズ統計学を利用した購買間隔

モデルに拡張して、スキャン・パネルデータを用いて実証分析を行った。結果から、「過去の購買量」が多くなると購買時期が遅くなり、価格が高くなると購買時期が遅くなるという結果が得られた。しかし価格については、消費者異質性を考慮したモデルでは、有意な結果は得られなかった。一方で、「チラシ」については、係数がプラスのものもマイナスのものもあり、一貫した結果が得られなかった。また、「家族人数」は人数が増えると購買間隔が短くなるという結果が一貫して得られた。今回の結果では、採用する確率分布や、消費者異質性を考慮するか否かによって、マーケティング変数の効果は変わるという結果が得られた。異質性を仮定しないモデルについて、情報量規準を利用してモデル選択を行った結果では、対数ロジスティック分布を仮定したモデルが最も適しており、次いでワイブル分布が適しているという結果となり、指数べきモデルが最もよい結果になった Seetharaman and Chintagunta (2003)とは異なる結果が得られた。しかし、本研究の実証分析では、1つのデータのみしか分析していないため、結果を一般化することは難しい。今後、さまざまなタイプのデータに応用することが期待される。

今回議論していない点として、購買間隔モデルにおける商品特性による違いがあげられる。特に日用品では、購買は繰り返し発生するが、その間隔は商品特性によって異なるはずである。例えば、コーヒーのような商品と、シャンプーなどの商品では、平均的な購買間隔は異なるはずであり、商品のタイプによって異なる確率分布が適していると考えられる。これらの検証を行うには、各商品カテゴリについて同じ条件でさまざまな確率分布を利用した分析を行い、知見を蓄積する必要がある。また、蓄積された結果についてメタ分析などを行うことで、結果の違いがどのような要因（商品属性等）によって生じているかを把握することができる。メタ分析を行うことで、今までにない新しいカテゴリの商品を発売した際に、どのような結果が得られるのか、などを予測することも可能になる。これは、今後の研究課題の一つであると

考えられる。

また、生活環境や、利用しているインターネット環境による違いについての検証も必要である。マーケティング・サイエンスはアメリカを中心に発達しており、多くの研究はアメリカにおける購買履歴データを利用している。しかし、アメリカと日本では生活環境が異なるため、購買傾向も異なることが予想され、結果の国際比較なども興味深いテーマである。また、同じ国内でも、都心部に生活する消費者と、地方（特に車社会）で生活する消費者は、購買傾向が異なることが想定される。さらに、近年ではインターネット通販が発達しているため、多くの消費者は店頭とインターネット通販を併用している。近年では、店頭購買とインターネット購買の双方をシングルソースで取得したデータも整備されつつあることから、これらのデータを利用して生活環境と利用チャネルにおける比較検証も重要な研究課題であると考えられる。

## 謝辞

本研究は、JSPS 科研費（19K20890）の助成を受けて実施したものです。

## 注

- 1) 記憶性の欠如（Lack of Memory）と呼ばれる。
- 2) 指数分布に従う確率変数の和が従う分布は Erlang 分布となる。
- 3) 離散比例ハザードモデルとは、連続時間を週などの離散時間に区切るものである。詳しくは、(4) 時間によって変動するマーケティング変数の扱いのパートで紹介する。
- 4) 複数のイベント（例えばブランド A,B の購買）が存在する場合、ある一つのイベント（例えばブランド A の購買）が観測されると、その他のイベント（例えばブランド B の購買）は観測されない。これを競合リスクと呼び、このような状況

で単独のイベントについて分析を行うと、推定値にはバイアスがかかることが知られており、競合リスクモデルが利用される。

- 5) 比例ハザードモデルでは、ハザード関数はベースライン関数と共変量等の効果の積で表現されるが、加法リスクモデルでは、ハザード関数はベースライン関数と共変量等の効果の和で表現される。
- 6) マルチステートモデルは、複数の種類のイベント（ここではブランド）間の遷移を組み込んだ生存時間解析であり、例えばブランド A からブランド B にスイッチする確率の時間変化を把握することができる。
- 7) バリエーションキミングとは、特定のブランドだけではなく、さまざまなブランドを購入することをいう。一方で、同じブランドを買い続けることを inertia（慣性）などと呼ぶ。
- 8) マルチステートモデルと競合リスクモデルは厳密には異なるが、区別しないこともある。
- 9) 統計学では、どのデータに欠測がおきているかを示す変数は欠測インディケータと呼ばれる。
- 10)  $\bar{b}$  の事後分布は正規分布に、 $V_b$  の事後分布は逆ガンマ分布に、 $\bar{\beta}$  の事後分布は多変量正規分布に、 $V_\beta$  の事後分布は逆ウィシャート分布になる。
- 11) マーケティング変数として前回購買時点の価格と今回の購買時点の価格の差を利用するため、各消費者の初回購買は解析から除外する。
- 12) ベイズ統計学においてモデルの当てはまりを評価する指標として一般的に利用されている。対数周辺尤度は大きいほどモデルの当てはまりがよく、DIC および BIC は小さいほど当てはまりがよい。

## 参考文献

阿部誠, 近藤文代. (2005). マーケティングの科学—POS データの解析—. 朝倉書店.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**(422), 669–679.

Allenby, G. M., Leone, R. P. and Jen, L. (1999). A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association*, **94**(446), 365–374.

Bijwaard, G. E., Franses, P. H. and Paap, R. (2006). Modeling purchases as repeated events. *Journal of Business & Economic Statistics*, **24**(4), 487–502.

Chatfield, C. and Goodhardt, G. J. (1973). A consumer purchasing model with Erlang inter-purchase times. *Journal of the American Statistical Association*, **68**(344), 828–835.

Chen, Y. and Steckel, J. H. (2012). Modeling credit card share of wallet: Solving the incomplete information problem. *Journal of Marketing Research*, **49**(5), 655–669.

Chintagunta, P. K. (1998). Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science*, **17**(3), 253–270.

Chintagunta, P.K. and Dong, X. (2006). Hazard/Survival models in marketing. in Grover, R. and Vriens, M. *The handbook of marketing research: Uses, misuses, and future advances*. Sage.

Chintagunta, P. K. and Haldrar, S. (1998). Investigating purchase timing behavior in two related product categories. *Journal of Marketing Research*, **35**(1), 43–53.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society: Series B*, **34**, 187–220.

Dobson, A. J. 著 田中富, 森川敏彦, 山中竹春, 富田誠 訳. (2008). 一般化線形モデル入門. 共立出版.

Ehrenberg, A. S. (1959). The pattern of consumer purchases. *Applied Statistics*, **8**(1), 26–41.

Fok, D. and Paap, R. (2009). Modeling category - level purchase timing with brand - level marketing variables. *Journal of Applied Econometrics*, **24**(3), 469–489.

Fok, D., Paap, R. and Franses, P.H. (2012). Modeling dynamic effects of promotion on interpurchase times. *Computational Statistics & Data Analysis*, **56**, 3055–3069.

Franses, P. H. and Paap, R. (2001). *Quantitative models in marketing research*. Cambridge University Press.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B*, **56**(3), 510–

514.

- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian data analysis 3rd*. Chapman & Hall/CRC.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. in *Bayesian statistics 4*, Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. eds., Oxford University Press.
- Gönül, F. and Srinivasan, K. (1993). Consumer purchase behavior in a frequently bought product category: Estimation issues and managerial insights from a hazard function model with heterogeneity. *Journal of the American Statistical Association*, **88**(424), 1219-1227.
- Grewal, R., Mehta, R. and Kardes, F. R. (2004). The timing of repeat purchases of consumer durable goods: The role of functional bases of consumer attitudes. *Journal of Marketing Research*, **41**(1), 101-115.
- Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research*, **25**(4), 342-355.
- Gupta, S. (1991). Stochastic models of interpurchase time with time-dependent covariates. *Journal of Marketing Research*, **28**(1), 1-15.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, **47**(1), 153-161.
- Helsen, K. and Schmittlein, D.C. (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Science*, **11**(4), 395-414.
- Herniter, J. (1971). A probabilistic market model of purchase timing and brand selection. *Management Science*, **18**(4), 102-113.
- 星野崇宏. (2013) 継続時間と離散選択の同時分析のための変量効果モデルとその選択バイアス補正: Webログデータからの潜在顧客への広告販促戦略立案. 日本統計学会誌, **43**(1), 41-58.
- Hosmer, D. W., Lemeshow, S. and May, S. (2011). *Applied survival analysis*. Wiley Blackwell.
- Ibrahim, J. G., Chen, M. H. and Sinha, D. (2005). *Bayesian survival analysis*. John Wiley & Sons, Ltd.
- 猪狩良介, 星野崇宏. (2016). Online-Offline チャネルにおける消費者の購買間隔と購買金額の同時モデリング. オペレーションズ・リサーチ, **61**(9), 589-599.
- Igari, R. and Hoshino, T. (2017). A survival analysis incorporating auxiliary information by a Bayesian generalized method of moments: Application to purchase duration modeling. *Journal of the Japanese Society of Computational Statistics*, **30**(1), 27-44.
- Igari, R. and Hoshino, T. (2018). A Bayesian data combination approach for repeated durations under unobserved missing indicators: Application to interpurchase-timing in marketing. *Computational Statistics & Data Analysis*, **126**, 150-166.
- Jain, D. C. and Vilcassim, N. J. (1991). Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Science*, **10**(1), 1-23.
- Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data 2nd*. Springer Science & Business Media.
- 小西葉子. (2006). 存続時間分析による美容院顧客の来店確率予測. 統計数理, **54**(2), 445-459.
- Leszczyc, P. T. P., Sinha, A. and Timmermans, H. J. (2000). Consumer store choice dynamics: An analysis of the competitive market structure for grocery stores. *Journal of Retailing*, **76**(3), 323-345.
- Manchanda, P., Dube, J. P., Goh, K. Y. and Chintagunta, P. K. (2006). The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, **43**(1), 98-108.
- Moe, W. W. and Fader, P. E. (2004). Capturing evolving visit behavior in clickstream data. *Journal of Interactive Marketing*, **18**(1), 5-19.
- Neslin, S. A., Henderson, C. and Quelch, J. (1985). Consumer promotions and the acceleration of product purchases. *Marketing Science*, **4**(2), 147-165.
- Newton, M.A. and Raftery, A. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B*, **56**(1), 3-48.
- 新美潤一郎, 星野崇宏. (2015). ユーザ別アクセス・

- パターン情報の多様性を用いた顧客行動の予測とモデリング. *応用統計学*, **44**(3), 121-143.
- Park, Y. H. and Fader, P. S. (2004). Modeling browsing behavior at multiple websites. *Marketing Science*, **23**(3), 280-303.
- Park, C. H., Park, Y. H. and Schweidel, D. A. (2014). A multi-category customer base analysis. *International Journal of Research in Marketing*, **31**(3), 266-279.
- Rossi, P. E., McCulloch, R. E. and Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science*, **15**(4), 321-340.
- 佐藤忠彦. (2015). マーケティングの統計モデル. 朝倉書店.
- 里村卓也. (2015). マーケティング・モデル (第2版). 共立出版.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**(2), 461-464.
- Seetharaman, P. B. (2004). The additive risk model for purchase timing. *Marketing Science*, **23**(2), 234-242.
- Seetharaman, P. B., Chib, S., Ainslie, A., Boatwright, P., Chan, T., Gupta, S. and Strijnev, A. (2005). Models of multi-category choice behavior. *Marketing Letters*, **16**(3-4), 239-254.
- Seetharaman, P. B. and Chintagunta, P. K. (2003). The proportional hazard model for purchase timing: A comparison of alternative specifications. *Journal of Business & Economic Statistics*, **21**(3), 368-382.
- Sinha, R. K. and Chandrashekar, M. (1992). A split hazard model for analyzing the diffusion of innovations. *Journal of Marketing Research*, **29**(1), 116-127.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, **64**(4), 583-639.
- Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**(398), 528-540.
- Telang, R., Boatwright, P. and Mukhopadhyay, T. (2004). A mixture model for Internet search-engine visits. *Journal of Marketing Research*, **41**(2), 206-214.
- 照井伸彦. (2008). ベイズモデリングによるマーケティング分析. 東京電機大学出版局.
- Vakratsas, D. and Bass, F. M. (2002). A segment - level hazard approach to studying household purchase timing decisions. *Journal of Applied Econometrics*, **17**(1), 49-59.
- Vilcassim, N. J. and Jain, D. C. (1991). Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity. *Journal of Marketing Research*, **28**(1), 29-41.
- 山口景子. (2014). 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**(1), 13-29.
- Wedel, M. Kamakura, W. A. DeSarbo, W. S. and Ter Hofstede, F. (1995). Implications for asymmetry, nonproportionality, and heterogeneity in brand switching from piece-wise exponential mixture hazard models. *Journal of Marketing Research*, **32**(4), 457-462.

