

CIRJE-J-298

周期性のある購買行動に適応する顧客生涯価値モデル：  
購買履歴データを使ってPareto/NBDモデルの限界を克服する

東京大学大学院経済学研究科  
阿部 誠

2020年4月

CIRJE ディスカッションペーパーの多くは  
以下のサイトから無料で入手可能です。  
[http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp\\_j.html](http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp_j.html)

このディスカッション・ペーパーは、内部での討論に資するための未定稿の段階にある論文草稿である。著者の承諾なしに引用・複写することは差し控えられる。

周期性のある購買行動に適応する顧客生涯価値モデル：  
購買履歴データを使って Pareto/NBD モデルの限界を克服する

阿部 誠

東京大学大学院経済学研究科  
東京都文京区本郷7-3-1  
Email: [abe@e.u-tokyo.ac.jp](mailto:abe@e.u-tokyo.ac.jp)

April 15, 2020

## 概要

顧客の離脱を観測できない非契約型 CRM では、BTYD (Buy Till You Die) モデルを使うことによって離脱を推測できる。その代表的なモデルである Pareto/NBD では、リセーとフリクエンシーという最小限の購買データから推測を可能にするために、購買行動に強い仮定を置いている。しかし、購買がランダムに発生するという仮定は、周期性のある店舗、カテゴリー、商品の購買には相応しくない。また、購買率と離脱率に関する2つのガンマ分布が独立という仮定では、購買行動と離脱行動の関係が無視されている。

本研究では、個々の購買の時点と金額を含んだ顧客購買履歴データを用いて、周期性の有り無しいずれの購買行動にも対応する柔軟な顧客生涯価値モデルを構築する。先行研究が豊富な Pareto/NBD 系モデル同様、離脱はランダム、1回当たりの購買金額は顧客内で対数正規分布にしたがうと仮定する一方、購買には周期性に対応するロジスティック閾値モデルを採用する。提案モデルからは、生涯価値や購買の規則性など、CRM に有用な様々な顧客別のマーケティング指標が得られる。

実証分析では、ヘアサロンの購買（訪問）履歴データを用いて、ランダム購買を仮定した個人モデル(PE)と混合分布による集計モデル (Pareto/NBD) とで比較した。その結果、データへのフィットと予測精度に関しても、パラメータの推定に関しても、提案モデルが Pareto/NBD や PE より優れていた。さらに顧客維持戦略への応用として、生涯価値の増加を最大にする最適介入レベルを顧客別に算出した。

**Keywords: customer lifetime value (CLV), churn, memoryless, cyclic, Pareto/NBD, Poisson**

## **Customer Lifetime Value Model that Accommodates Cyclic Purchase Behavior: Overcoming the limitation of Pareto/NBD models using Complete Purchase History**

In non-contractual CRM whereby customer churn cannot be observed, BTYD (Buy Till You Die) models permit one to infer the churn. Amongst these models, the most popular one, Pareto/NBD posits strong assumptions on purchase behavior in order to allow its model estimation from a minimal amount of information, namely, customers' recency and frequency (RF) data. The assumptions, however, place two serious restrictions. First, the Poisson purchase is not appropriate for stores, categories, and products that exhibit cyclic transaction behavior. Second, the independence of two gamma mixture distributions for the purchase and churn rates ignores the association between purchase and churn behaviors.

This research proposes a customer lifetime (CLV) model that can accommodate both memoryless and cyclic purchase behaviors from customers' complete purchase history. Identical to extensively studied Pareto/NBD and its variant models, the proposed model assumes random churn and stochastic spending per transaction following a lognormal distribution. In contrast, purchase behavior, to address its cyclicality, is captured by a logistic threshold model. This CY (cyclic) model provides customer-specific marketing metrics that are useful for one-to-one marketing, including CLV and purchase cyclicality.

Using purchase history from hair salon customers, the proposed model (CY) is compared against two models, both of which assume memoryless purchase: one is an individual-level Poisson purchase / Exponential churn (PE) model, and the other is a Pareto/NBD that captures customer heterogeneity through independent gamma mixture distributions. CY model resulted in superior performance over PE and Pareto/NBD in terms of both fit/prediction and parameter estimate. In order to demonstrate application to a retention tactic, the model derived customer-specific optimal level of interception that maximizes the return on CLV.

## 1. はじめに

顧客関係管理(CRM)で重要な概念である顧客生涯価値(CLV)を算出するには、顧客の離脱率を把握する必要がある (Blattberg & Deighton 1996)。しかしながら、年会費などの支払い義務のない非契約型ビジネスの場合、顧客は単に購買を止めるだけで、離脱は直接、観測できない。購買がしばらく無くても、それは次回までの購買サイクルがたまたま長いだけで、顧客は離脱していないのかも知れない。

非契約型 CRM において離脱を推測する代表的な消費者行動モデルは (Buy Till You Die: BTYD) と呼ばれ、その名称どおり、「顧客は離脱するまで購買を続け離脱後は顧客としてカムバックしない」仮定をおいている。BTYD の中でも、購買と離脱はランダムに発生するという単純な仮定に基づいた Pareto/NBD モデル (Schmittlein, Morrison & Colombo 1987) は、そのシンプルさから、多数の拡張モデルが提案、研究されてきた。たとえば、R のパッケージ “BTYD” や “BTYDplus” には、さまざまなモデルが組み込まれている (表 1)。

表 1: R のパッケージに含まれている代表的な BDTY モデル

| モデル                     | 論文   |
|-------------------------|--|
| <b>MBG/NBD</b>          | Batislam, Denizel & Filiztekin (2007), Hoppe and Wagner (2007) |
| <b>BG/NBD</b>           | Fader, Hardie & Lee (2005b)                                    |
| <b>BG/BB</b>            | Fader, Hardie & Shang (2010)                                   |
| <b>Pareto/NBD</b>       | Fader, Hardie & Lee (2005a)                                    |
| <b>Pareto/NBD (HB)</b>  | Ma and Liu (2007)  |
| <b>Pareto/NBD (Abe)</b> | Abe (2009)   |
| <b>Pareto/GGG</b>       | Platzer and Reutterer (2016)                                   |

Pareto/NBD とその類型モデルの特徴は、消費者行動の購買と離脱がランダムに発生することである。これは、購買回数がポアソン分布、生存時間が指数分布にしたがうことを意味する。さらに、購買率を表すポアソン分布のパラメータと離脱率を表す指数分布のパラメータは、消費者によって異なる値であることを考慮して、それらは独立なガンマ分布 (混合分布) にしたがう。

Pareto/NBD 類型の利点としては、強い仮定のためモデルがシンプルで、顧客のリセナシーとフリクエンシー (RF データ)のみから推定可能なことだ。反面、この利点は表裏一体で、以下の3つの弱点につながる。

(1) 前回購買からの期間に関係なく購買がランダムに発生する仮定は、周期性のある店舗、カテゴリー、商品の購買には相応しくない。しかし、購買周期性を把握するには、その顧客の購買間隔の情報が必要だが、それは RF データにはふくまれていない。

(2) 購買率と離脱率に関する2つの混合分布が独立という仮定は、購買行動と離脱行動に関係が無いことを意味する。しかし、カテゴリーによっては、購買頻度の高い顧客の方が離脱率が低いことは十分に想定される。

(3) 個別対応の CRM では、パラメータの値が顧客別に得られると特に有用である。しかし、ポアソン分布と指数分布の2つのパラメータを推定するためには、RF データに含まれる3つのデータポイント（観測期間、リセンサー、購買回数）では情報が少ない。これを克服するためにベイズ推定や階層ベイズモデルが用いられるが、推定が不安定な場合がある。

これらの弱点を踏まえ、本研究の目的は、一回一回の購買の時点と金額を含んだ顧客の購買履歴データを用いて、周期性のある購買行動にも適応した柔軟な顧客生涯価値モデルを構築することにある。

論文は以下のように構成されている。まずは、先行研究、特に Pareto/NBD と対比しながら、本研究のアプローチを提示する。次にモデルの詳細と、そこから導出される生涯価値をも含めた顧客別マーケティング指標を説明する。その後、CRM の顧客維持におけるモデルの活用例として、最適な介入戦術を紹介する。そして、実証研究として、ヘアサロンの会員カードから収集された顧客購買履歴データにモデルを適用して経営上の示唆を得る。最後に研究の結論とモデルの限界を述べる。

## 2. フレームワークと既存研究の中での位置付け

### 2. 1. 購買モデル

顧客によって、規則的なサイクル（たとえば毎週末）で購買する人がいる一方、ほぼランダムなタイミングで購買する人もいるだろう。一人の顧客の購買記録を一定の期間内で集計した RF データと違って、購買履歴データからは、その顧客の全ての購買間隔を知ることが出来る。したがって、その顧客がどの程度、規則的な周期で購買しているのかを把握することができる。

マーケティング・サイエンスでは、購買間隔モデルに生存分析を用いることが多く、代表的な購買間隔の密度関数として、ワイブル分布、ガンマ分布、一般化ガンマ分布などが使われる(Allenby, Leone & Jen 1999, Gupta 1991, Helsen & Schmittlein 1993, Jain & Vilcassim 1991)。これらの密度関数はシェイプ・パラメータの値によっては形状が非

連続的に変化するため、パラメータの推定が不安定になる傾向が見られる。Pareto/NBD モデルでは、購買間隔はワイブル分布の特殊なケース（シェイプ・パラメータが1）である指数分布となっており、これはランダム購買を仮定している。

我々の目的はモデルの実装化なので、個人レベルでの推定がロバストで安定していることは重要である。そのため本研究では、以下のアプローチをとる。

（1）購買間隔の情報が個人レベルで十分に得られるヘビーユーザーに焦点をあててモデルを構築する。購買間隔のデータが少ないライトユーザーへの拡張は、事前情報を組み込んだベイズ推定によって対応可能である。

（2）購買間隔の累積密度関数は、シンプルでパラメータ数が必要最小限のロジスティック分布（この先、ロジスティック閾値モデルと呼ぶ）を用いる。推定されたパラメータからは、顧客ごとに購買サイクルの規則性を指標化することができる。

## 2. 2. 離脱モデル

顧客は、対象（店舗、カテゴリー、商品、etc.）に対する関心や嗜好を失ったり、競合へのスイッチ、転居、死去など、様々な理由から離脱する。したがって、Pareto/NBD モデルと同様、顧客の離脱はランダムに発生すると仮定する。これは生存分析の枠組みでは、ハザード関数が定数であることを意味し、生存時間は指数分布に従う。

## 2. 3. 金額モデル

Abe (2016)に従って、顧客内の1回当たりの購買金額は、対数正規分布にしたがうと仮定する。その理由は、購買金額が正の値をとり、パラメータ推定がガンマ分布より安定しており、階層モデルへの拡張が容易なためである。

## 2. 4. 顧客の異質性

表2は、Pareto/NBDに基づいた代表的な顧客生涯価値モデル、Schmittlein and Peterson (1994) (以後SPと呼ぶ)、Fader, Hardie and Lee (2005a) (以後FHLと呼ぶ)、Abe (2016)を、本研究のモデルと対比させたものである。既存の3モデルはすべてRFMデータを使い、購買行動はランダムであるポアソン・プロセスを、離脱行動もランダム性を仮定している。顧客の異質性に関しては、SPとFHLでは、それぞれのパラメータが独立なガンマ混合分布にしたがうことで対応しており、推定には最尤法が使われている。一方、Abe(2016)は一人ひとりのパラメータをベイズ法で推定することで異質性に対処している。本研究では、個々の購買データを使い、購買発生は周期性を考慮したロジスティック閾値モデル、離脱はランダム性を仮定し、一人ひとりのパラメータを最尤法で推定する。

表 2： Pareto/NBD に基づいた代表的な顧客生涯価値モデルの比較

| モデル | 特徴  | 混合分布モデル<br>Pareto/NBD<br>SP (1994) | 混合分布モデル<br>Pareto/NBD<br>FHL (2005) | 個人レベル<br>Abe (2016) | 個人レベル<br>本研究 |
|-----|-----|------------------------------------|-------------------------------------|---------------------|--------------|
| 購買  | 行動  | ポアソン                               | ポアソン                                | ポアソン                | 閾値モデル        |
|     | 異質性 | ガンマ分布                              | ガンマ分布                               | 個人レベル               | 個人レベル        |
|     | 周期性 | ×                                  | ×                                   | ×                   | ○            |
| 離脱  | 行動  | ランダム                               | ランダム                                | ランダム                | ランダム         |
|     | 異質性 | ガンマ分布                              | ガンマ分布                               | 個人レベル               | 個人レベル        |
| 金額  | 行動  | 正規分布                               | ガンマ分布                               | 対数正規分布              | 対数正規分布       |
|     | 異質性 | 正規分布                               | ガンマ分布                               | 個人レベル               | 個人レベル        |
| データ |     | RFM                                | RFM                                 | RFM                 | 全購買          |
| 推定法 |     | 最尤法                                | 最尤法                                 | 階層ベイズ               | 最尤法          |

金額 (M データ) のモデル化では、SP は normal-normal モデル、つまり一人の顧客内の購買金額は正規分布、顧客間の異質性はその平均パラメータが正規分布にしたがうと仮定している。FHL は gamma-gamma モデル、つまり顧客内の購買金額はガンマ分布、顧客間の異質性はその平均パラメータがガンマ分布にしたがうと仮定している。Abe(2016)は、顧客内の購買金額は対数正規分布を仮定し、一人ひとりのロケーション・パラメータを階層ベイズ法で推定することで異質性に対応している。本研究では、顧客内の購買金額は対数正規分布を仮定し、一人ひとりのロケーション・パラメータは最尤法で推定する。

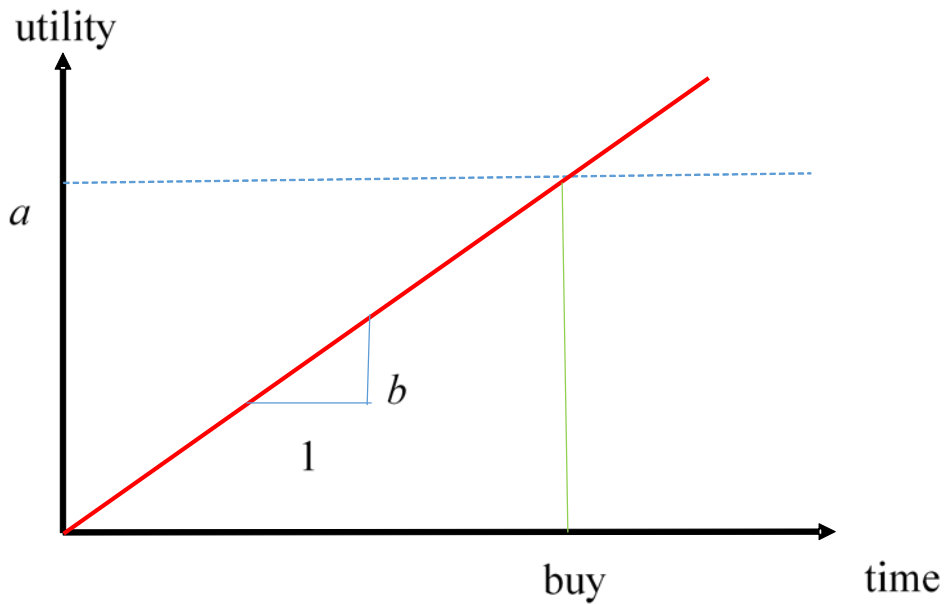
### 3. モデル

#### 3. 1. 購買モデル — ロジスティック閾値モデル

時間とともに線形に増加する効用 (*utility*) が、閾値  $a$  を超えると購買が発生するモデルを考える (図 1)。観測されない要因を考慮して、効用と閾値にはノイズ ( $\varepsilon, \eta$ ) を加える。この 2 つのノイズは独立同一の極値分布に従うと仮定すると、前回の購買から  $t$  時点までに購買が発生する確率  $p(t)$  は式 1 のロジスティック分布で表される。



図 1： 購買モデル - ロジスティック閾値モデル



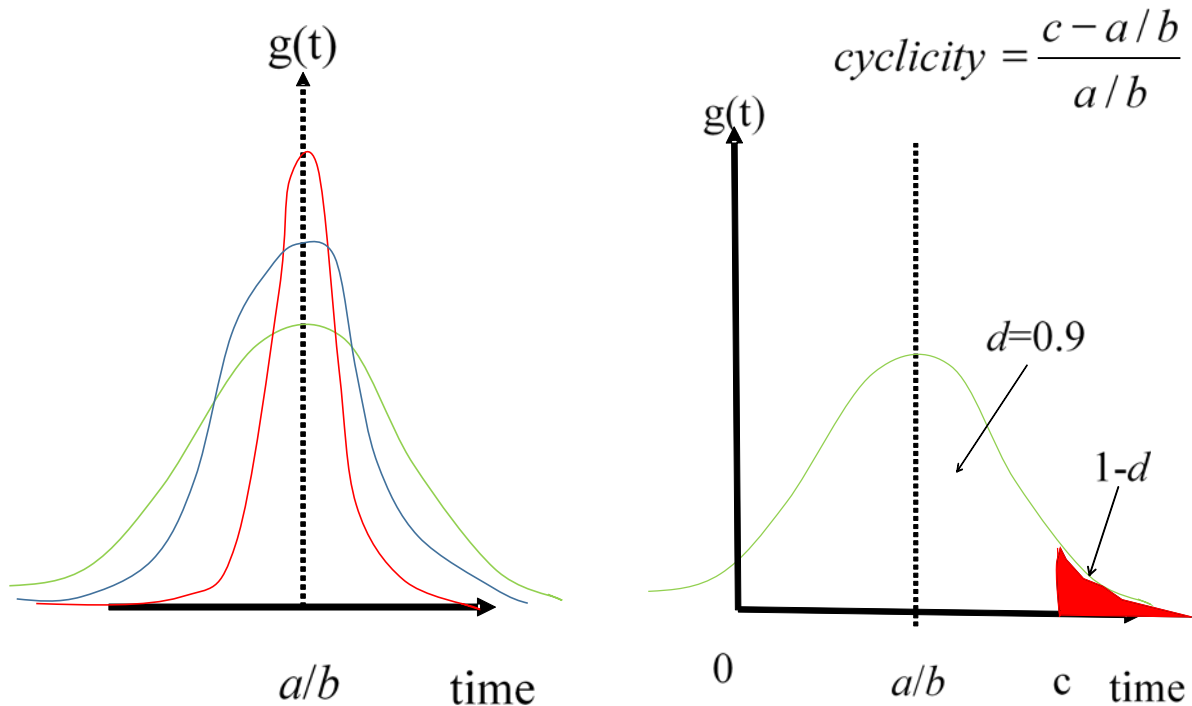
$$(1) \quad p(t) = \text{prob}(bt + \varepsilon > a + \eta) = \text{prob}(\eta - \varepsilon < -a + bt) = \frac{e^{-a+bt}}{1 + e^{-a+bt}}$$

たとえばヘアサロンへの訪問（購買）の例であれば、 $b$  は単位時間当たりの髪の毛の伸び率と解釈できる。このシンプルなモデルは、いくつかの拡張が可能である。まず、外的な説明変数  $x$ （週末、天気、イベントなど）を、 $bt+cx+\varepsilon$  の形で効用に組み込むことができる。さらに、顧客による毛の伸び率の違いを考慮して、 $b$  を個人属性  $d$  の関数  $b(d)$  とした階層モデルに発展させることができる。

購買間隔の密度関数は、式 1 を  $t$  で微分したもので、図 2 になる。モードは  $a/b$  で、 $a$  と  $b$  のスケールが大きくなるとモードの周辺でより尖った形状になる。これは、購買サイクルがより規則的になることを意味しているため、この規則性を解釈しやすいマーケティング指標で表現したものが、図 3 の *cyclicity* (周期性) である。

図 2： 購買間隔の密度関数

図 3： *cyclicity* (周期性)



これは、次回の購買が  $(100 \times d)\%$  の確率で発生する時間  $c$  を、平均購買間隔  $a/b$  で標準化したものである。たとえば  $d=0.9$ 、 $c=k \times a/b$  であれば、 $cyclicity=k-1$  となり、次回の購買は平均購買間隔の  $(cyclicity+1)$  倍以内に  $90\%$  の確率で発生すると解釈できる。 $cyclicity$  は正の値をとり、小さな値ほど、より規則的な購買サイクルを表わす。

### 3. 2. 離脱モデルと金額モデル

ランダム離脱の仮定から、生存時間  $\tau$  の密度関数  $f(\tau)$  は、離脱率  $\mu$  の指数分布にしたがう (式 2)。

$$(2) \quad f(\tau | \mu) = \mu e^{-\mu\tau} \quad \tau \geq 0$$

顧客の 1 回当たりの購買金額  $s$  は、対数正規分布にしたがう。

$$(3) \quad \log(s) \sim N(\log(m), \omega^2) \quad s > 0$$

式 3 では、 $N$  は正規分布を表し、 $\omega^2$  は 顧客内 の対数購買金額の分散である。ある顧客の 1 回当たりの購買金額の期待値は、 $E[s] = \exp(\log(m) + \omega^2/2)$  になる。

購買、離脱、金額の3モデルのパラメータ  $\{a, b, \mu, m, \omega^2\}$  は、顧客の異質性から、一人ひとり違う値をとると仮定し、購買履歴データから最尤法によって推定される。

### 3. 3. 顧客生涯価値 (CLV)とその他のマーケティング指標

観測終了時点での顧客の生存確率は、式4で表される(導出は付録を参照)。

$$(4) \quad P(\text{alive at } T) = \frac{1 - G(r)}{\{1 - G(r)\} + e^{\mu r} \int_0^r \frac{\mu e^{-\mu t}}{e^{-a+bt} + 1} dt}$$

ここでは、初回の購買時点を0とした時の観測終了時点を  $T$ 、リセシーを  $r$ 、 $G(t)$ は購買間隔の累積確率密度(式1のロジスティック閾値モデル  $p(t)$  と同等)を表す。

顧客が一定期間内  $w$  に購買する回数の期待値は、平均間隔  $a/b$  の周期性があるため、その正確な導出は複雑である。式5のように  $E(\psi)$  を計算することで近似できるが(付録を参照)、 $w$  が購買間隔に比べてある程度、長くないと不正確になる。平均購買間隔  $a/b$  までに購買が発生する確率は  $1/2$  のため、最後の項で  $0.5$  が減算されている。

$$(5) \quad E[X(w) | a, b, \mu] \approx \frac{b}{a} E[\psi] = \frac{b}{a\mu} (1 - e^{-\mu w}) - 0.5 \quad \text{ただし、} \psi = \min(\tau, w).$$

一定期間内  $w$  の期待総購買金額は、式5の期待購買回数  $E[X(w) | a, b, \mu]$  と1回当たりの期待購買金額  $E[s | m, \omega]$  の積から式6になる。

$$(6) \quad E[\text{sales}(w) | a, b, \mu, m, \omega] = E[s | m, \omega] E[X(w) | a, b, \mu] = e^{\log(m) + \omega^2 / 2} \left[ \frac{b}{a\mu} (1 - e^{-\mu w}) - 0.5 \right].$$

顧客購買データからはマージンやコストに関する情報は分からないため、本研究では、価値を利益ではなく売上げと定義して、顧客生涯価値(CLV)を導出をする。連続時間に基づいた Pareto/NBD モデルでは、ある顧客の CLV は一般に以下で表される。

$$CLV = \int_0^{\infty} V(t)R(t)D(t)dt$$

ここで、 $V(t)$ は時点  $t$  における顧客価値（期待売上げ）、 $R(t)$ は生存関数（顧客が  $t$  期間以上、生存する確率）、 $D(t)$ は時間に関する割引率である（FHL; Rosset, Neumann, Eick & Vatnik 2003）。提案モデルでは、 $V(t)=b/a E[s]$ （ただし  $s$  は対数正規分布にしたがうため、 $E[s]=\exp[\log(m)+\omega^2/2]$ ）、 $R(t)=\exp(-\mu t)$  になる。割引の年率  $d$  を連続時間で複利計算すると、 $D(t)=\exp(-\delta t)$ 、ただし時間の単位は「年」で、 $\delta=\log(1+d)$  となる。これらを代入すると、CLV は以下の式で表すことができる。

$$(7) \quad CLV = \int_0^{\infty} V(t)R(t)D(t)dt = \int_0^{\infty} \frac{b}{a} e^{\log(m)+\omega^2/2} e^{-\mu t} e^{-\delta t} dt = \frac{\frac{b}{a} e^{\log(m)+\omega^2/2}}{\mu + \delta}$$

したがって、購買履歴データから各顧客の  $\{a, b, \mu, m, \omega^2\}$  が推定できれば、CLV は式 7 を使って計算することができる。

### 3. 4. 顧客維持戦略

顧客維持のための投資額とその維持率との関係を仮定することによって、CLV を最大化する顧客維持投資レベルが算出できる。Blattberg & Deighton (1996) にしたがって、もし顧客が離脱していた場合、介入レベル  $c$  によって、 $r(c)$  の確率で呼び戻せると考えよう。

$$r(c) = R(1 - e^{-kc}) \quad \text{where } c \geq 0$$

関数  $r(c)$  のパラメータ、 $R$  と  $k$  は過去の維持介入結果から決める。データが存在しない場合は、マネジャーに対していくつかの単純な問に回答してもらう決定解析法やデルファイ法から推定できる(Blattberg & Deighton 1996、古川、守口、阿部 2011)。

顧客が離脱しているかは観測されないため、本モデルで算出した顧客の生存確率を  $p$  とすると、介入後の生存確率  $p_a$  は  $p_a = p + (1-p)r(c)$  になる。よって、介入レベル  $c$  による CLV の増収  $\Delta$  は式 8 で表される。

$$(8) \quad \begin{aligned} \Delta &= (p_a V - c) - pV \\ &= (1-p)r(c)V - c \quad \text{ただし } V = \text{顧客生涯価値} \end{aligned}$$

顧客ごとに、CLV の増加が最大になるような介入レベル  $c^*$  は、式 8 の一階条件を解

いた式 9 になる。

$$(9) \quad c^* = \frac{1}{k} \ln[(1-p)kRV]$$

式 9 を解釈すると、CLV が大きい顧客や離脱している確率の高い顧客ほど呼び戻す価値が高いため、最適投資額は大きくなる。逆に CLV が小さい顧客や離脱している可能性が低い顧客では、投資に見合うリターンが得られずに、最適投資レベルが低くゼロになる場合もある。

### 3. 5. 推定

購買間隔データには、観測期間による打ち切りがあるため、購買と離脱プロセスを同時に考慮する必要がある。  $n$  回目の購買間隔を  $t_n$  とすると尤度関数は式 10 で表せる（導出は付録を参照）。

$$(10) \quad L(a, b, \mu | t_n, r, T) \propto \prod_n g(t_n) e^{-\mu T} \left[ \{1 - G(r)\} + e^{\mu r} \int_0^r \frac{\mu e^{-\mu t}}{e^{-a+bt} + 1} dt \right]$$

金額モデル 3 式の  $\{m, \omega^2\}$  は、正規分布のパラメータなので、最尤推定値は式 11 の標本平均と標本分散になる。ここで、 $x$  は購買回数を表す。

$$(11) \quad \begin{aligned} \overline{\log(m)} &= \frac{1}{x} \sum_{n=1}^x \log(s_n) \\ \omega^2 &= \frac{1}{x-1} \sum_{n=1}^x [\log(s_n) - \overline{\log(m)}]^2 \end{aligned}$$

## 4. 実証分析

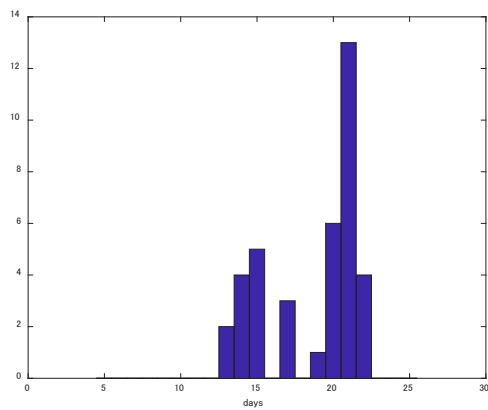
### 4. 1. データ

検証には、あるヘアサロンの FSP メンバーの購買（訪問）履歴データを用いた。その理由は、当該カテゴリーの購買サイクルは規則性が強いため（数週間おき）、提案モデルの利点がランダム購買を仮定した Pareto/NBD モデルの弱点に対して顕著に表れると考えられるからである。

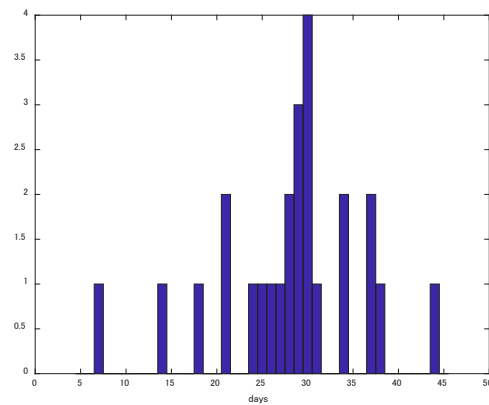
2015年7月1日から9月30日の3ヵ月間にFSPのメンバーになった顧客の中で、推定期間12ヵ月間(2015/7/1~2016/6/30)に5回以上、購買のあったヘビーユーザーを分析した。対象顧客307名の購買間隔の平均は46.9日、標準偏差は24.1日である。その先12ヵ月間(2016/7/1~2017/6/30)の購買履歴データはモデルの検証に用いた。対象顧客の延べ購買回数は、推定期間が2179回、検証期間が1649回である。

図4は、ヘビーユーザートップ3名の購買間隔のヒストグラムを示したものである。Pareto/NBDの仮定するランダム購買(ポアソン・プロセス)であれば、指数分布(モードが0)になるはずだが、購買に規則性があるため、形状はそれとは大きく異なっているのが分かる。

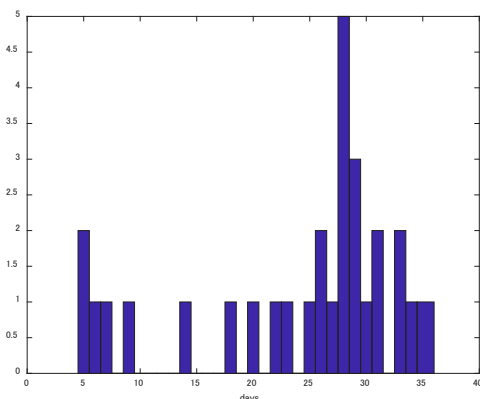
図4： 典型的な顧客の購買間隔



ID=22



ID=95



ID=140

## 4. 2. モデルの検証

提案モデル CY(cyclic) では、購買に周期性を考慮し、離脱はランダム、金額モデルは対数正規分布を仮定して、個人レベルのパラメータを当てはめている。ベンチマーク・モデルには、離脱プロセスと金額プロセスにおける仮定は同じだが購買はランダムという2つのモデルを検討する。一つは、これに個人レベルのパラメータを当てはめたポアソン購買・指数生存モデル PE (Poisson/Exponential)、もう一つは個人の異質性を独立なガンマ混合分布でモデル化した Pareto/NBD である。

顧客の離脱や CLV は観測できないため、正解としての教師信号にはならない。モデルのパフォーマンスを評価するには、検証期間中に観測された購買回数や購買金額をモデルの予測値と比較するしかない。

検証する3つのモデル (CY、PE、Pareto/NBD) はいずれも確率モデルなので、一人ひとりの購買予測精度は顧客によって大きなバラツキがある。このバラツキを平坦化するため、推定期間中の購買回数ごとに顧客をグループ化して、グループごとに平均予測精度を比較する (Fader, Hardie & Lee 2005a)。

最初に、データへのフィットを確認するために、x 軸に推定期間中 (week 1~52) のリピート購買回数 ( $x=4, 5, \dots$ )、y 軸にその顧客の推定期間中に予測される平均購買回数と平均総購買金額をそれぞれプロットしたものが図5と図6である。購買回数も購買金額も、CY と PE は Pareto/NBD より大幅にフィットがよく、CY は PE より少し優れている。

図5： 推定期間中の平均購買回数の比較

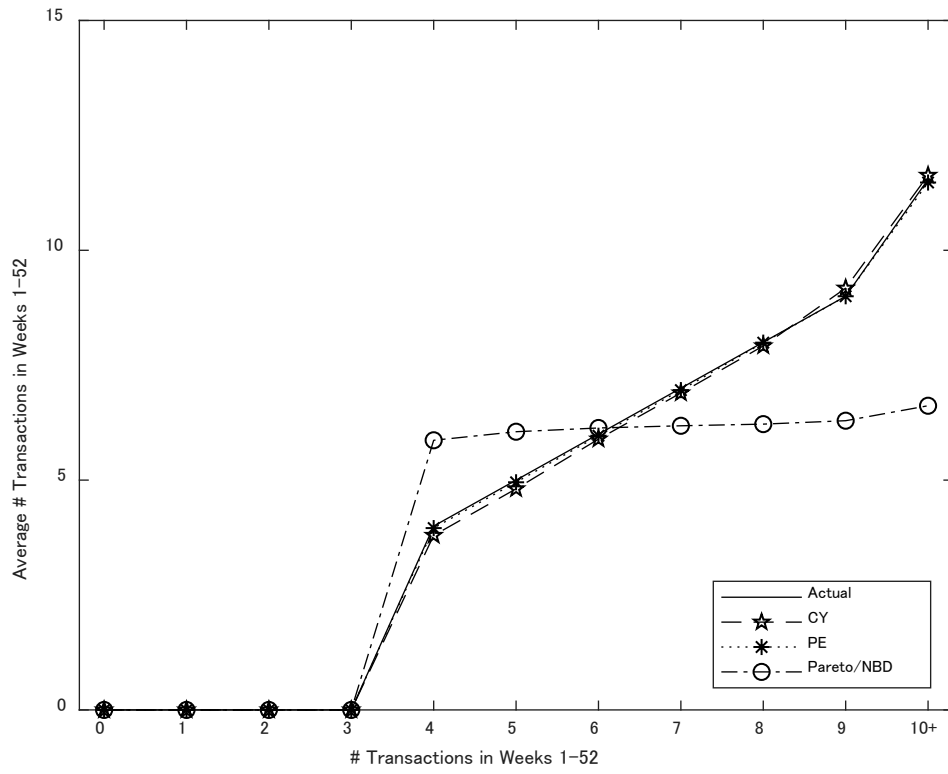
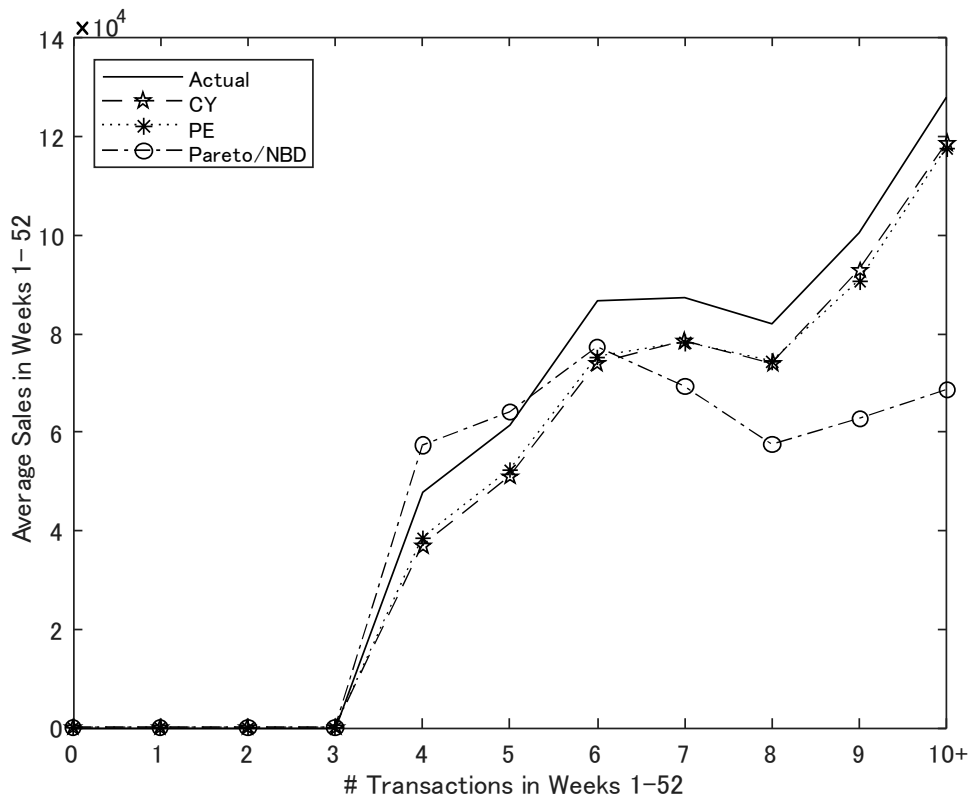


図 6： 推定期間中の平均総購買金額の比較





次に、予測精度を確認するために、x軸に推定期間中のリピート購買回数、y軸にその顧客（x=4, 5, …）の検証期間中（week 53~104）に予測される平均購買回数と平均総購買金額をそれぞれプロットしたものが図7と図8である。購買回数も購買金額も、CYとPEはPareto/NBDより大幅に予測精度がよく、CYはPEより少し優れている。

図7： 検証期間中の平均購買回数の比較

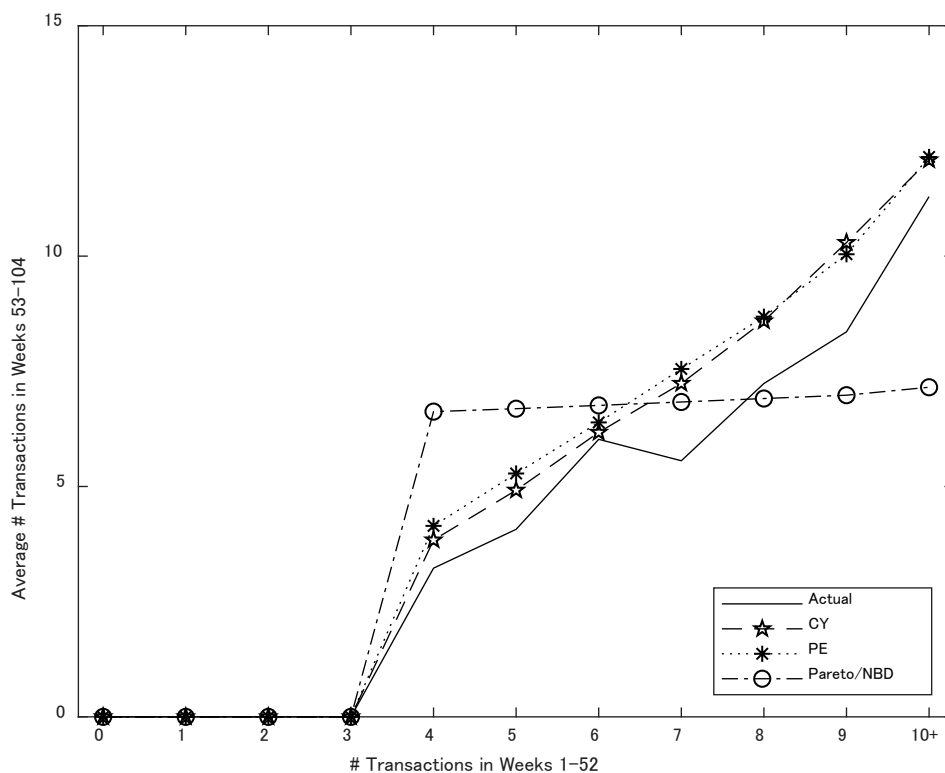


図8： 検証期間中の平均総購買金額の比較

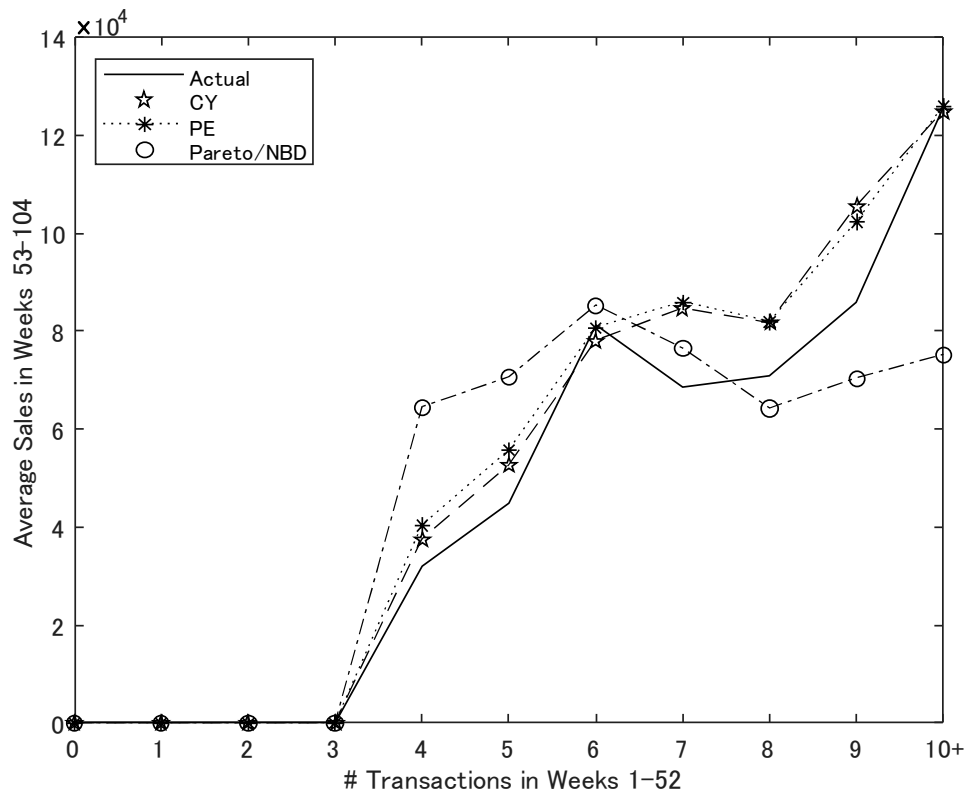


表 3 は、3 つのモデルの非集計レベルにおける精度指標として、顧客別に予測された検証期間中の購買回数(金額)と実際に観測された購買回数(金額)を相関係数と平均二乗誤差(MSE)で評価したものである。購買回数と購買金額の両方で、相関係数、MSE いずれも、CY、PE、Pareto/NBD の順に予測精度が高かった。

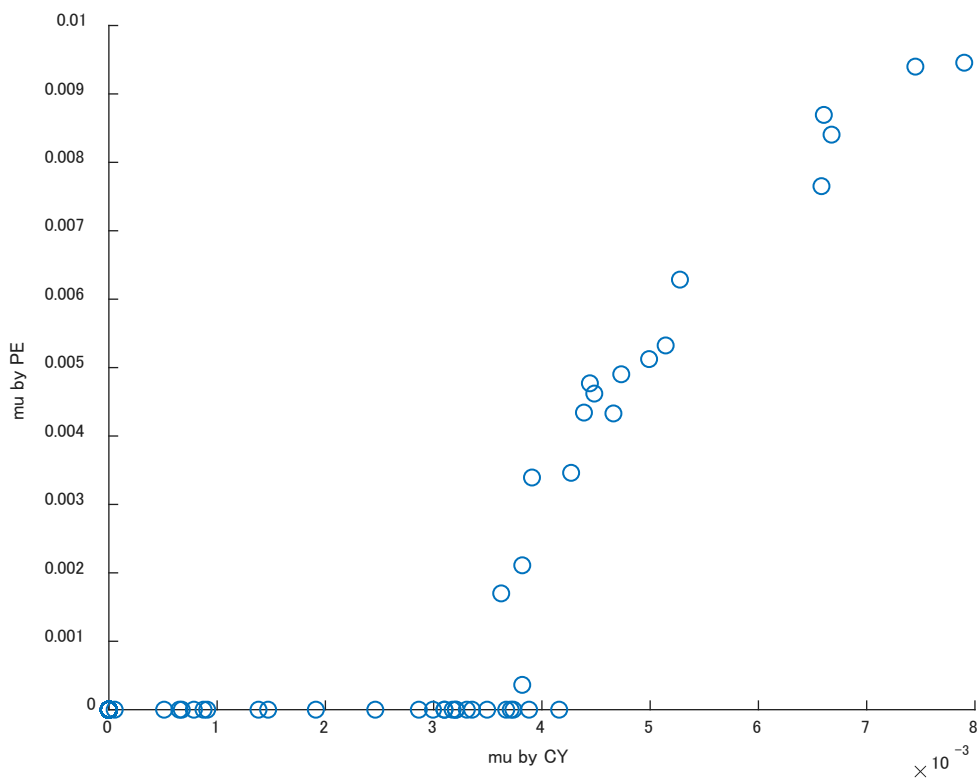
表 3： 検証期間中の購買回数と購買金額の予測精度

|      | 基準                    | CY      | PE      | Pareto/NBD |
|------|-----------------------|---------|---------|------------|
| 購買回数 | correlation           | 0.6446  | 0.6197  | 0.5770     |
|      | MSE                   | 10.4191 | 10.9865 | 16.8072    |
| 購買金額 | correlation           | 0.7423  | 0.7343  | 0.5345     |
|      | MSE ( $\times 10^9$ ) | 1.4038  | 1.4466  | 2.2449     |

まとめると、購買回数と購買金額に関しては、集計と非集計いずれのレベルにおいても、個人レベルの2つのモデル(CYとPE)は混合分布を使ったモデル(Pareto/NBD)よりパフォーマンスが大きく優れていた。

個人レベルの2つのモデルを比較すると、データへのフィットと予測精度に関しては、周期性を仮定したCYはランダム購買を仮定したPEより若干、優れていたが、その違いは小さかった。しかしながら、パラメータの推定値は大きく異なり、CYと比較してPEの顧客別離脱率 $\mu$ は、0の値が圧倒的に多く、推定が不安定(標準誤差が大きい)であった(図9)。その理由は、CYモデルの推定が、4つ以上の購買間隔(つまり購買回数5回以上)とリセンサーのデータを使っているのに対して、PEモデルではRFデータに含まれる3つの情報(購買回数、推定期間、リセンサー)だけから推定されているためである。

図9: CYとPEの顧客別離脱率

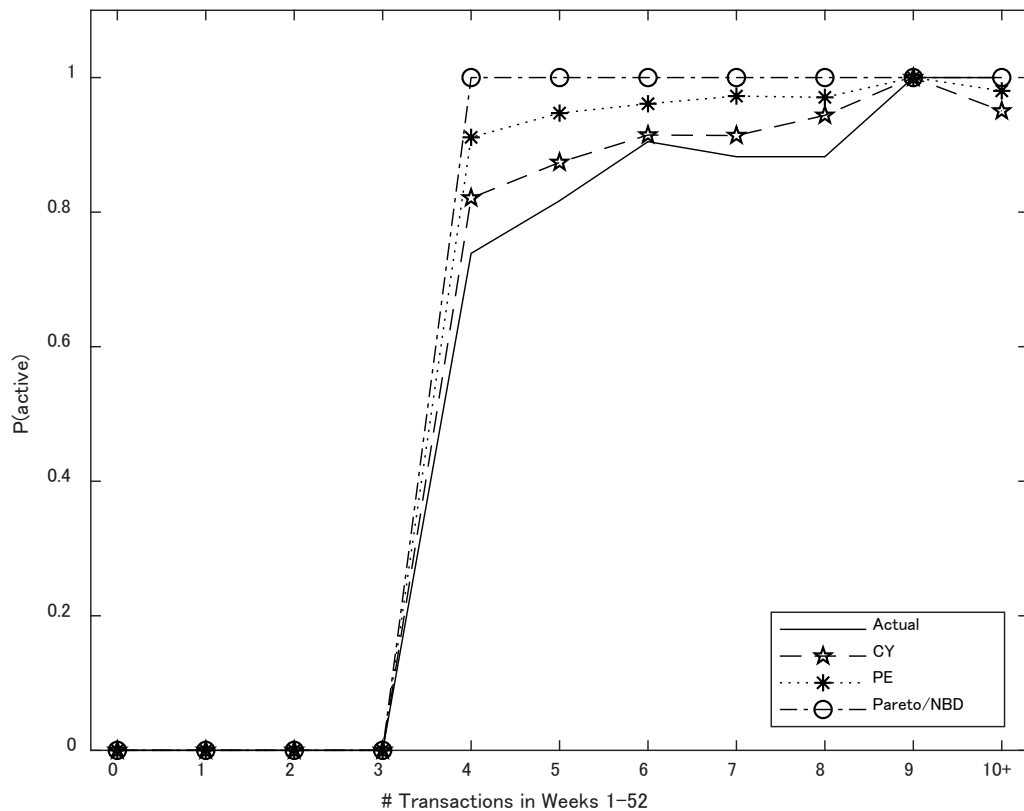


Pareto/NBDがCYやPEより大きく劣る理由は、離脱率 $\mu$ の異質性を表す混合分布の推定に問題があるためである。PEモデルによる顧客別離脱率 $\mu$ の推定値に0が多く含まれていたため、それにフィットするガンマ混合分布のシェイプ・パラメータは1より小さく( $7.2 \times 10^{-6}$ )、離脱率の最頻値は0であった。つまり、大多数の顧客は離脱率が極端に低く推定されている。それを反映して、推定期間直後の顧客の平均生存確

率が、Pareto/NBD では、ほぼ 1.0 となっている（図 10）。対象顧客が購買頻度の高いヘビーユーザーのみなので、ガンマ分布では異質性を柔軟にフィットさせられなかったことに原因があるようだ。

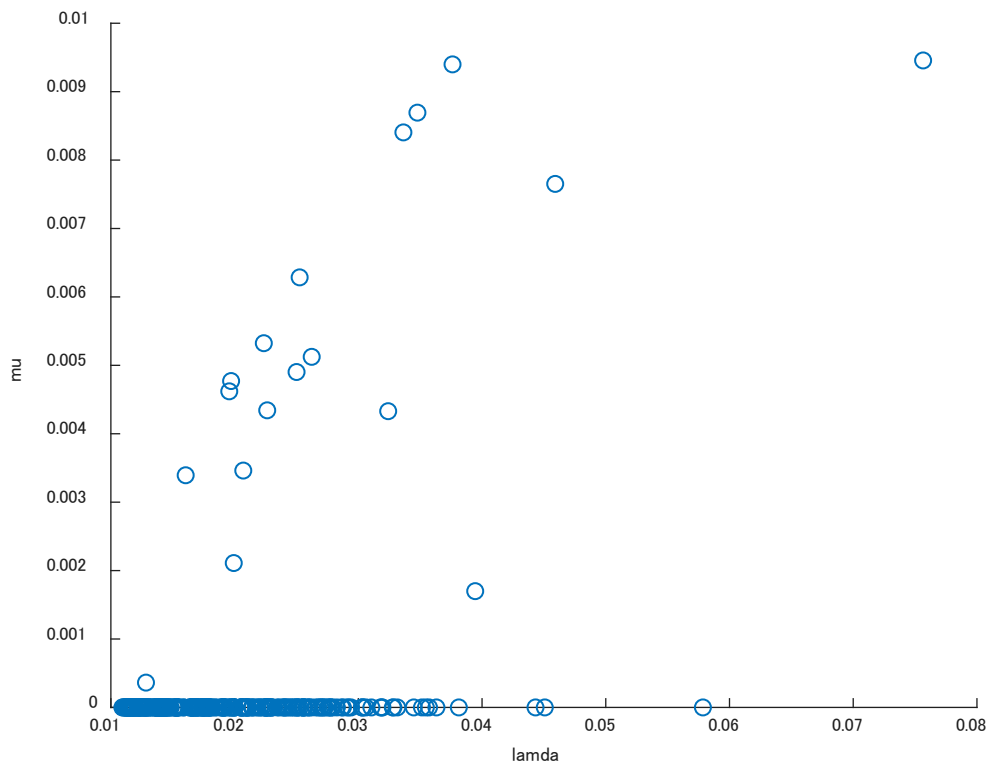
離脱率を顧客別に推定している CY と PE では混合分布が存在しないため、パラメトリック関数の形状による制約が課せられず、そのような問題は起きない。図 10 の Actual は、推定期間中に購買が無かった顧客は全員、離脱したと見なした場合の生存確率なので、実際（観測不可）より過少な値をとることに注意して欲しい。

図 10： 推定期間直後の顧客別生存確率



さらに、Pareto/NBD では $\lambda$ と $\mu$ の混合分布は独立と仮定されているが、PE が推定した顧客レベルの $\lambda$ と $\mu$ の散布図（図 11）から、その仮定が満たされていない（相関係数は 0.41）ことが分かる。

図 11： PE が推定した顧客レベルの $\lambda$ と $\mu$ の散布図



以上、3モデルの検証結果から、データへのフィットと予測精度に関しても、パラメータの推定に関しても、購買の周期性を考慮した個人レベルのモデルが一番優れていたため、この先はCYモデルの推定結果を考察する。

## 5. 結果

表4はCYモデルから算出した20人（CLVがトップ10とボトム10）の顧客レベルのマーケティング指標である。これらの指標は、購買、離脱、金額モデルのパラメータ $\{a, b, \mu, m, \omega^2\}$ 、一回当たりの平均購買金額 (avespend)、周期性 (cyclicality)、観測終了時点での生存確率 (pactive)、検証期間中の期待購買回数 (ce)と期待総購買金額(me)、そしてCLVである。CLVは割引率を年率15%に設定して算出している。最後の3行は、これら指標の対象顧客307名の平均値、最大値、最小値である。

表4： 顧客レベルのマーケティング指標

| ID      | $a$      | $b$      | $\mu$    | $m$      | $\omega^2$ | avespend | cyclicity | pactive  | ce       | me       | CLV      | coupon   |
|---------|----------|----------|----------|----------|------------|----------|-----------|----------|----------|----------|----------|----------|
| 1       | 6.330382 | 0.230804 | 3.56E-14 | 16603.58 | 0.295763   | 19249.79 | 0.347092  | 1        | 12.77135 | 245845.9 | 1832483  | 0        |
| 2       | 5.829972 | 0.159729 | 3.67E-14 | 21008.8  | 0.128693   | 22405.08 | 0.376884  | 1        | 9.472844 | 212239.9 | 1602748  | 0        |
| 3       | 4.697085 | 0.153495 | 3.49E-14 | 15985.02 | 0.125565   | 17020.77 | 0.467785  | 1        | 11.39509 | 193953.3 | 1452264  | 0        |
| 4       | 6.071109 | 0.186609 | 3.91E-14 | 16324.73 | 0.177528   | 17840.03 | 0.361915  | 1        | 10.68833 | 190680.2 | 1431733  | 0        |
| 5       | 10.22071 | 0.241459 | 3.47E-14 | 23047.37 | 0.007711   | 23136.41 | 0.214978  | 1        | 8.099294 | 187388.5 | 1427114  | 0        |
| 6       | 7.027876 | 0.236249 | 3.34E-14 | 13966.93 | 0.299692   | 16224.77 | 0.312644  | 1        | 11.73625 | 190417.9 | 1424049  | 0        |
| 7       | 4.169653 | 0.096995 | 3.37E-14 | 13239.85 | 1.091553   | 22851.29 | 0.526956  | 1        | 7.967363 | 182064.5 | 1387905  | 0        |
| 8       | 9.650654 | 0.585385 | 2.87E-14 | 8530.75  | 0.009398   | 8570.93  | 0.227676  | 1        | 21.57945 | 184956   | 1357419  | 0        |
| 9       | 14.84113 | 0.552757 | 3.48E-14 | 13301.34 | 0.037045   | 13550.01 | 0.14805   | 1        | 13.05722 | 176925.4 | 1317675  | 0        |
| 10      | 9.919763 | 0.308    | 3.1E-14  | 15709.23 | 0.02783    | 15929.35 | 0.2215    | 1        | 10.80192 | 172067.6 | 1291364  | 0        |
|         |          |          |          |          |            |          |           |          |          |          |          |          |
| 298     | 1.728396 | 0.054513 | 0.004661 | 4006.081 | 0.319892   | 4700.926 | 1.271251  | 0.016686 | 0.083873 | 394.2794 | 29396.27 | 3182.617 |
| 299     | 1.550122 | 0.030467 | 0.005143 | 8100     | 0          | 8100     | 1.417453  | 0.047716 | 0.130444 | 1056.599 | 28808.61 | 3129.335 |
| 300     | 1.5896   | 0.033091 | 0.004388 | 6532.268 | 0.00424    | 6546.132 | 1.38225   | 0.023757 | 0.078005 | 510.6286 | 28560.04 | 3145.834 |
| 301     | 5.164415 | 0.105657 | 0.003175 | 4536     | 0          | 4536     | 0.425455  | 0.179421 | 0.702369 | 3185.944 | 26078.57 | 2876.059 |
| 302     | 1.97063  | 0.041917 | 0.00427  | 5391.403 | 0.006077   | 5407.809 | 1.114986  | 0.038388 | 0.131623 | 711.7918 | 24721.89 | 2983.294 |
| 303     | 1.851191 | 0.029434 | 0.003908 | 6044.274 | 0.001514   | 6048.852 | 1.186925  | 0.068248 | 0.176588 | 1068.157 | 22411.39 | 2851.091 |
| 304     | 0.908227 | 0.017802 | 0.006604 | 7685.743 | 0.047646   | 7871.038 | 2.419246  | 0.016801 | 0.036957 | 290.8919 | 22080.05 | 2890.702 |
| 305     | 12.13719 | 0.202164 | 0.003205 | 3240     | 0          | 3240     | 0.181032  | 0.146483 | 0.45094  | 1461.045 | 15039.77 | 2354.994 |
| 306     | 12.13719 | 0.202164 | 0.003205 | 3240     | 0          | 3240     | 0.181032  | 0.146483 | 0.45094  | 1461.045 | 15039.77 | 2354.994 |
| 307     | 0.996714 | 0.023422 | 0.00745  | 1080     | 0          | 1080     | 2.204469  | 0.022027 | 0.053854 | 58.1624  | 3240.152 | 928.6464 |
|         |          |          |          |          |            |          |           |          |          |          |          |          |
| average | 7.540367 | 0.160521 | 0.000506 | 9.09584  | 0.12344    | 10504.25 | 0.486755  | 0.888215 | 6.250155 | 66525.39 | 509290.4 | 487.8623 |
| max     | 47.75274 | 1.143904 | 0.007901 | 10.08396 | 2.025695   | 26302.11 | 3.52041   | 1        | 21.57945 | 245845.9 | 1832483  | 4759.155 |
| min     | 0.624139 | 0.017802 | 2.54E-15 | 6.984716 | 0          | 1080     | 0.046013  | 0.000529 | 0.002947 | 18.75062 | 3240.152 | 0        |

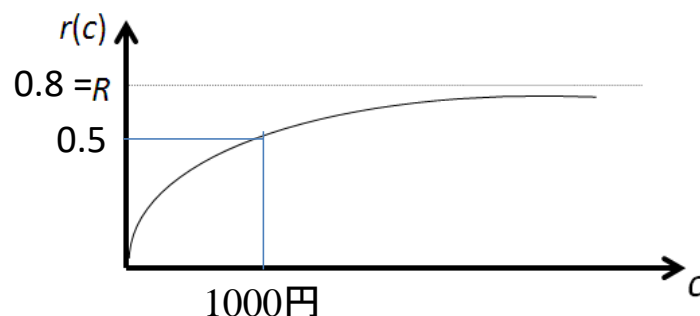
表4の右2列目のCLVから一列ずつ左に向かって解釈していこう。307名の平均顧客生涯価値は約51万円だが、最大は約183万円、最小は3240円である。検証期間中(2016/7/1~2017/6/30)に予測される総購買金額は、平均で66500円(最大245800円、最小19円)である。同様に、検証期間中に予測される購買回数は、顧客によって大きく異なるが、平均で6.25回(最大21.6回、最小0.003回)である。推定期間終了の

2016年7月1日時点における生存確率は顧客によって0.00053~1.0と大きく異なるが、平均は0.8882なので、1年後の平均離脱率は11.2%になる。顧客ベースと呼ばれる生存している顧客の期待人数は、273 (= 307 × 0.8882) 名である。

cyclicality(周期性)を見ると、典型的な顧客の場合、次回の購買は平均購買間隔の1.49倍以内に90%の確率で発生する。ただし次回の購買が90%の確率で発生する期間が、平均購買間隔の1.05倍である規則的な顧客から、平均購買間隔の4.52倍である非規則的な顧客まで、個人差は大きい。一回あたりの期待購買金額は、平均が10504円、最大が26302円、最小が1080円となっている。残り左5列は、パラメータ $\{a, b, \mu, m, \omega^2\}$ の推定値である。

顧客維持介入として、顧客が離脱していた場合、投資レベル $c$ によって確率 $r(c)$ で呼び戻せる図8のような指数曲線を考える( $k=0.000981$ )。3.4節で述べたように、この指数曲線の形状は、実際には決定解析法などで推定する。図8の解釈は、1000円のクーポン(投資)を送付することによって離脱顧客の半分は呼び戻せるが、いくら高額なクーポンを送付しても救出率は80%で頭打ちになる、というものである。

図8. 維持介入投資額 $c$ と救出率 $r(c)$ の指数関数



推定期間直後の2016年7月1日時点に介入する場合、CLVの増加が最大になる介入レベル $c^*$ が表4の一番右の列(coupon)である。トップ10の顧客はいずれも離脱の確率が0なので、介入すべきでない、つまり投資の回収がCLVの増加では見込めない $c^*=0$ 円であった。一方、ボトム10の顧客はこの時点での生存確率が低いため、介入による救済がコストに見合う以上のメリットをもたらす。最適レベルは顧客によって0円から4,760円と異なり、平均は488円である。最適介入レベルの公式9から、CLVが高く生存確率の低い顧客には、より強い介入が効果的であることが分かる。

## 6. 結論

本研究では、使用するデータを RFM から個々の購買の時点とその金額を含んだ顧客購買履歴に拡張することによって、周期性のある購買行動にも適応した柔軟な顧客生涯価値モデルを提案した。先行研究が豊富な行動モデル BTYD に基づき、他の Pareto/NBD 系モデル同様、離脱はランダム、1 回当たりの購買金額は顧客内で対数正規分布にしたがうと仮定する一方、購買は周期性に対応するよう新たなモデルを提案した。

購買は、誤差を含んだ効用が時間とともに線形で増加し、それが閾値を超えた時に発生するロジスティック閾値モデルを仮定した。このシンプルなモデルは、パラメータの推定が安定している他に、(1) 購買間隔に影響を与える説明変数を含めることができる、(2) 時間に対しての増加率に説明変数を組み込んだ階層化に拡張できる、(3) 顧客別に推定される 2 つのパラメータから、その顧客の購買サイクルの規則性を指標化できる、などの利点がある。

対象がヘビーユーザーであれば、推定には最尤法が使える。購買回数の少ないライトユーザーが対象であればベイズ推定法が必要になる。推定したモデルからは、CRM に有用な様々な顧客別のマーケティング指標が得られる。これらは、一回当たりの期待購買金額、周期性、観測終了時点での生存確率、検証期間中の期待購買回数と期待総購買金額、そして CLV である。

ヘアサロンの FSP メンバーの購買（訪問）履歴データを用いて、購買の周期性を考慮した個人レベルの CY モデルを、ランダム購買を仮定した個人モデル(PE)と混合分布による集計モデル (Pareto/NBD)とで比較した。その結果、データへのフィットと予測精度に関しても、パラメータの推定精度に関しても、提案モデルが Pareto/NBD や PE より優れていた。

そして、CRM での実装を想定し、顧客維持介入への応用例を示した。ここでは、顧客が離脱していた場合、介入レベル  $c$  によって  $r(c)$  の確率で呼び戻せる場合、CLV の増加が最大になる最適介入レベル  $c^*$  を顧客別に算出した。

今後の研究拡張の方向としては、購買間隔に影響を与える説明変数を含めたり、時間に対しての増加率に関する説明変数を組み込む階層化などであろう。また、本研究では単純化のために、消費者行動の購買／離脱プロセスと金額プロセスが顧客内では独立であると仮定したが、この 2 つのプロセスの顧客内における相関を考慮することは、CLV をより正確に推定するために役立つであろう。例えば Jen, Chou, and Allenby (2009) では、金額の高い購買の後には次回の購買が遅れるという現象を観測している。



## 付録

### 1. 生存確率（式 4）と尤度関数（式 10）の導出

生存確率は、ベイズルールを使って購買履歴から式 A1 のように導出できる。

$$\begin{aligned}
 P(\tau > T | \text{購買履歴}) &= P(\text{alive} | \text{history}) \\
 \text{(A1)} \quad &= \frac{P(\text{alive} \& \text{history})}{P(\text{alive} \& \text{history}) + P(\text{dead} \& \text{history})}
 \end{aligned}$$

ここで、 $\tau$  は生存時間、 $T$  は観測期間である。

ロジスティック閾値モデルでは、購買間隔時間の累積密度関数  $G(t)$  と密度関数  $g(t)$  は式 A2 で表される。

$$\begin{aligned}
 G(t) \equiv p(t) &= \frac{e^{-a+bt}}{1 + e^{-a+bt}} \\
 \text{(A2)} \quad g(t) &= \frac{dG(t)}{dt}
 \end{aligned}$$

ここから式 A1 の分母の 2 項は、それぞれ以下になる。

$$\begin{aligned}
 P(\text{alive} \& \text{history}) &= P(t_n \text{ interpurchases} \& \text{nopurchase between } [T-r, T]) \\
 &= \prod_n g(t_n) \{1 - G(r)\} e^{-\mu r}
 \end{aligned}$$

$$\begin{aligned}
 P(\text{dead} \& \text{history}) &= \int_{t_x}^T P(t_n \text{ interpurchases} \& \text{nopurchase between } [t_x, t] \& \text{die at } t \in [t_x, T]) dt \\
 &= \prod_n g(t_n) \int_0^r \{1 - G(t)\} \mu e^{-\mu(T-r+t)} dt
 \end{aligned}$$

したがって、生存確率は式 4 になる。

$$\text{(4)} \quad P(\text{alive at } T) = \frac{1 - G(r)}{\{1 - G(r)\} + e^{\mu r} \int_0^r \frac{\mu e^{-\mu t}}{e^{-a+bt} + 1} dt}$$

生存関数は、式 A1 の分母なので、式 10 で表せる。

$$(10) \quad L(a, b, \mu | t_n, r, T) \propto \prod_n g(t_n) e^{-\mu T} \left[ \{1 - G(r)\} + e^{\mu r} \int_0^r \frac{\mu e^{-\mu t}}{e^{-a+bt} + 1} dt \right]$$

## 2. (式 5) の $E[\psi]$ where $\psi = \min(\tau, w)$ の導出

$$\begin{aligned} E[\psi] &= E[\psi | \tau > w]P[\tau > w] + E[\psi | \tau < w]P[\tau < w] \\ &= w e^{-\mu w} + \int_0^w t f(t | t < w) dt \cdot (1 - e^{-\mu w}) \\ &= \frac{1 - e^{-\mu w}}{\mu} \end{aligned}$$

$$\text{where } f(t | t < w) = \frac{\mu e^{-\mu t}}{1 - e^{-\mu w}} \quad (0 < t < w)$$

## 参考文献

Abe, Makoto (2009), “Counting Your Customers’ One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model”, *Marketing Science*, 28, 541-553.

Abe, Makoto (2016) “Deriving Customer Lifetime Value from RFM Measures: Insights into Customer Retention and Acquisition”, *From Little’s Law to Marketing Science: Essays in Honor of John D. C. Little*, (ed.) John R. Hauser and Glen L. Urban, MIT Press, chapter 5, 127-169.

Allenby, G. M., R. P. Leone & L. Jen (1999) “A Dynamic Model of Purchase Timing with Application to Direct Marketing”, *Journal of the American Statistical Association*, 94 (446), 365-374.

Blattberg, R. and J. Deighton (1996) “Manage Marketing by the Customer Equity”, *Harvard Business Review*, 74 (4), 136-144.

Batislam, E. P., M. Denizel, and A. Filiztekin (2007) “Empirical Validation and Comparison of Models for Customer Base Analysis”, *International Journal of Research in Marketing*, 24 (3), 201–209.

Fader, P. S., B. Hardie, and Lee, K. L. (2005a) “RFM and CLV: Using Iso-Value Curves for Customer Base Analysis”, *Journal of Marketing Research*, 42, 415-430.

Fader, P. S., B. Hardie, and Lee, K. L. (2005b) “Counting Your Customers” the Easy Way: An Alternative to the Pareto/NBD Model”, *Marketing Science*, 24 (2), 275-284.

Fader, P. S., B. Hardie, and J. Shang (2010) “Customer-Base Analysis in a Discrete-Time Noncontractual Setting”, *Marketing Science*, 29 (6), 1086-1108.

Gupta, S. (1991) “Stochastic models of interpurchase time with time-dependent covariates”, *Journal of Marketing Research*, 28 (1), 1-15.

Helsen, K., and D. C. Schmittlein (1993) “Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models”, *Marketing Science*, 12 (4), 395-414

Hoppe, D., and U. Wagner. 2007. “Customer Base Analysis: The Case for a Central Variant of the BetaGeometric/NBD Model”, *Marketing Journal of Research and Management*, 3 (2), 75–90.

Jain, D. C., & Vilcassim, N. J. (1991) “Investigating household purchase timing decisions: A conditional hazard function approach”, *Marketing Science*, 10 (1), 1-23.

Jen, L., C. -H. Chou, and G. M. Allenby (2009) “The Importance of Modeling Temporal Dependence of Timing and Quantity in Direct Marketing,” *Journal of Marketing Research*, 46, 482-493.

Ma, S. -H., and J. -L. Liu (2007) “The MCMC Approach for Solving the Pareto/NBD Model and Possible Extensions”, in *Third International Conference on Natural Computation (ICNC 2007)*, IEEE, vol. 2, 505-512.

Platzer, M., and T. Reutterer (2016) “Ticking Away the Moments: Timing Regularity Helps to Better Predict Customer Activity”, *Marketing Science*, 35 (5), 779-799.

Schmittlein, D. C., D. G. Morrison, and R. Colombo (1987), “Counting your customers: Who are they and what will they do next?”, *Management Science*, 33 (1), 1-24.

Schmittlein, D. C. and R. A. Peterson (1994), “Customer Base Analysis: An Industrial Purchase Process Application”, *Marketing Science*, 13 (1), 41-67.

Rosset, S., E. Neumann, U. Eick, and N. Vatnik (2003), “Customer Lifetime Value Models for Decision Support”, *Data Mining and Knowledge Discovery*, 7 (July), 321-339.

古川、守口、阿部 (2011)、「新版：マーケティング・サイエンス入門」、有斐閣.