CIRJE-J-121

CRM のデータ分析に理論とモデルを組み込む
消費者行動理論にもとづいた RF 分析

東京大学大学院経済学研究科
阿部　誠

2004 年 12 月

CRMのデータ分析に理論とモデルを組み込む

消費者行動理論にもとづいたRF分析

阿部　誠

東京大学　大学院経済学研究科・経済学部

流通情報

11月24日

Incorporating Theory and Modeling in Data Analysis for CRM: RF Analysis based on a Consumer Behavior Model

Abstract

While RFM analysis is popular among practioners, ad-hoc rules are often employed to judge whether customers are alive or not. Because customers do not declare explicitly when they are dead, a company infers a customer is dead if she did not make any purchase, for example, for over three months. Even with the same period of nonpurchase, however, customers with a long interpurchase time need not be worried for death whereas those with a short interpurchase time could be dead. Hence, it is very important to account for customer heterogeneity when assessing the survival of customers. In this research, using standard RF data, I will derive the survival probability of an individual customer based on the common hypotheses on consumer behavior.

１．はじめに

需要過剰の市場では、単に質が高くて価格が適切なだけでは不十分である。　個々の消費者をよりよく理解し、彼らのニーズに細かく応え、それを超えた製品やサービスを提供しなければ、消費者に選択・購買してもらうことができない。　成熟した経済社会で特に重要なことは、個々の消費者（顧客）の違い―商品に関する選好やマーケティング刺激に対する反応の異質性―を十分に認識し、それに適切に対応することである。　マーケティングでは、セグメンテーション、ポジションニング、ターゲティングと呼ばれる差別化された商品の提供や顧客によって異なったマーケティング活動などが、早い時期から行われてきた。

近年の情報技術の発達によって大量の顧客データを容易に収集、保存できるようになったが、多くの企業はその分析に手を焼いている状況である。　データマイニングなどによる自動化された全数分析が「花」の時代だ。　しかしながらデータマイニングも現場の仮説が生かされないと、得られる知見があまりにも当然か、あるいはたまたまデータのノイズに反応した再び起こり得ないような突飛なもので、十分な効果を上げられないことが多い。　「顧客はひとりひとり違う」ことを認識して CRM や One-to-One などの個に対するより精緻なマーケティングを実践するためには、顧客の特性を個人別に推定することが重要になる。　するとあれほど大量にあったデータでさえ、顧客一人当たりで見るとその情報量は限られており、データは逆に足りないというパラドックスが生じるのである。　それでは、この個人別では少ないデータからその顧客の特性を探るにはどうすればよいのであろうか？　それは、データに理論やモデルを組み込むことである。

もうひとつのモデルの利点として、直接には観測されない変数の値を推測することができることが挙げられる。　そして多くの場合、その変数を知ることが実務上でも様々なメリットをもたらす。　代表的な例として、離散的選択モデルにおける効用値がある(Guadagni and Little 1983)。　観測されるのはブランドが選択される、されないの 0/1 の二項変数であるが、その背後にブランドに対する効用という連続的な潜在変数を仮定することによって、消費者のブランド選好への理解が深まり、マーケティング活動へのインプリケーションが得られる。

CRM で重要な概念である顧客の生涯価値(customer lifetime value)を計算するには、顧客の離脱率または維持率を把握することが必要である (Blattberg and Deighton 1996)。しかし離脱する顧客は単に購買を止めるだけで、特に年会費などの支払い義務がなければ、わざわざ離脱を申告することは稀だ。　通常このような場合、企業は独自の経験則に基づいて、例えば顧客が３ヶ月購買しなければ離脱したと判断したりする。　実務家の間でよく使われる RFM 分析では、(RECENCY = 3 ヶ月)のようなアドホックで一

律なルールが基本になっているが、ここには２つの大きな問題がある。　第１に、このルールが主観的なことである。 なぜ 2 ヶ月や 4 ヶ月でなく､3 ヶ月なのだろうか？ 生死は 0/1 の変数で表せるが、ブランド選択の効用と同様、RECENCY に関連した連続的な潜在変数を仮定することによってその閾値から生死の判断を客観的に下せないであろうか？　2 つ目の問題は、マーケティングのスピリットを忘れて、全ての顧客は同質と見なしていることである。　同じ 3 ヶ月の RECENCY でも、購買間隔が長い顧客は離脱の心配が無いが､購買間隔が短い顧客は離脱している可能性が高いであろう。つまり離脱率の推測に顧客の異質性に配慮する必要があるだろう。

今回の論文では RFM 分析で使われる RF の部分のデータから、一般的な消費者行動の仮定に基づいて、ある時点での顧客の生存確率を理論とモデルに基づいて推定する。図１は、企業にとっての顧客の魅力度(貢献度)を BCG のポートフォリオ・マトリクスとして、既存の RF 分析を描いたものである。　リーセンシーと購買頻度(フリークエンシー)は独立に考慮され、あるリーセンシーの閾値(例えば 3 ヶ月)を境に優良顧客(Star と Cash Cow)と離脱顧客(Problem Child と Dog)が分かれている。　これに対して、RECENCY と FREQUENCY の両方を同時に考慮すると、図 2 のように解釈が変わってくる。　既存の分析では Star と区分されていた購買頻度が高くかつ最近購買した顧客(左上の区分)の中でも、購買頻度に見合うようなリーセンシーを示さない顧客は Problem Child として早急な対応が求められる。　また、既存の分析では Dog と区分されていた購買頻度が低くかつ最近購買していない顧客(右下の区分)の中でも、購買頻度に対して比較的最近購買した顧客は、企業に予想外の売上げをもたらしたとして Cash Cow に分類される。　既存の RF 分析では、これら 3 角形の 2 セグメントの顧客に対して企業は特別な配慮をしなかったため、利得機会を失っていた。

< 図 1 >

< 図 2 >

本研究では、最終購買からどのくらいの期間が経っていたら離脱の危険性が高いのかを、その顧客の購買頻度を考慮にいれて求める。　実務上のメリットとしては大きく 2 つある。　ひとつは、通常は観測されない離脱の確率が顧客別に得られるため、一人一人の顧客生涯価値がより正確に計算できる。　2 つ目は、現時点での生存確率が顧客別に得られるため、それを足し合わせれば現在のアクティブな顧客の期待総数、つまりその企業の持つ顧客ベースの計算が可能になる (Schmittlein, Morrison and Colombo 1987)。　顧客ベースは登録会員数などと違って企業の会計データには直接は反映されないので､財務諸表を補う長期的な観点に立った企業の診断指標のひとつとして非常に有益である。

2. モデル

2.1. 消費者行動の仮定

[仮定 1] 顧客の生存時間は指数分布に従う

これは、離脱(死亡)が過去の生存期間に関係なくランダムに起きるというメモリレス・プロセスを意味する。　この仮定の妥当性は、離脱が企業に対する飽き、競合企業への乗り換え、転居、死去などの様々な理由によって起きることと、一度購買が観測されるということは顧客の生存が確認されて離脱プロセスがリセットされることから支持される。　生存時間をτとすると、この確率分布は(1)式のように表される。

(1) $$f(\tau) = \mu e^{-\mu\tau} \qquad \tau \geq 0$$

ここでは、μ は指数分布のパラメータで、E[τ]=1/μ なので、μは大雑把に「離脱率」を表すと解釈できる。

[仮定 2] 購買はポアソン・プロセスに従う

この仮定では、購買が過去に、何時起きたかに関係なくランダムに発生することを意味する。　Ehrenberg (1972, 1988) の研究以来、この 0 次の購買プロセスは多くのデータでロバストなことが確認されている(Bass, Givon, Kalwani, et. al 1984)。　ただしこの仮定は周期性のある購買には当てはまらないため、単一カテゴリーよりは多カテゴリーの購買行動、例えばデパートでの購買などに適用するべきである。　τ期間以上生存した顧客のT期間の購買回数xは(2)式で表される。

(2) $$P[x \mid \lambda, \tau > T] = \frac{(\lambda T)^x}{x!} e^{-\lambda T} \qquad x = 0, 1, 2, ..$$

ここでは、λはポアソン・プロセスのパラメータで、E[x]=λT なので、λは「単位期間あたりの購買頻度」と解釈できる。

[仮定 3] 顧客の異質性

生存時間の分布と購買頻度の分布を表すパラメータμとλの値は、顧客ごとに異なる。

2.2. 生存確率の導出

2001 年 1 月 1 日から 2002 年 1 月 31 日までの観測期間中における 1 人の顧客の購買履歴に基づいて、本日 2002 年 1 月 31 日時点のこの顧客が生存している確率を考えてみよう。　初回の購買は 2 月 14 日に、4 回目の購買が 12 月 1 日に発生したとして、この顧客の購買履歴を視覚化したものが図 3 に示されている。　*t* は初回から最終の購買までの期間、*T* は初回の購買から 1 月 31 日までの期間を、そして *x* は観測期間中の購買回数を表す。　リーセンシーは(*T*-*t*) で表されるので、この顧客の場合は 60 日である。

< 図 3 >

この表記に基づいて、この顧客が 1 月 31 日に生存している確率、つまり生存期間τが *T* より大きくなる確率を第 2.1.節の 3 つの仮定から導くと(3)式のように表せる。詳細は付録を参照していただきたい。

(3)
$$P[\tau > T \mid \lambda, \mu, T, t] = \frac{1}{1 + \frac{\mu}{\lambda + \mu}\left[e^{(\lambda+\mu)(T-t)} - 1\right]}$$

最終購買が 1 月 31 日に発生していれば、リーセンシーは 0 なので生存確率は 1 になる。　最終購買が 1 月 31 日以前の場合の生存確率は、リーセンシーの他にこの顧客のμとλが分かれば計算できる。 μとλは仮定 3 で示したように購買履歴から顧客別に推定する。

また、顧客の生涯価値を計算するための 1 年後の維持率は、(1)式のτの分布に基づいて 1 年間以上生存する確率を積分すると exp(-μ)になる。

2.3. モデルの推定

仮定 1 と 2 の生存プロセスと購買プロセスから図 3 のような購買履歴 (*x*, *t*, *T*) が観測される確率は(4)式の尤度関数で表される(付録参照)。

(4)
$$L(x, t, T \mid \lambda, \mu) = \lambda^x \left\{ \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)T} + \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t} \right\}$$

この尤度を最大にするパラメータμとλが最尤推定値である。　顧客ごとに 2 つのパラメータを 3 つの購買履歴データ (*x*, *t*, *T*) から推定するため、理論的には可能である。

本論文では、最尤推定値をグリッド・サーチによって直接求めた。

3．FSP データへの応用

3.1．データの概要

上記のモデルを、某百貨店のフリークエント・ショッパー・プログラム(FSP)によって収集された購買データに適用して、顧客の生存確率を求めてみた。　観測期間は 2001 年 1 月 1 日から 2002 年 1 月 31 日までで、50 人の顧客を無作為に抽出した。 (x, t, T) という 3 つの購買データから、顧客別にμとλの最尤推定値を計算し、その結果に基づいて顧客の 1 年後の維持率と 2002 年 1 月 31 日時点での生存確率と求めた。

3.2．推定結果

表 1 の各行は 1 人の顧客に該当し、列はそれぞれ顧客 ID、購買データ (x, t, T)、μとλの推定値、2002年1月31日時点の生存確率、そして1年後の維持率が報告されている。1年に数回しか購買しない顧客もいれば、2日に1回ちかく購買するような顧客もいる。この百貨店は都心のターミナル駅に直結しているため、仕事帰りにデパ地下で食品を買う客も多い。　このような購買頻度の大きな違いをリーセンシー・データの分析に反映させて生存確率を計算することの有用さは明白であろう。

<表 1>

図4は顧客別に推定された50組のμとλの値を散布図としてプロットしたものである。広い範囲での散らばりからも、顧客の異質性を考慮することがいかに重要であるかが分かる。

< 図 4 >

約半数の顧客のμが 0、つまり永久に離脱しないと推定されている。　一見不自然なようであるが、モデル構造を考慮すると納得できる結果である。　このモデルでは離脱データの無い状況において理論に基づいて離脱が予測されているため、通常の購買サイクルがしばらく途切れると離脱の可能性が高いと判断される(付録参照)。　したがって、この擬似的な離脱、具体的にはリーセンシーの期間が最低でも 3〜4 回分の購買サイクルになる状態、が観測されない場合は離脱していないと見なされ、生存期間を表す指数分布のパラメータμが 0 と推定される。

それを踏まえた上で、サンプル全体の平均記述統計を報告しておく。　$1/\lambda$の平均を

計算することによって平均購買間隔は 12.2 日、$1/\mu$の平均を計算することによって顧客の平均維持期間は 2.2 年、また 1 年後の平均維持率は 79.1%であることが分かる。この百貨店の2002年1月31日時点における顧客の平均生存確率は0.74で、これは 50 人のサンプルに基づいた顧客ベースでは 37.1 人に該当する。

4．結論

顧客の離脱は直接には観測されない。　企業は RECENCY を使った経験則(例えば 3 ヶ月購買がなければ離脱)に頼って、この判断を下しているのが現状である。　本論文では、既存の RF 分析に消費者行動理論に基づいたモデルを組み込むことによって、消費者の異質性を考慮し、この観測されない離脱の確率を推定することが可能であることを提案した。　仮定された行動理論は、過去のマーケティング研究で十分に検証され、ロバストであるポアソン購買プロセスとメモリレスな生存プロセスである。

表 1 のような結果は、実務での CRM や One-to-One マーケティング活動に非常に有益である。　たとえば 2002 年 1 月 31 日の時点で生存確率が0.3~0.7の離脱過程にある顧客群に対してクーポンなどの DM を送付することによって、販促活動の効率化を計ることができる。　生存確率を継続的に更新・モニターすることによって、生存確率がある閾値を下回った顧客全員に DM を送付することによって離脱を未然に防ぐようなアクションなどもとれる。　また、顧客別の維持率に基づいて顧客生涯価値を計算すれば、より正確な CRM が実践可能になる。

この論文で紹介したモデルの限界を述べておく。　まずは言うまでもないが、消費者行動の仮定 1 と 2 が当てはまらない状況では、このモデルは機能しない（Chatfield and Goodhardt 1973)。　したがって、適切な業界やカテゴリーを選択することが必要である。　第 2 に、3.2 節で触れたが、擬似的な離脱状態が観測されないとその顧客のμのパラメータが 0 と推定されてしまうことである。　第 3 に、パラメータの数に対してデータ数が非常に少ないため、推定の標準誤差が大きいのは容易に想像できよう。問題はその大きさが分からないために、誤差がどれほど深刻なのか、そしてそれに基づいてマーケティング・アクションを起こした時のリスクが評価できない。　上記の 2 番目と 3 番目の問題は、より複雑なベイズ推定法を用いることによって克服できる(Abe 2004)。

本研究では、購買履歴データが入手できる既存顧客へのアクションという形で CRM への応用が念頭に置かれている。　購買履歴の存在しない新規顧客に対してのマーケティング活動の提案として、このモデルを拡張できないであろうか？　既存顧客のデータを使って購買と生存プロセスのパラメータμとλをデモグラフィック変数などの比較的容易に入手できる情報と関連付けることができれば、優良顧客のプロファイルを

購買履歴なしで後者の情報のみから形成することが可能になる。　これには階層ベイズ法を用いることで対応できる。

IBM や SAP に見られるように、今やマーケティングは情報産業である。　顧客データを有効に生かし、情報処理を差別化することによって新たな知識を創造する企業こそが、競合の一歩先を行き競争優位に立つことができる。　クロス表や記述統計などの集計分析、主成分分析のような情報集約、決定木やデータマイニングによる自動化データ探索を超えて、理論とモデルを組み込んだ顧客データの分析が、今後の CRM や One-to-One マーケティングに一層重要となるであろう。

付録：　生存確率と尤度関数の導出

まず、観測された購買履歴から顧客が生存している確率をベイズの定理に基づいて(A1)式のように表す。

$$
\begin{aligned}
P(\tau > T \mid \lambda, \mu, x, t, T) &= P(\text{生存} \mid \text{履歴}) \\
&= \frac{P(\text{生存} \,\&\, \text{履歴})}{P(\text{履歴})} \qquad \textbf{(A1)} \\
&= \frac{P(\text{履歴} \mid \text{生存})\, P(\text{生存})}{P(\text{履歴} \,\&\, \text{生存}) + P(\text{履歴} \,\&\, \text{死亡})}
\end{aligned}
$$

すると、生存時間が指数分布なので

$$
P(\text{生存}) = P(\tau > T) = e^{-\mu T}
$$

となる。　さらに、

$$
\begin{aligned}
P(\text{履歴} \mid \text{生存}) &= P(x\text{回目の購買が}t\text{に起きる}\,\&\,[t,T]\text{に購買が起きない}) \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda t} \times e^{-\lambda(T-t)} \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda T}
\end{aligned}
$$

$$P\left(\text{履歴}\,\&\,\text{死亡}\right)=\int_t^T P\left(x\text{回目の購買が}t\text{に起きる}\,\&\,[t,y]\text{に購買が起きない}\,\&\,y\in[t,T]\text{に死亡}\right)dy$$

$$=\int_t^T \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda\, t} \times e^{-\lambda(y-t)} \times \mu e^{-\mu y} dy$$

$$=\frac{\lambda^x t^{x-1}}{\Gamma(x)} \mu \int_t^T e^{-(\lambda+\mu)\,y} dy$$

$$=\frac{\lambda^x t^{x-1}}{\Gamma(x)} \frac{\mu}{\lambda+\mu} \left\{ e^{-(\lambda+\mu)\,t} - e^{-(\lambda+\mu)\,T} \right\}$$

を(A1)式に代入すると、下の式が得られる。

$$P\left(\tau > T \mid \lambda, \mu, x, t, T\right) = \frac{\dfrac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda T} \times e^{-\mu T}}{\dfrac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda T} \times e^{-\mu T} + \dfrac{\lambda^x t^{x-1}}{\Gamma(x)} \dfrac{\mu}{\lambda+\mu} \left\{ e^{-(\lambda+\mu)\,t} - e^{-(\lambda+\mu)\,T} \right\}}$$

$$= \frac{1}{1 + \dfrac{\mu}{\lambda+\mu} \left\{ e^{(\lambda+\mu)(T-t)} - 1 \right\}}$$

また、尤度関数は (A1) 式の分母のパラメータに依存する部分なので、下の式のように表せる。

$$L(x,t,T \mid \lambda,\mu) \propto P\left(\text{履歴}\right)$$

$$= \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda T} \times e^{-\mu T} + \frac{\lambda^x t^{x-1}}{\Gamma(x)} \frac{\mu}{\lambda+\mu} \left\{ e^{-(\lambda+\mu)\,t} - e^{-(\lambda+\mu)\,T} \right\}$$

$$= \frac{\lambda^x t^{x-1}}{\Gamma(x)} \left\{ \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)T} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} \right\}$$

参考文献

Abe, Makoto (2004), “ＲＭF Analysis based on a Consumer Behavior Model: A Hierarchical Bayes Approach,” presented at the 2004 Marketing Science Conference, Amsterdam.

Bass, Frank M., Moshe M. Givon, Manohar U. Kalwani, David Reibstein, Gordon P. Wright (1984), “An investigation into the order of the brand choice process,” *Marketing Science*, 2

(4), 267-187.

Blattberg, Robert C. and John Deighton (1996), "Manage marketing by the customer equity test," *Harvard Business Review*, 74 (4), 136-144.

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

Ehrenberg, A. S. C. (1972), *Repeat-Buying: Theory and Applications*, Amsterdam; North-Holland.

Ehrenberg, A. S. C. (1988), *Repeat-Buying: Facts, Theory and Data*, 2nd Ed. New York; Oxford University Press.

Guadagni, Peter M. and John D. C. Little (1983), "A logit model of brand choice calibrated on scanner data," *Marketing Science*, 2 (3), 203-238.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting your customers: Who are they and what will they do next?" *Management Science*, 33 (1), 1-24.

図１．既存の **RF** 分析

図 **2.** 本研究で提案している **RF** 分析

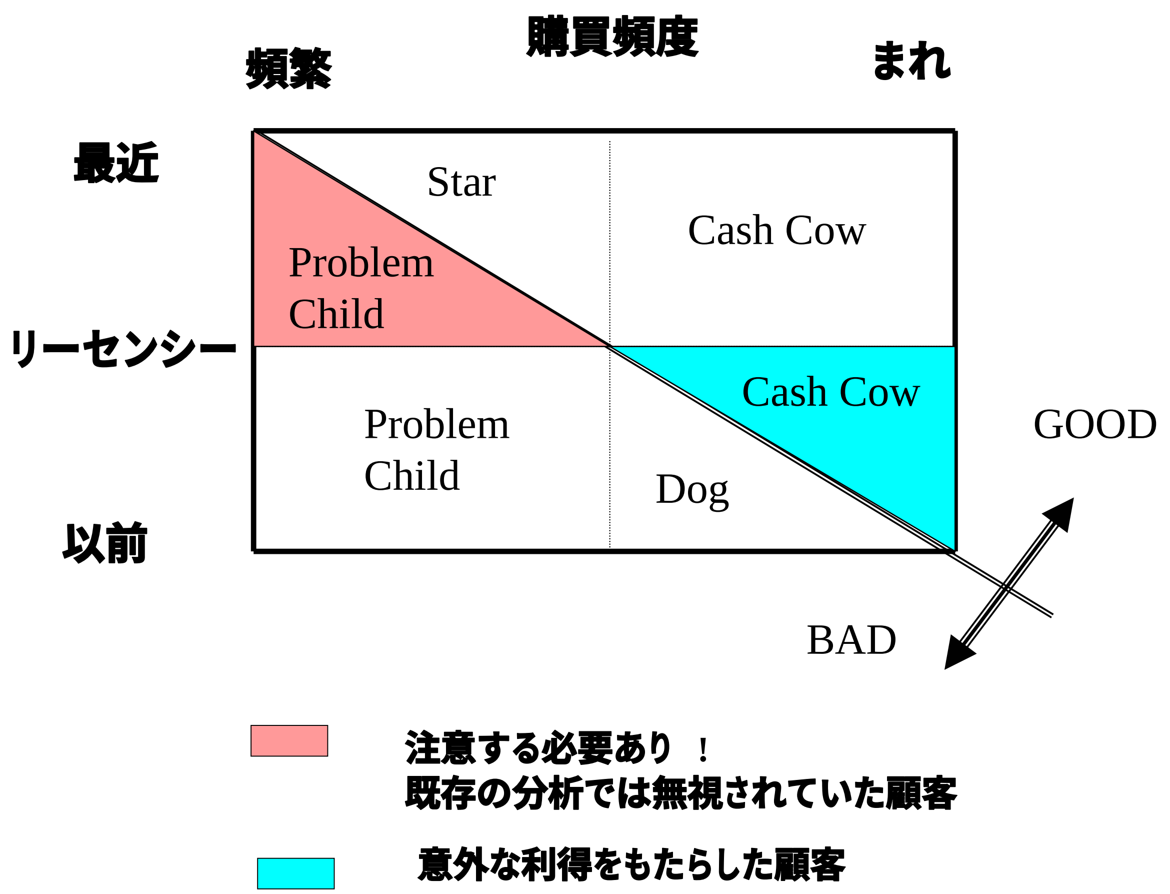

図 3.　購買履歴データ

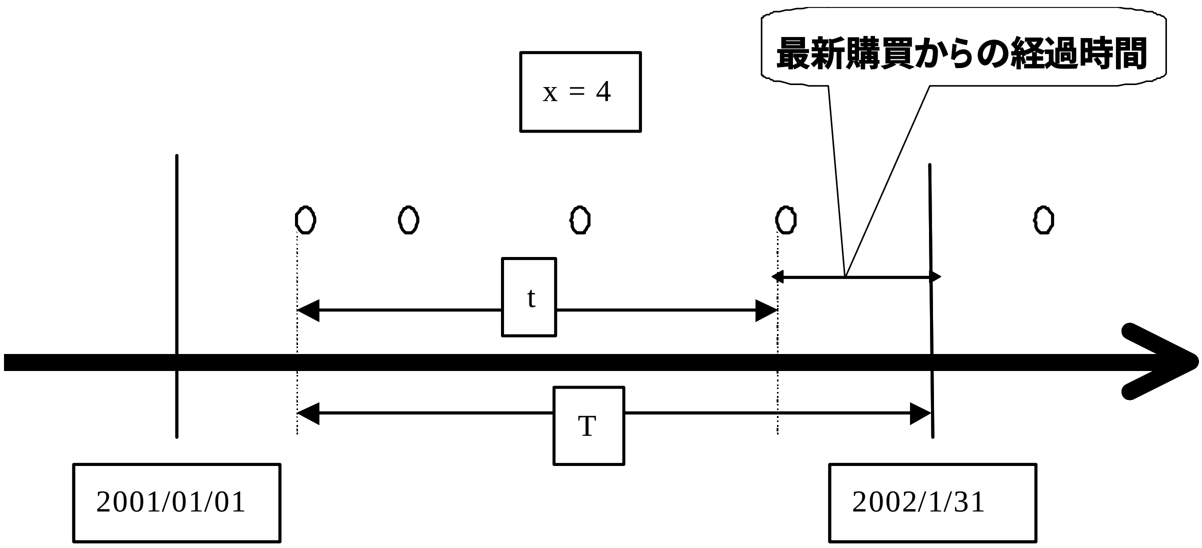

図 4.　顧客別に推定された50組のμとλの散布図

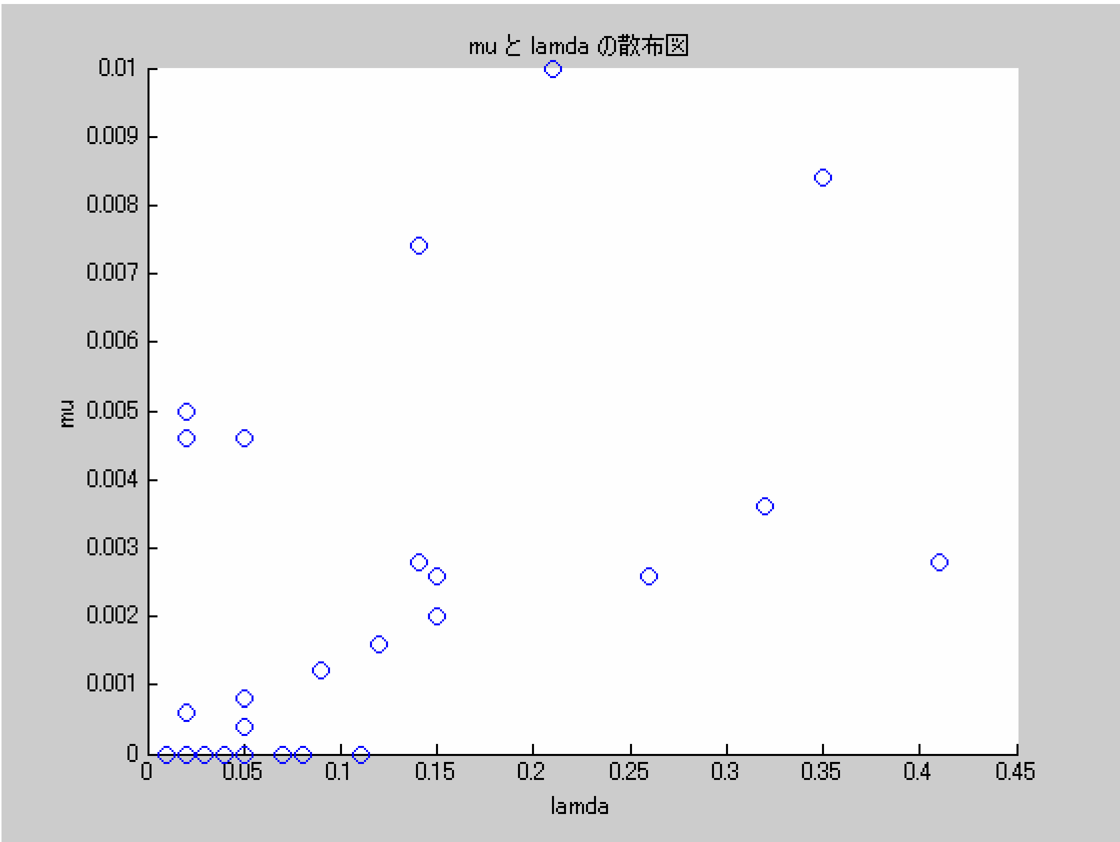

表１．推定結果

顧客ID	x	t	T	λ	μ	生存確率	1年後維持率
1	4	266	299	0.01	0	1.0000	1.0000
2	4	80	130	0.03	0	1.0000	1.0000
3	19	323	386	0.05	0.0004	0.8460	0.8642
4	20	348	379	0.05	0	1.0000	1.0000
5	26	332	372	0.07	0	1.0000	1.0000
6	53	346	381	0.15	0.002	0.2720	0.4819
7	33	350	390	0.08	0	1.0000	1.0000
8	6	295	346	0.02	0	1.0000	1.0000
9	12	322	373	0.03	0	1.0000	1.0000
10	50	339	389	0.14	0.0028	0.0389	0.3599
11	147	353	384	0.41	0.0028	0.0004	0.3599
12	11	357	388	0.03	0	1.0000	1.0000
13	43	347	385	0.12	0.0016	0.4304	0.5577
14	4	310	354	0.01	0	1.0000	1.0000
15	32	342	387	0.09	0.0012	0.5605	0.6453
16	19	355	386	0.05	0	1.0000	1.0000
17	93	359	390	0.26	0.0026	0.0286	0.3871
18	11	339	386	0.03	0	1.0000	1.0000
19	4	335	389	0.01	0	1.0000	1.0000
20	7	28	61	0.21	0.01	0.0152	0.0260
21	12	248	287	0.04	0	1.0000	1.0000
22	21	228	264	0.08	0	1.0000	1.0000
23	10	190	307	0.05	0.0046	0.0196	0.1866
24	90	274	309	0.32	0.0036	0.0011	0.2687
25	4	192	247	0.02	0	1.0000	1.0000
26	6	114	149	0.04	0	1.0000	1.0000
27	42	117	150	0.35	0.0084	0.0003	0.0466
28	9	326	379	0.02	0	1.0000	1.0000
29	6	340	381	0.02	0	1.0000	1.0000
30	41	355	387	0.11	0	1.0000	1.0000
31	14	278	318	0.04	0	1.0000	1.0000
32	16	275	337	0.05	0.0008	0.7399	0.7468
3	6	251	363	0.02	0.0006	0.7915	0.8033
34	4	207	322	0.01	0	1.0000	1.0000
35	19	130	306	0.14	0.0074	0.0000	0.0671
36	11	308	365	0.03	0	1.0000	1.0000
37	4	269	315	0.01	0	1.0000	1.0000
38	4	82	117	0.03	0	1.0000	1.0000
39	4	167	372	0.02	0.0046	0.0336	0.1866
40	5	159	198	0.03	0	1.0000	1.0000
41	4	336	389	0.01	0	1.0000	1.0000
42	4	141	314	0.02	0.005	0.0628	0.1612
43	7	224	276	0.03	0	1.0000	1.0000
44	43	345	378	0.11	0	1.0000	1.0000
45	3	231	383	0.01	0	1.0000	1.0000
46	44	280	315	0.15	0.0026	0.2203	0.3871
47	32	347	382	0.08	0	1.0000	1.0000
48	26	347	378	0.07	0	1.0000	1.0000
49	6	317	354	0.02	0	1.0000	1.0000
50	11	340	378	0.03	0	1.0000	1.0000
平均	22.12	270.88	327.5	0.0762	0.00122	0.7412	0.7907