

マーケティングにおける結果データ動的活用のためのベイジアンモデリング

佐藤 忠彦

マーケティング分野ではマイクロ・マーケティングと呼ばれる“個に”焦点を当てた活動がその有効性から注目されている。また、消費者の行動の結果であるデータの蓄積が進んでいる。マイクロ・マーケティングを高度化するためには、それらデータから個に関する詳細な情報を抽出し、その活動に生かすことが必要である。特に、結果データからの動的観点での情報抽出は今後重要度の増す領域である。それを実現する手段がベイジアンモデリングである。本稿ではそれらを背景としてなされた分析事例とマーケティング研究の今後の展望を示す。

キーワード：マイクロ・マーケティング，ベイジアンモデリング，データ同化

1. はじめに

今日、マーケティングを取り囲む様々な環境が大きく変化している。特に二つの変化が重要である。一つ目はデータ環境の変化であり、二つ目はマーケティング分野におけるビジネスモデルの変化である。

店舗型の小売業や Web 上の商用サイトでは、消費者の行動の結果であるデータが自動的にかつ大規模に蓄積されるようになってきている。そのデータの構造を模式的に示したものが図1である。通常、図1の消費者軸方向に集計したデータを POS データと呼び、“いつ”、“何を”、“いくらで”、“何個”販売したかの情報を得ることができる。小売業に限定して考えれば、POS データに関してはほぼ 100% の店舗でデータを取得できる。現在でも小売業の様々な意思決定にはこの POS データが用いられている。一方で、Frequent Shoppers Program (FSP) 実施小売業や Web 上の商用サイトの増加に伴い、POS データよりも詳細な情報をもつ ID 付き POS データと呼ばれるデータの蓄積が進んできている (図1)。ID 付き POS データには、“誰が”、“いつ”、“何を”、“いくらで”、“何個”購買したかという情報が含まれている。ただし、ID 付き POS データには詳細な情報は含まれているものの、全体としては非常にスパースなデータベース構造である点は、解析上留意しなければならない。

マーケティングが対象とする商品販売、サービス提供などの分野では、長年、多くの消費者に最大公約的に評価される商品 (サービス) に絞り大量生産・販売し、規模の経済の原理で多くの利益を創出していると考えられてきた (少品種大量生産・販売)。しかし、市場の成熟化と消費者嗜好の多様化に伴って、昨今、その形式が多品種少量生産・販売へと変化している。少品種大量生産・販売型のビジネスモデルで、利益を創出することが困難になってきているのである。そのため供給者は、ビジネスモデルを供給者サイドの論理によるビジネスモデル (良いものを作れば売れるはず) から、需要サイドの論理によるビジネスモデル (消費者の欲するものを作れば売れるはず) に、すなわち顧客視点のビジネスモデルに変化させていかねばならない。この変化に対応するために、メーカーは、消費者や顧客のニーズを個別に的確に捉え、個に特化した細分化されたマーケティング活動を実現しなければならない状況にある。これらは、マーケティング実務における“マス・マーケティングからマイクロ・マーケティングへ”の変化により確認できる。マイクロ・マーケティングを高度化するためには、“個”に関する顕在化したものばかりでなく、潜在的なものも含めた詳細な情報が必要不可欠である。

二つ目のマーケティングの環境変化に対応し、それらの活動を高度化するためには、前述した消費者の行動の“結果データ”から、そこに内在する潜在的な高次情報をも抽出し、消費者の理解を深めることが求められる。これを実現するためには、統計的モデリングの技法を用いて、様々なマーケティング現象を“上手

さとう ただひこ
筑波大学 大学院ビジネス科学研究科
〒112-0012 文京区大塚 3-29-1

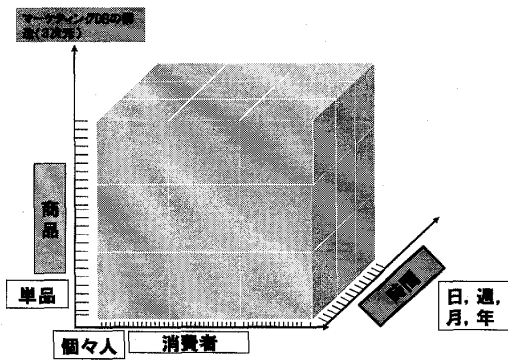


図1 結果データのデータベース構造

に”表現し、それに基づき個に関する詳細な情報を得ることが必要である。この際活用される統計モデルの技法が、後述のベイジアンモデリングである。

マーケティング分野において、これまでなされている研究のほとんどは、時間方向の構造を考慮せず、消費者のクロスセクショナルな行動の解明のみに焦点を当てなされている。さらに、当該分野では、“消費者は異質”という仮定の下での研究は数多くなされているが、一方、“時間的に異質”の仮定の下での研究は非常に少ない。しかし、現場やそこから得られるデータをつぶさに観察すると“時間的異質性”も存在することは明らかであり、この視点でのマーケティング研究は今後重要になっていくものと考えられる。そのため、本稿では“動的”な結果データの活用に関心をもち、解析事例を紹介し、この方向性でのマーケティング研究の今後の展望を示す。

2節では、ベイジアンモデリングを簡単に紹介する。3節には、結果データの動的活用事例を示す。4節はまとめと今後の展望である。

2. ベイジアンモデリング

本節では、マーケティング分野で活用の進む、ベイジアンモデリングの概要を示す。なお、以降では y_n , x_n , φ は、観測データ、パラメータ、超パラメータ（パラメータのパラメータ）をそれぞれ示すこととする。

従来の統計モデリングでは少ないパラメータで表現されるモデルが良いモデルとされてきた。いわゆる“けちの原理”である。どのようなデータであったとしても平均と分散のみで規定される正規分布から得られたと仮定するようなものである。

一方、パラメータ数を増やせば統計モデルの記述能力は向上するが、汎化能力と呼ばれる将来のデータの予測能力が低下する。この問題への対策として、パラ

メータ x_n についても統計モデル $p(x_n|\varphi)$ を想定するのがベイズモデルである。この $p(x_n|\varphi)$ をベイズ統計では事前分布と呼ぶ。この事前分布の導入により、ベイズの定理

$$p(x_n, \varphi|y_n) = \frac{p(y_n|x_n)p(x_n, \varphi)}{p(y_n)} \propto \frac{p(y_n|x_n)}{\text{尤度}} \frac{p(x_n|\varphi)}{x_n \text{ の事前分布}} \frac{p(\varphi)}{\varphi \text{ の事前分布}} \quad (1)$$

を用いることで、想定した事前分布 $p(x_n, \varphi)$ がどのように修正されるのか、つまりパラメータに関する不確実性がデータによりどの程度修正されたのかを観察するのである。この $p(x_n, \varphi|y_n)$ を事後分布と呼ぶ。ここで、データ y_n の発生確率 $p(y_n)$ は x_n, φ によらない数値になるので、事後分布 $p(x_n, \varphi|y_n)$ は(1)式右辺2行目に比例する。この仕組みにより、多数のパラメータも安定して推定できるようになり、結果として高い予測能力とデータ記述能力を同時に持つ総合的な統計モデルを構成できる。この一連のモデル化の行為を、通常、ベイジアンモデリングと呼ぶ。なお、(1)式を用いる際に、 n を時点と考えると時変パラメータを、個人と考えると個人ごとのパラメータを推定できることになる。本稿で紹介する話題は、時系列解析におけるベイズモデルである状態空間モデルを適用したものである。なお、本稿では紹介を行っていないが階層ベイズモデルのマーケティングへの適用に関しては文献[4][7]が参考になる。

3. 分析事例

3.1 状態空間モデル

3.1節では、本稿の事例で活用している時系列解析におけるベイズモデルである一般状態空間モデルを簡単に紹介する。

(一般状態空間モデル)

時刻 n の観測された時系列ベクトルを y_n 、未知の状態ベクトルを x_n とすると、以下に示す2つの式により一般状態空間モデルが定義される。

$$[\text{システムモデル}] \quad x_n \sim q(x_n|x_{n-1}), \quad (2)$$

$$[\text{観測モデル}] \quad y_n \sim r(y_n|x_n). \quad (3)$$

ただし、 $q(x_n|x_{n-1})$ は x_{n-1} を与えた場合の x_n の条件付き分布、 $r(y_n|x_n)$ は x_n を与えた場合の y_n の条件付き分布を示す。 $q(x_n|x_{n-1})$, $r(y_n|x_n)$ は(1)式の x_n の事前分布、尤度にそれぞれ対応する。

ここで導入した一般状態空間モデルは線形・ガウス

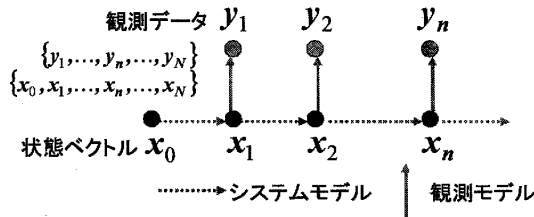


図2 線形・ガウス型状態空間モデルの鎖状グラフ

型の時系列や非線形時系列、離散時系列などの様々な時系列データを統一的な枠組みで取り扱うことを可能にするモデルである。図2は、(2)および(3)式で定義される一般状態空間モデルのイメージ理解のために、線形・ガウス型の状態空間モデルを例に鎖状グラフを示したものである。

図2に示す線形・ガウス型の状態空間モデルの状態推定は、Kalman フィルタと呼ばれるアルゴリズムで行われる。それ以外のより一般的な状態空間モデルの状態推定は、非ガウス型フィルタや粒子フィルタといったアルゴリズムを用いて推定を行う。この周辺は文献[3]を参照されたい。

3.2 動的売上反応モデルと参照価格

本小節では一つ目の分析事例を示す。以降に示す解析の詳細は文献[5]を参照のこと。ここで用いる結果データはPOSデータであり、モデル化は線形・ガウス型の状態空間モデルの枠組みでなされている。

マーケティング分野において、通常よくなされる解析は、売上反応分析と呼ばれるものである。これは、売上に対して値引きなどの販売促進活動がどの程度効果的かを探るもので、通常、売上と値引きなどの販売促進施策の状況を示す変数間の関係を静的な回帰モデルで表現することで、その効果を推定するのである。この構造であれば通常POSデータに含まれる“観測変数”のみを用いて簡便な評価ができる。しかし、このアプローチでは、消費者行動理論における参照価格の理論や消費者反応の動的変化を取り込めておらず、そのため、個を意識したマーケティング活動高度化に必要とされる、消費者の深い理解は実現できない。

参照価格とは、消費者個々に内在する商品に対する値頃感を示すものとして定義され、統計的な意味においては潜在変数となる。消費者行動論によれば、消費者は価格そのものというよりは、参照価格と実際に提示された価格の差によって購買意思決定を行うということが繰り返し述べられてきている。また、販売状況をすべて同一の状況に設定したとしても、時期が異なる

れば売上に大きなばらつきが生じる。この状況を説明するためには、販売促進活動の効果が動的に変化すると仮定する必要がある。本事例では、参照価格と消費者反応の動的変化をモデル化し、売上反応モデルに取り込むことで、知見の抽出を行っている。

本分析事例での課題は、参照価格の構造と反応係数の時間変化をどのようにモデル化するかである。まず、参照価格に関しては(4)式のように定式化する。

$$RP_n = (1-a)AP_{n-1} + aRP_{n-1} \quad (4)$$

(4)式中、 RP_n , AP_n , a は n 時点の参照価格、売価、参照価格のキャリー・オーバーを示すパラメータをそれぞれ示す。 a は最尤法で推定される。この設定の下で、点数PI (来店客千人当たり販売点数) y_n の規定構造を下記(5)式のように定式化する。これが状態空間モデルにおける観測モデルになる。

$$\log(y_n) = t_n + p_n + (\alpha_n^1 + \beta_n^1 E_n) Z_n^1 + (\alpha_n^2 + \beta_n^2 E_n) Z_n^2 + \varepsilon_n, \varepsilon_n \sim N(0, \sigma^2) \quad (5)$$

なお、 E_n は第 n 時点でのエンド陳列実施の有無を示し、価格変数 Z_n^1 , Z_n^2 は(6)式のように定義する。

(ロス変数)

$$Z_n^1 = \begin{cases} \frac{AP_n - RP_n}{100}, & AP_n > RP_n \text{ の時} \\ 0, & \text{その他} \end{cases}$$

(ゲイン変数)

$$Z_n^2 = \begin{cases} \frac{RP_n - AP_n}{100}, & RP_n > AP_n \text{ の時} \\ 0, & \text{その他} \end{cases}$$

さらに、 t_n , p_n はトレンド、週周期成分を示し、 α_n^1 , β_n^1 , α_n^2 , β_n^2 は説明変数の影響度を示す時変係数である。ここで、 $x_n = (t_n, p_n, p_{n-1}, \alpha_n^1, \beta_n^1, \alpha_n^2, \beta_n^2)^t$ とすると、時変係数を達成するモデルは(7)式で与えられ、状態空間モデルにおけるシステムモデルになる (平滑化事前分布)。

$$x_n = F \cdot x_{n-1} + G \cdot v_n \quad (7)$$

(7)式中、 F , G は既知の行列であり、 v_n は時変係数を達成するために導入された確率項で、システムノイズと呼ばれる (詳細は文献[5]を参照のこと)。

図3は、推定された店舗レベル参照価格と売価の系列を示している。参照価格の推移はモデルにより推定されたもので、観測されているものではない ($RP_n = 0.95RP_{n-1} + 0.05AP_{n-1}$)。

通常マーケティング分野で採用されている固定係数のモデルを採用した場合には、売上の時間経過とともに売上が落ち込むような構造や値引きに対する消費者

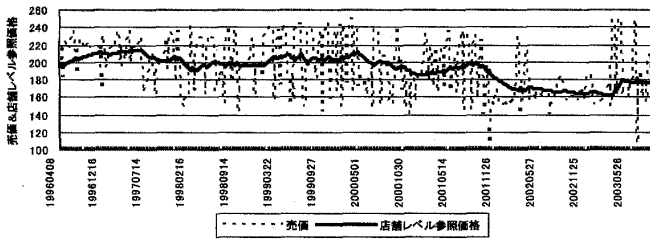


図3 店舗レベル参照価格 (推定結果)

反応の動的変化は的確に捉えることができない。また、通常のアプローチでは、説明変数の構造をパラメータ推定と同時に行うのは大きな困難がある。紹介した動的モデルを用いればこれらの課題に対応できる。その意味で、本分析事例のアプローチは有用である。

3.3 動的離散選択モデルと家庭内在庫量

本小節では二つ目の分析事例を示す。以降に示す解析の詳細は文献[6]を参照のこと。ここで用いる結果データはID付POSデータであり、モデル化は一般状態空間モデルの枠組みでなされている。

図1に示した構造のデータベースには、消費者が購買した(サービスを受けた)時のデータのみが蓄積される。そのため、これらデータは、消費者数は多いが消費者ごとのサンプル数(時点方向)は少数という特徴を持っていた。さらに、消費者ごとのサンプル数は異なっている。そのため、マーケティング分野における離散選択モデルを用いた研究では、基本的に動的構造は考慮されてきていない。しかし、時系列方向にもデータの蓄積が進んだ現在、この状況は大きく変化している。

本紹介事例は、消費者個々の来店とカテゴリ購買生起の2変数離散時系列データのモデル化を行い、個人ごとにそれら行動の動的メカニズムの解明を試みたものである。

第 n 日に消費者 i が来店するかどうかを $y_{i,n}^1$ (来店; $y_{i,n}^1=1$, 非来店; $y_{i,n}^1=0$)、あるカテゴリを購買するかどうかを $y_{i,n}^2$ (購買; $y_{i,n}^2=1$, 非購買; $y_{i,n}^2=0$)とし、それぞれベルヌーイ過程に従うと仮定し、次のようにモデル化する。

$$P(y_{i,n}^1, y_{i,n}^2) = P(y_{i,n}^2 | y_{i,n}^1) P(y_{i,n}^1) \\ = (p_{i,n}^2)^{y_{i,n}^2} (1 - p_{i,n}^2)^{1 - y_{i,n}^2} (p_{i,n}^1)^{y_{i,n}^1} (1 - p_{i,n}^1)^{1 - y_{i,n}^1} \quad (8)$$

(8)式中、 $p_{i,n}^1, p_{i,n}^2$ は消費者 i の第 n 日における来店確率およびカテゴリ購買生起確率をそれぞれ示し、それぞれロジットモデルで定式化する。これが一般状態空間モデルの観測モデルになる。システムモデルは、時変パラメータを並べたベクトルを状態ベクト

ル x_n とし、(7)式と同等の形式で与える。ただし、観測モデルが線形・ガウス型ではないため、モデルの推定には粒子フィルタを用いる。以降では、カテゴリ購買生起モデルに焦点を当て議論する。

消費者のカテゴリ購買生起モデルは、効用(潜在変数)の概念を導入し、その効用の確定項($u_{i,n}^2$)により(9)式のように選択確率を定式化できる。

$$p_{i,n}^2 = \frac{y_{i,n}^2 \exp(u_{i,n}^2)}{1 + \exp(u_{i,n}^2)} \quad (9)$$

(9)式で示すカテゴリ購買生起モデルでは、消費者の商品選択行動ではなく、その商品が含まれる例えば醤油、インスタントコーヒーといったカテゴリの購買生起行動のモデル化を狙っている。そのため、商品個々の販売促進活動とは別の、消費者個々に内在する変数をも考慮しなければならない。消費者行動理論の文脈でなされた研究では、当該行動には家庭内在庫量が強く影響することが示されている。ただし、家庭内在庫量も参照価格同様潜在変数であり、直接は観測できない。

本分析事例での課題は、潜在変数である家庭内在庫量と反応係数の時間変化をどのようにモデル化するかである。反応係数のモデル化に関しては、3.2節の事例と原理的には同様であるので割愛する。家庭内在庫量に関しては次のようにモデル化する。

第 n 日の個人 i に対する家庭内在庫量、 INV_n^i は下記のように定義される。

$$INV_n^i \equiv INV_{n-1}^i + PurQty_{n-1}^i - Consump_{n-1}^i \quad (10)$$

(10)式中、 $PurQty_n^i, Consump_n^i$ は第 n 日の個人 i の対象カテゴリの購買量および消費量をそれぞれ示す。さらに、家庭内在庫量に影響を及ぼす家庭内消費量についても(11)式のようにモデル化する。(11)式中、 C^i は個人ごとの当該カテゴリの平均消費量(ホールドアウトサンプルで算定)および b^i, fp^i は最尤法で推定されるパラメータである。

$$Consump_n^i = INV_n^i \left(\frac{(C^i)^{b^i}}{(C^i)^{b^i} + (INV_n^i)^{fp^i}} \right) \quad (11)$$

(10)および(11)式によって、家庭内在庫量と消費量の相互依存性を表現できる。

図4は、個人ごとに推定される家庭内在庫量と消費量の動的変動を示している。消費量が家庭内在庫量の変動に伴って変動する様子が分かる。

上述の推定結果のマーケティング上の意味については、紙幅の都合上割愛するが、本事例で紹介した個に関する情報抽出は、本アプローチではじめて実現でき

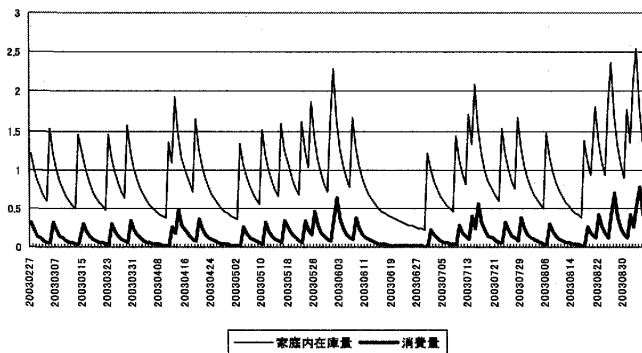


図4 家庭内在在庫量，消費量推定結果

るものである。当該モデルでは、入力データとして“基本的に来店したかどうか”と“カテゴリー購買生起があったかどうか”のみを用い、その背後に内在している“個に”関する情報を抽出している。これらの情報は、前述のマイクロ・マーケティングにおいて有効に活用される可能性をもつ。その意味で、本分析事例のアプローチは有用である。

4. まとめと今後の展望

本稿では、マーケティング分野における環境変化に適應し、マイクロ・マーケティング活動を高度化するために、そこで必要となると考えられる“個に”関する情報抽出に有効であるベイジアンモデリングを紹介し、その分析事例を示した。以降には、当該分野における動的知識発見実現のための今後の展望を述べる。

来店や購買といった最終結果にいたるプロセスに関する情報（嗜好，市場反応，習慣，頭の中の情報処理過程等）は直接測定（観察）されていない。それらのほとんどはそもそも動的に変動する潜在的な変数であり、しかも本質的に個人ごとに必ずしも正確に測定できるものでもない。しかしながら、最近のマーケティング研究では、その周辺の解明が興味の対象となっている。

マーケティング分野では、これら消費者の潜在的特性およびそれらが消費者の行動に与える影響について消費者行動，企業行動などの観点から膨大な演繹的研究がなされ、マーケティング理論や消費者行動理論として体系づけられてきた。最近になってようやく、それら理論に基づいて計算機上でシミュレーションを実施することで、複雑な現象を解明しようとする試みがなされるようになってきている。このシミュレーションモデルは理論から導かれたモデルであるので、演繹モデルである。一方、本稿で紹介した分析事例では、

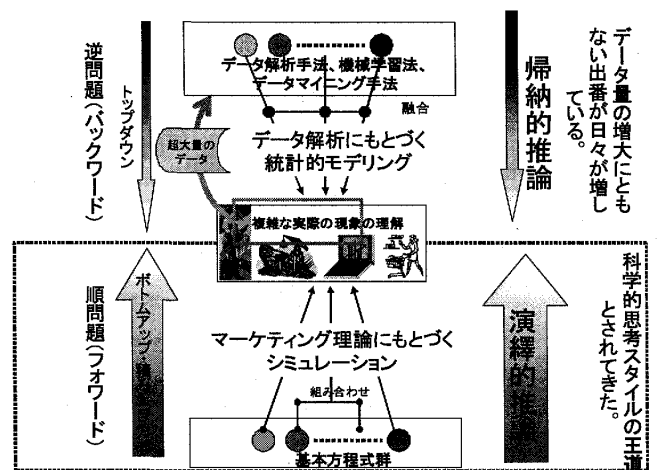


図5 演繹と帰納の融合

データからそこに内在するメカニズムを推定する帰納的な形式で推論を行った。演繹アプローチにしても帰納アプローチにしても、どちらか一方だけでは複雑なマーケティングにおける動的現象の解明には不十分である。もしマーケティング理論や消費者行動理論に基づく演繹的なシミュレーションモデルと本稿の分析事例で示したようなデータにもとづく帰納的な統計モデルが融合できれば、結果データの更なる高度活用が期待される。

前述の課題を解決するには、理論に基づく潜在変数の時間発展を与えるシミュレーションモデルを、一般状態空間モデルのシステムモデルとしてプラグインすれば、マーケティング理論のモデル内への明示的取り込みが可能になる。これらの考え方は、ベイジアンモデルの一種で“データ同化”と呼ばれ、昨今研究が進んでいる統計手法である。マーケティング分野で、データ同化手法を用いた研究は現在までなされていない。しかし、この手法の持つ柔軟性は極めて高く、その意味で当該分野でも本手法は有効であると考えられる。本手法を用いたマーケティング研究の進展が望まれる。なお、データ同化に関して紙幅の都合上本稿ではその詳細を紹介していないが、詳しくは文献[1]を参照されたい。

マーケティング分野において、演繹推論の立場をとる研究者は非常に多い。理論に基づき仮説を立て、データによりその仮説を検証するという立場である。一方、統計科学の諸手法を用いてデータから帰納的にマーケティング現象を捉えモデル化しようという立場の研究者もいる。すなわち、現象を支配している関係式，経験則を観測データから推定していくといった立場である。現在まで多くのマーケティング研究は、どちらか

一方の立場でなされているものがほとんどである。しかし今後は、これら立場を融合した新たなパラダイムでのマーケティング研究が必要となるであろう。図5には、マーケティングにおいて演繹と帰納の融合を達成するための概念を模式的に示している。そこで最も重要な役割を担うのがデータ同化を含むベイジアンモデリングである。それが、演繹と帰納の融合、すなわち、理論、観測データ、経験的知識などの異種多様な情報の統合と個別化を実現する鍵となる技術になる[2]。

謝辞 本稿で示した内容はすべて、統計数理研究所の樋口知之先生との共同研究の成果によっている。この場を借りて御礼を申し上げる次第である。

参考文献

[1] 樋口知之, 「データ同化: その概念と計算アルゴリズム

ム」, 『統計数理』, Vol. 53, No. 2, (2004), 211-229.

[2] 樋口知之監修・著, 『統計数理は隠された未来をあらわにする』, 東京電機大学出版局, (2007).

[3] 北川源四郎著, 『時系列解析入門』, 岩波書店, (2005).

[4] P.E. Rossi, G.M. Allenby and R. McCulloch, Bayesian Statistics and Marketing, John Wiley and Sons, (2005).

[5] 佐藤忠彦, 樋口知之, 「動学的売上反応モデルによるPOSデータの解析」, 『マーケティング・サイエンス』, Vol. 15, No. 1・2 (2008), 1-26.

[6] 佐藤忠彦, 樋口知之, 「動的個人モデルによる購買生起行動の解析」, 『マーケティング・サイエンス』, Vol. 16, No. 1・2 (2009), 49-73.

[7] 照井伸彦著, 『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, (2008).