

# POSデータのロジットモデルによる 解析とその評価

—インスタント・コーヒーのPOSデータを用いて—

川口 康之

## 1. はじめに

80年代半ば以降、レジスターのPOS化進展に伴い消費者の購買行動の詳細な分析が可能となった。これは消費者別の購買行動のデータを必要とするロジット分析において、特に重要な意味を持つ。

本研究では、POSデータを用いてロジット・モデルに基づくインスタントコーヒーのブランド選択行動について解析を行ない、内挿、外挿予測によりモデルの適合度の評価を行う。具体的には、平成5年のデータに基づき内挿を行い、平成6年のブランドシェアを外挿により予測することを目的とする。

## 2. ロジットモデル

商品  $j$  ( $j = 1, \dots, J$ ) に対する効用は確定的部分  $A_j$  と確率的部分  $\varepsilon_j$  で以下のような関数で表されると仮定する。

$$U_j = \sum \alpha_i x_{ij} + \varepsilon_j = A_j + \varepsilon_j$$

ただし

$\alpha_i$  : パラメータ

$x_{ij}$  : マーケティング変数 (説明変数)

$\varepsilon_j$  : 独立で同一極値分布タイプ1に従う

とする。すなわち  $\varepsilon_j$  の分布関数は

$$F(U) = \Pr[\varepsilon_j \leq U] = \exp[-\exp(U - A_j)]$$

かわぐち やすゆき 東京工業大学経営工学科

〒152 東京都目黒区大岡山 2-12-1

与えられる。この分布に互いに独立に従う  $J$  個の  $U_j$  の標本をとる。その最大値は次の分布に従う。

$$F_{\max}(U) = \Pr \left[ \max_j U_j \leq U \right] = \exp \left( -e^{-U} \sum_{j=1}^J e^{A_j} \right)$$

よって、商品  $k$  を選ぶ確率  $P_k$  は

$$\begin{aligned} P_k &= \int_{-\infty}^{\infty} \int_{-\infty}^{U_k} f(U_k) f_{\max}(U) dU dU_k \\ &= \frac{e^{A_k}}{\sum_{j=1}^J e^{A_j}} \end{aligned}$$

で表される。ただし  $f$  は  $F$  の密度関数である。これをロジットモデルと呼ぶ。

モデルの長所としては、どんな  $A_j$  をとっても  $e^{A_j}$  は正なので  $0 \leq P_j \leq 1$  が常に成り立つ。逆に短所としては、IIA (Independence from Irrelevant Alternatives) 問題、消費者層の相違、効用関数の仮定の妥当性などが挙げられる。

## 3. 本研究でのモデル

本研究では、上記の問題点 (解析区間、クラス分け、等) を考慮に入れたモデルに基づいて解析を行った：

1. 日ごとにデータの解析を行い、選択確率を求める。
2. 用いるデータはすべての商品に共通な価格掛率、エンド (店頭販売)、チラシの有無とする。なお、解析の自由度を上げるために、各商品を説明する新たな2つの説明変数  $y_{ij}$  を導入する。
3. 消費者を4又は5クラスに分ける。

よって、クラス  $k$  の商品  $j$  の  $t$  期の効用関数を

$$U_{jtk} = \sum \beta_{ik} y_{ij} + \sum \alpha_{ik} x_{ijt} + \varepsilon_{jtk}$$

で表し、これに基づいて選択確率  $P_{jtk}$  を推定し、

$$\sum_k \sum_t P_{jtk} N_{tk}$$

をその年の商品  $j$  の購入量とする。ただし、 $P_{jtk}$  はクラス  $k$  の  $t$  期の商品  $j$  の選択確率、 $N_{tk}$  はクラス  $k$  の  $t$  期の総購入量、 $\alpha_{ik}$  と  $\beta_{ik}$  はパラメータである。

さらに説明変数  $y_{ij}$  は、選択行動がマルコフ連鎖により説明できると仮定し、選択行動の推移確率行列  $\mathbf{P}$  の定常分布  $\pi = (\pi_i)$  により特徴づける。

$$y_{1j} = \pi_j, \quad y_{2j} = \pi_j(1 - P_{jj})$$

但し  $\sum \pi_j = 1$  で

$$\pi = \pi \mathbf{P}$$

また、 $y_{1j}$  は  $t$  期に商品  $j$  を購入する確率、 $y_{2j}$  は  $t$  期に商品  $j$  を購入し、 $t+1$  期に商品  $j$  以外を購入する確率である。

次にクラス分けの手法について述べる。はじめに1個人=1クラスと仮定して、各個人ごとにパラメータを推定する。但し、推定パラメータの不安定性から、個人  $k$  が1年間に選んだ商品群を個人  $k$  の想起集合と仮定する。この推定パラメータ・ベクトル  $\alpha_k, \beta_k$  はそれぞれ個人  $k$  のセールスプロモーション (以下 SP と呼ぶ) に対する反応、商品群に対する反応を表している。すなわち

$$\gamma_k = (\alpha_k, \beta_k)$$

の位置でブランドロイヤルティ、SP 反応を含む個人  $k$  の特性を測ることができる。よって、個人間のベクトル  $\gamma_k$  の距離でクラスター分析を行い、クラス数を個人数より減少させ、小数データに基づく推定量の不安定性を除去した。具体的には最長距離法または最短距離法でクラスター分析を行い、観測値と理論値の確率2乗誤差の総和、または確率2乗誤差の加重平均から最適な分析基準並びにクラス数を決定した。

なお推定されるパラメータは次のように解釈できる。

|               |                |
|---------------|----------------|
| $\beta_{1k}$  | 商品が消費者を保持する力   |
| $\beta_{2k}$  | 商品が消費者を手放す度合い  |
| $\alpha_{1k}$ | 負ならば、値引きに対する反応 |
| $\alpha_{2k}$ | エンドに対する反応      |
| $\alpha_{3k}$ | チラシに対する反応      |

## 4. 推定結果

### 4.1 内挿の評価

解析の結果、最長距離法でクラス分けを行う方が当てはまりの良いことが分かった。但し評価は選択確率の理論値と観測値との2乗誤差の総和をとって行なう (表1)。また、表2はクラスに入る人数のサイズを示しているが、この表よりクラス分けは4クラスより進んでいない。よって5クラス分割した時点でクラス分けは終了する。

表 1:

|    |              |             |
|----|--------------|-------------|
| 内挿 | 3クラス分け (最短法) | 0.077915685 |
| 外挿 | 3クラス分け (最長法) | 0.048923373 |
| 内挿 | 5クラス分け (最長法) | 0.051575133 |

表 2:

|      | 3クラス | 4クラス | 5クラス |
|------|------|------|------|
| クラス1 | 666  | 666  | 666  |
| クラス2 | 102  | 77   | 77   |
| クラス3 | 26   | 26   | 23   |
| クラス4 |      | 25   | 25   |
| クラス5 |      |      | 3    |

最長距離法で5クラスに分割し、さらにクラスごとに説明変数  $y_{jk}$  を導入し推定購入量を見ていく。各クラスの購入サイズを表3に示す。

表 3:

| クラス  | 1    | 2   | 3   | 4   | 5  |
|------|------|-----|-----|-----|----|
| 選択回数 | 4803 | 412 | 169 | 151 | 15 |

以上の結果より、各クラスの意味付けをパラメータ・ベクトルの値から考察する。

表4は各クラスのパラメータ値を示している。表4よりSPが効いているのは、クラス1のみである。 $\beta_k$  の大小関係と大きさから、クラス1はブランドロイヤルティの小さい、SPに敏感な消費者クラスと考えられる。

逆に、クラス2とクラス5はブランドロイヤルティの大きなクラスと考えられる。

クラス3は、SPがかなり負に働いていることから、ロイヤルティの小さい人で、個人の選択行動を持っているクラス、クラス4はコーヒーの愛好家ではないかと考えられる。

表 4: パラメータの値

|       | $j \rightarrow j$ | $j \rightarrow k$ | 掛け率           | エンド           | チラシ           |
|-------|-------------------|-------------------|---------------|---------------|---------------|
|       | $\beta_{1k}$      | $\beta_{2k}$      | $\alpha_{1k}$ | $\alpha_{2k}$ | $\alpha_{3k}$ |
| クラス 1 | 4.999             | 8.052             | -1.260        | 0.000         | 0.000         |
| クラス 2 | 0.657             | -2.026            | 0.601         | -0.239        | -0.239        |
| クラス 3 | 12.118            | 12.507            | 26.235        | -2.844        | 0.000         |
| クラス 4 | -10.553           | 0.000             | 1.806         | -0.233        | 0.000         |
| クラス 5 | 23.435            | 5.349             | 4.963         | -1.790        | -1.538        |

よってクラス 1 以外は過去 (平成 5 年) の経験分布をもって、外挿 (平成 6 年の各ブランドのシェアの推定) を行う方が良い値が出てくると予測できる。

理由としては、効用関数

$$U_{jtk} = \sum \beta_{ik} y_{ij} + \sum \alpha_{ik} x_{ijt} + \varepsilon_{jtk}$$

の第 2 項は  $t$  に依存しているが、第 1 項は依存していない。実際 SP を行うことにより商品の効用が下がるとは考えられないので、第 2 項が負の効用を示すクラスは、第 1 項のみで効用が決まると考えられる。すなわち、そのクラスの選択行動は過去の経験分布に従って予想した方が良い結果が期待できる。以下推定したパラメータを用いて外挿 (平成 6 年) について考えていく。

## 4.2 外挿の評価

データは平成 6 年 7 月までのものを用いた。各クラスの人数のサイズを表 5 に示す。クラス 0 は平成 6 年に新たに加わった消費者なので解析の対象から外す。

表 5:

| クラス | 0 | 1   | 2  | 3  | 4  | 5 | 合計  |
|-----|---|-----|----|----|----|---|-----|
| 人数  | 2 | 618 | 64 | 22 | 22 | 3 | 731 |

前節で考察した事柄について、次の 2 つのケースで外挿を行った：

ケース 1：すべてのパラメータをモデルで推定

ケース 2：クラス 1 のみモデル、残りのクラスは経験分布で推定

前節で予想した通り、ケース 2 のほうが良い当てはまりをみせる (図 1)。よってこれを最終結果とする (表 6)。なお

不一致係数：0.058627

であった。

さらにクラス毎に見ていくと、最大サイズのクラス 1 はモデル、経験分布共にほとんど差が無いのに対し、

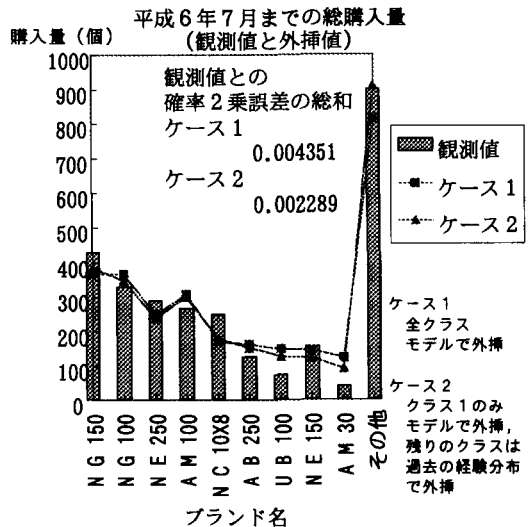


図 1:

表 6:

| ブランド               | 総購入量  |       | 確率 (%) |     |
|--------------------|-------|-------|--------|-----|
|                    | 外挿    | 観測値   | 外挿     | 観測値 |
| ネスカフェゴールド 150g     | 382.7 | 427   | 13     | 15  |
| ネスカフェゴールド 100g     | 347.6 | 329   | 12     | 12  |
| ネスカフェエクセラ 250g     | 234.5 | 288   | 8      | 10  |
| AGF マキシム 100g      | 298.0 | 264   | 10     | 9   |
| ネスカフェアブチーノ         | 177.3 | 246   | 6      | 9   |
| AGF プレنديー 250g    | 152.9 | 124   | 5      | 4   |
| UCC ザブレンド 144/100g | 126.2 | 73    | 4      | 3   |
| ネスカフェエクセラ 150g     | 122.2 | 154   | 4      | 5   |
| AGF マキシム 30g       | 91.5  | 40    | 3      | 1   |
| その他                | 911.1 | 899   | 32     | 32  |
| 合計                 | 2,844 | 2,844 | 100    | 100 |

残りのクラスは明らかに経験分布のほうが良い当てはまりをみせる (表 7)。

特に人数の 87%、選択回数の 87% を占めるクラス 1 のモデルによる推定が成功していることから、本研究の有効性が実証されたと考える。

## 5. 結論と今後の課題

本研究ではロジットモデルを中心に消費者の商品選択行動を考察した。結果として、消費者間の特異性を除くことにより、解析精度をかなり向上できることが明らかになった。

クラス 1 の消費者はロジットモデルにより選択確率を推定することができた。クラス・サイズにかんがみても、このクラスをマーケティング戦略の対象として SP を構築することが重要となる。また今回は、各個人の説明変数のパラメータによってクラスター分析を

表 7: 観測値と理論値との 2 乗誤差の総和

|       | 経験分布    | モデル     |
|-------|---------|---------|
| クラス 1 | 0.00240 | 0.00360 |
| クラス 2 | 0.01440 | 0.11863 |
| クラス 3 | 0.02330 | 0.12190 |
| クラス 4 | 0.03460 | 0.43444 |
| クラス 5 | 0.04222 | 0.83625 |

行ったが、このとき、想起集合を仮定した事がうまく特異層を排除出来た原因になったと思われる。

また、モデルでは、エンド、チラシに対する反応を示す消費者層を検出することは出来なかったが、同じデータで分析を行っている他のグループも同じ結果を出している。考えられる点は 2 つある。1 つは、本研究で用いたデータはスーパーマーケットのデータである。これは消費者クラスの特徴から考えると、値引きに敏感な層を中心に分析している可能性が高い。したがって、もしコンビニエンスストアのデータを利用すればさらに面白い結果が得られると予想される。また最初からエンド有り、チラシ有り、を 1 つの商品 (1 つの商品に対し最大 4 種の商品とする) として解析を行えば、これらの SP の有効性ははっきり検出できるものと考えられる。もう 1 点は、このデータの問題で、その他の商品のサイズが大きすぎることにある。その為、本来 SP に反応しているクラスが存在していても、

規模の効果がそれが消えてしまった可能性がある。対象商品を増やせば、外挿の解析の精度は落ちるであろうが、それでも上位 2 ないし 3 クラスに絞ったより実戦的な具体的戦略の構築には役立つであろう。

なお、拡張モデルは外から説明変数を導入したが、2 段階の最小化問題を解くことにより、ブランド力を示す値を変数としてそのつど推定すれば、より正確なモデル、クラスを構築する事が出来るはずである。これらの問題を今後の課題として、本論文を終わる。

## 参考文献

- [1] 小川孔輔:法政大学産業情報センター:POSとマーケティング戦略, 有斐閣 (1993).
- [2] 片平秀貴:新しい消費者分析-LOGMAP の理論と応用, 東京大学出版会 (1991).
- [3] 中西正雄:消費者行動分析のニュー・フロンティア, 誠文堂新光社 (1984).
- [4] 阿部周造:消費者行動, 千倉書房 (1978).
- [5] 小川孔輔, 片平秀貴, 杉田善弘, 西尾チヅル:オペレーションズ・リサーチ Vol.39, No.4. 特集:マーケティングの展開, 日本オペレーションズ・リサーチ学会 (1994).
- [6] J. Eliashberg, G.L. Lilien: HANDBOOKS IN OPERATIONS RESEARCH AND MANAGEMENT SCIENCE, Vol.5, MARKETING, chapter 6, NORTH-HOLLAND (1993).

## ● 書評者募集 ●

下記の書籍の書評をしていただける方を募集しています。ご希望の方は編集委員の栗田治 (慶応義塾大学) までお申し出ください。

|                           |         |
|---------------------------|---------|
| 江村潤朗ほか「技術革新と情報化人材育成」      | 日科技連出版社 |
| 合原一幸ほか「カオス応用システム」         | 朝倉書店    |
| 坂和正敏ほか「ソフト最適化」            | 朝倉書店    |
| 岩澤孝雄ほか「経営革新と産業ネットワーク」     | 日科技連出版社 |
| ポール・ルック編「ソフトウェア信頼性ハンドブック」 | 日科技連出版社 |
| ブライアン・ジョイナー「第 4 世代の品質管理」  | 日科技連出版社 |
| 大村平「ビジネス数学のはなし」           | 日科技連出版社 |