

Online-Offline チャン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

猪狩 良介, 星野 崇宏

1. はじめに

近年、店頭に加え EC サイトなどの Online チャンネルも重要な販売チャネルとなっており、Online と Offline の双方を自社製品の販売チャネルとして展開している企業は多い。このような環境下において、Online から Offline へと顧客の行動を促すマーケティング施策や、Online 情報接触行動の Offline 購買行動への影響を踏まえた戦略が重要であるとされている。

また、近年では製造と販売を一貫して実施するプライベート・ブランド（以下 PB）が多く見られるようになってきている。特に PB では、会員登録やアプリのダウンロードなどにより、リアル店舗での購買行動と Online 上の購買行動の双方をシングルソースで捉えやすい。これにより、どの消費者はリアル店舗で商品を購入し、どの消費者は EC サイト上で商品を購入するか、またどの消費者はリアル店舗・EC の双方で購入するかを把握することが可能になる。また、Online と Offline における行動の相互作用も考えられる。たとえば、初めはリアル店舗で商品を購入していた消費者が、学習することで EC サイトで購入するようになる、といった行動の変化もありえる。あるいは、EC サイトで購入していた人のブランドロイヤルティが高まることで、リアル店舗への来店につながる、といった行動も考えられる。これらを把握することで、Online からリアル店舗への誘導や、リアル店舗から Online サイト訪問を促すといった OtoO 戦略を効率的に行うことができる。さらに Online と Offline では消費者の購買行動が異なると考えられる。たとえば、店頭ではシャ

ンプーは自社ブランドで購入し、トリートメントは競合ブランドを購入していた消費者が、EC ではまとめて買いをすることでトリートメントも自社ブランドで購入するようになることも考えられる。このように、異なるチャンネルへ顧客を促すことで、その消費者の自社製品全体の支出額の増加やロイヤルティ向上の可能性も十分に考えられる。

本研究では、店頭と Online の二つの購買チャンネルにおける購買間隔と購買時の購入金額に焦点を当てる。消費者は店頭と Online の二つのチャンネルのうちどちらかを選択するため、店頭と EC サイトは競合チャンネルとなる。競合チャンネルがある場合には、それぞれのチャンネルについて単独の購買有無をモデリングするだけでなく、購買タイミングの競合を考える必要がある。たとえば、ある購買機会で EC 購買が発生した場合、競合イベントである店頭購買は観測されない。複数のイベントにおいて、一方のイベントが観測されると他方のイベントが観測できない状況を競合イベントと呼び、この状況下で単独の解析を行うと、期待した結果が得られないことがある。また、どのチャンネルで購入したかを考慮しない単一イベントモデルでは、EC 購買と店頭購買の関係性は考慮することができない。

そこで本研究では、Online と Offline 双方のチャンネルにおける購買行動をシングルソースで捉えた PB のデータを用い、チャンネルを競合イベントと捉えた繰り返しのある購買間隔モデルと、購買した際の購買金額の同時モデルを提案する。店舗と Online における購買タイミングに加えて、実務的には消費者の購買金額の予測も重要であることから、間隔と金額を同時にモデリングする。競合イベントを捉えるモデルとしては、競合リスクモデル [1, 2] を、購買金額を捉えるモデルには回帰モデルを用いる。競合リスクモデルを用いることで、実際には観測されない潜在的なイベントも考慮した分析ができる。たとえば、店頭購買が観測されている消費者においても EC 購買に対する潜在的な購買発生率が存在し、店頭のみを利用している消費者に

いがり りょうすけ

慶應義塾大学大学院経済学研究科
〒108-8345 東京都港区三田 2-15-45

igariryosuke@yahoo.co.jp

ほしの たかひろ

慶應義塾大学経済学部・大学院経済学研究科
〒108-8345 東京都港区三田 2-15-45

hoshino@econ.keio.ac.jp

受付 16.1.22 採択 16.7.11

対して、もしマーケティング変数を適切に操作すれば EC 購買が起りやすくなる、といったことを捉えることができる点が特徴である。さらに、消費者の異質性を捉えるために同時モデルに潜在クラスモデルを導入し、クラスへの所属を予測する階層ベイズモデルへと拡張する。PB に焦点を当てる理由として、その商品の購買が自社の店舗や EC サイトのみでしか行えないクローズドな世界であること、会員登録などにより複数のチャネルにおける購買履歴をシングルソースで取得しやすいことなどが挙げられる。

2. 先行研究

2.1 マルチチャネルにおける購買行動モデル

マルチチャネルにおける購買モデルについては、売上などの集計データを用いた研究（たとえば文献 [3, 4] など）と、個人レベルの購買データを用いた研究の大きく二つがある。本研究は個人レベルのデータを用いた研究に該当する。個人レベルのデータを用いた研究としては、文献 [5] は購買履歴データを用いて、購買有無と購買量の同時モデルに加え、購買を行う場合はカタログと Online のどちらを選択するかというモデルを提案している。彼らのモデルは、購買生起と購買量・購買チャネルを捉える同時モデルを提案しているという点では本研究に近いが、購買間隔を扱った研究ではない。また、文献 [6] は、個人の特性や供給側の要因（どのようにコンタクトするか）といった情報を用いて、消費者が利用するチャネル数をモデル化している。彼らのモデルは、どのような消費者が複数のチャネルで購買を行うかを捉えるうえでは有効であるが、購買間隔を扱った研究ではない。また、文献 [7] はパネルデータを用い、複数チャネルの初回採用行動を多変数のハザードモデルで捉える方法を提案している。彼らのモデルは購買チャネルの初回採用行動に焦点を当てており、繰り返し購買間隔を捉える本研究とは目的が異なる。

2.2 購買間隔と同時モデリング

マーケティング分野において、生存時間解析の方法を用いた購買間隔 (Interpurchase Timing) の研究はこれまで多くされている。生存時間解析を用いることで、購買間隔に影響を与えるマーケティング変数の効果を測定することが可能になる。これにより、消費者の来店時期や購入時期にクーポンや DM を送付するといった、適切なタイミングでマーケティング活動を行うことが可能になる。文献 [8, 9] は比例ハザードモデルを用いて購買間隔を捉える研究を行っている。また、

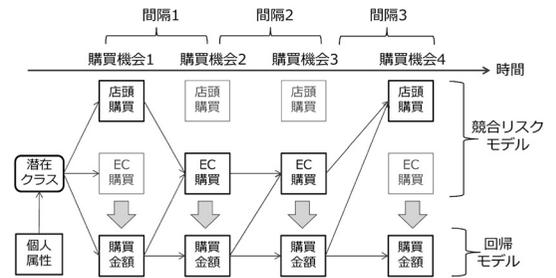


図 1 モデル全体像

文献 [10] はさまざまなタイプのハザードモデルを用いて購買間隔モデルの比較を行っている。一方で、文献 [11] は計数過程モデルを用いた繰り返し購買の間隔モデルを提案している。

さらに、購買間隔のモデルとブランド選択などの同時モデルも提案されている。同時モデルを用いることで、購買間隔と選択行動や購買数量、金額などの関係性の把握や、また間隔と反応双方の行動に共通する要素を抽出することができる。文献 [12] は、ハザードモデルと多項ロジットモデルを組み合わせ、購買間隔とブランド選択の双方を捉えた同時モデルを提案した。さらに文献 [13] は、Web サイトの訪問間隔とコンバージョン行動を捉えるために、ハザードモデルと 2 項選択モデルの同時モデルを提案している。

3. モデル

本研究では、繰り返しのある購買データに対し、競合リスクモデルを用いた Online (EC サイト) と Offline (店頭) における購買間隔と回帰モデルを用いた購買金額の同時モデルを構築する。さらに、消費者の異質性を説明するために潜在クラスモデルを導入し、階層ベイズモデルで表現する。競合リスクモデルとしては文献 [1, 2] などがあり、競合リスクモデルを含む同時モデルとしては、文献 [14] は最尤法による推定法を、文献 [15] はベイズ統計学による推定法をそれぞれ提案している。また、競合リスクモデルに潜在クラスモデルを導入したモデルとしては文献 [16] が存在するが、競合リスクモデルを含む同時モデルに潜在クラスを導入した事例はなく、さらにそれをベイズ推定した研究もない。本研究は、Online と Offline における繰り返しのある購買間隔を競合リスクモデルを用いて表現している点で新しい。また、提案モデルは競合リスクを含む同時モデルに潜在クラスを導入している点、さらにそれをベイズ推定している点で、統計モデルの観点からも新規性がある。図 1 にモデルの全体像を示す。

3.1 潜在クラスの導入

まずは、消費者の異質性を潜在クラスモデルによって表現する。潜在クラスモデルにより、マーケティングにおけるセグメンテーションの考えをモデルに組み込むことができる。消費者 $i (i = 1, \dots, n)$ が潜在クラス $m (m = 1, \dots, M)$ に所属する際のインディケータを $z_i = m$ と表記する。各潜在クラス m は異なるパラメータをもち、セグメントによる異質性をモデル上で表現可能である。

さらにマーケティングでは、消費者がどの潜在クラスに所属するかを予測することが重要である。そこで、デモグラフィック属性を用いて潜在クラスへの所属を説明する階層モデルを用いる [17, 18]。これにより、デモグラフィック属性のみわかっている人がどのクラスに所属するかを予測することができ、潜在顧客の行動予測に活用できる。

個人 i がクラス m に所属する確率を多項ロジットモデルを用いて式 (1) のように表現する。

$$p(z_i = m) = \frac{\exp(\mathbf{d}_i^T \boldsymbol{\gamma}_m)}{\sum_{l=1}^M \exp(\mathbf{d}_i^T \boldsymbol{\gamma}_l)} \quad (1)$$

ここで、 \mathbf{d}_i はデモグラフィック属性、 $\boldsymbol{\gamma}_m$ はその係数ベクトルである。ロジットモデルの識別性のため、 $\boldsymbol{\gamma}_1 = \mathbf{0}$ という制約をおく。

3.2 競合リスクモデルを用いた購買間隔モデル

Online と Offline チャネルにおける消費者の購買間隔を競合リスクモデルにより捉える。競合リスクモデルでは、複数のイベント $k (k = 1, \dots, K)$ があるなかで、観測される購買間隔 T は $T = \min(T_1, \dots, T_K)$ となり、最も購買間隔が短いイベントが観測されることから、これらはイベントの打ち切りと同じように考えることができる。打ち切りを考慮せずに解析を行うと、推定値にはバイアスが生じることが知られており、これを無視して競合イベントを単一イベントとして解析をすると、情報のある打ち切りと同じ状況となり、単純に EC や店舗の行動だけ見ると大きなバイアスが生じる [19, 20]。なお、本研究では競合イベントは EC サイトと店頭の一つのため、 $K = 2$ となる。

潜在クラス m に所属する消費者 i の $j (j = 1, \dots, J_i)$ 回目の繰り返し購買におけるイベント (チャネル) k のハザード関数は式 (2) となる。

$$\begin{aligned} \lambda_{mk}(t) &= \lim_{\Delta t \rightarrow 0} \frac{p(t \leq T_{ij} < t + \Delta t, \delta_{ij} = k | T_{ij} \geq t, z_i = m)}{\Delta t} \\ &= \lambda_{mk0}(t) \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}_{mk}) \end{aligned} \quad (2)$$

ここで、 T_{ij} は前回購買時からの経過日数を表す。なお、観測開始から最初のイベント (購買) をデータの始まりとし、初回から 2 回目までの購買間隔を T_{i1} とする。このように、今回はパネルデータを利用して繰り返し購買を捉えるため、生存時間解析において一般的な左右の打ち切りの状況はない。しかし、競合リスクモデルを用いているため、一方のイベントの購買が発生するともう一方のイベントについては、見かけ上右打ち切りの状況となる。また、 $\lambda_{mk0}(t)$ はイベント k とクラス m によって異なるベースラインハザード関数であり、 \mathbf{x}_{ij} は時間共変量、 $\boldsymbol{\beta}_{mk}$ はイベント k とクラス m によって異なるパラメータである。

ベースラインハザード関数 $\lambda_{mk0}(t)$ には、指数分布やワイブル分布などのパラメトリックモデルや、特定の分布を仮定しないノンパラメトリックモデルなどが利用可能である。今回はパラメータの値によりさまざまな形状を表現することが可能で、かつ比例ハザードモデルでも表現可能なワイブル分布をベースラインハザード関数に仮定する (式 (3))。

$$\lambda_{mk0}(t) = \alpha_{mk} t^{\alpha_{mk} - 1} \quad (3)$$

ここで、 α_{mk} はイベント k とクラス m によって異なる形状パラメータであり、 $\alpha_{mk} > 0$ である。

また、生存関数 $S(t)$ は式 (4) となる。

$$S(t) = \exp\left\{-\int_0^t \left(\sum_{k=1}^K \lambda_{mk}(u)\right) du\right\} \quad (4)$$

ここで確率密度関数は、式 (5) となる [1, 2]。

$$\begin{aligned} p(T_{ij} | z_i = m) &= \prod_{k=1}^K \lambda_{mk}(T_{ij})^{\delta_{ijk}} S(T_{ij}) \\ &= \prod_{k=1}^K \left[\alpha_{mk} T_{ij}^{\alpha_{mk} - 1} \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}_{mk}) \right]^{\delta_{ijk}} \\ &\quad \times \exp\left\{-\exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}_{mk}) T_{ij}^{\alpha_{mk}}\right\} \end{aligned} \quad (5)$$

ここで、 δ_{ijk} は継続時間 T_{ij} が観測される際のイベントが k のときに 1 をとり、それ以外のときは 0 をとるインディケータである。

3.3 購買金額のモデル

次に、購買が発生した時点において、消費者がどれだけ購入するかという購買金額のモデルを線形回帰モデルを用いて式 (6) のように表現する。

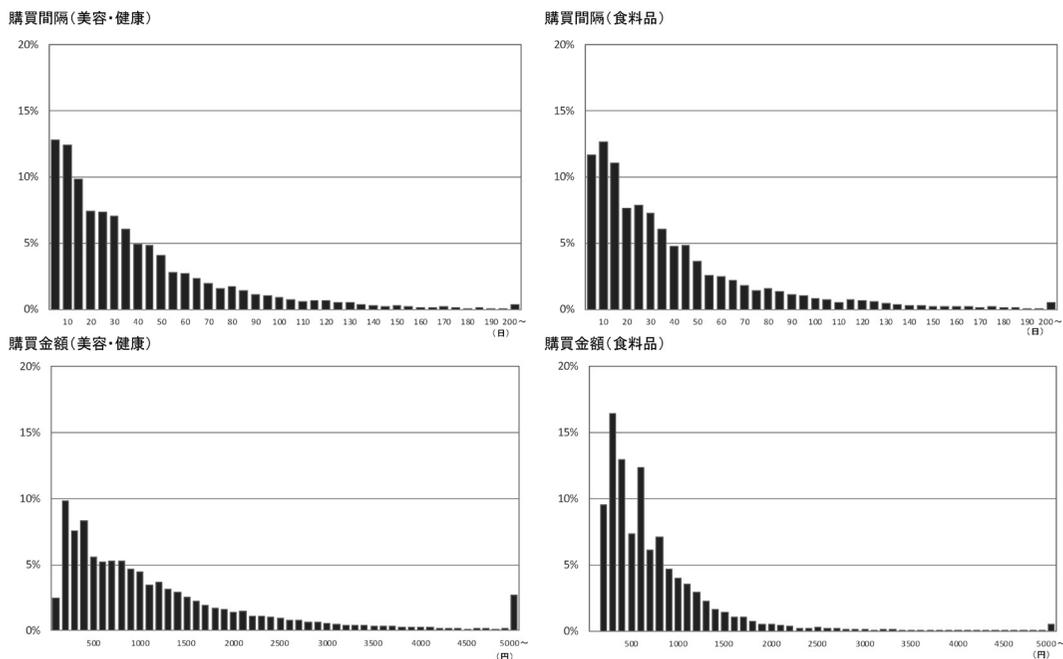


図2 購買間隔および購買金額の分布

$$Y_{ij} = \mathbf{w}_{ij}^{\top} \boldsymbol{\phi}_{mk} + \epsilon_{ijk}, \epsilon_{ijk} \sim N(0, \sigma_{mk}^2) \quad (6)$$

ここで、 \mathbf{w}_{ij} は時間共変量であり、 $\boldsymbol{\phi}_{mk}$ と σ_{mk}^2 はクラス m とイベント k によって異なるパラメータである。また、共変量 \mathbf{w}_{ij} に前回購買からの経過時間 T_{ij} を含める。これにより、クラスによる購買金額の傾向の違いと、購買タイミングと購買金額の関係性、マーケティング変数の効果の違いなどを把握することができる。

なお、確率密度関数は式 (7) となる。

$$p(Y_{ij}|T_{ij}, z_i = m) \propto \prod_{k=1}^K \left[\sigma_{mk}^{-1} \exp\left\{-\frac{1}{2\sigma_{mk}^2} (Y_{ij} - \mathbf{w}_{ij}^{\top} \boldsymbol{\phi}_{mk})^2\right\} \right]^{\delta_{ijk}} \quad (7)$$

3.4 尤度関数と推定法

以上より、提案モデル全体の尤度関数は式 (8) となる。

$$L = \prod_{i=1}^n \left\{ \sum_{m=1}^M p(z_i = m) \times \prod_{j=1}^{J_i} p(Y_{ij}|T_{ij}, z_i = m) p(T_{ij}|z_i = m) \right\} \quad (8)$$

モデルはベイズ統計学の枠組みで表現し、パラメータの推定にはマルコフ連鎖モンテカルロ法 (以後 MCMC 法) を用いる。MCMC 法の詳細は付録に示す。

4. 実証分析

4.1 データ概要と共変量

実証分析として、ある PB メーカーの Online と Offline における購買履歴をシングルソースで捉えたデータを用いる。この PB メーカーは、全国に多くの店舗を構えており、さらに EC サイトも運営している。また会員登録により両チャネルにおける購買行動を把握可能である。

分析には、2013 年 5 月から 2014 年 6 月までの約 1 年間のデータを用い、消費財の「美容・健康カテゴリ」と「食料品カテゴリ」の 2 カテゴリを用いる。「美容・健康カテゴリ」はシャンプーなどのヘアケア製品や化粧品などが含まれる。「食料品カテゴリ」は、加工食品や調味料などが含まれる。なお、観測期間内で 5 回以上購買があった消費者を対象とした。解析サンプルは「美容・健康カテゴリ」についてパラメータ推定用の in-sample として $n = 4508$ 、予測精度の検証用の out-sample として $n = 2417$ を、「食料品カテゴリ」の in-sample として $n = 3750$ 、out-sample として $n = 2724$ のサンプルを抽出した。また、購買間隔は日単位のデータを、購買金額は円単位のデータを用いる。なお、購買金額は解析の際には対数変換した値を用いる。購買間隔および購買金額のヒストグラムを図 2 に示す。

次に、解析に用いる変数を定義する (表 1 を参照)。

表 1 変数一覧

モデル	変数名	概要
時間共変量 (ハザード)	会員ステージ	なし (基準カテゴリ), シルバー, ゴールド以上
	途中の訪問/来店有無	前回から今回までの間の EC サイト訪問と店頭来店
	ln (前回購入金額)	前回購入時の金額 (円) の対数
	前回購入チャンネル	前回購入にて同一のチャンネルか否か
時間共変量 (回帰)	会員ステージ	なし (基準), シルバー, ゴールド以上
	途中の訪問/来店有無	前回から今回までの間の EC サイト訪問と店頭来店
	ln (前回購入金額)	前回購入時の金額 (円) の対数
	前回購入チャンネル	前回購入にて同一のチャンネルか否か
	購買間隔	今回の経過時間
個人属性 (多項ロジット)	性別	男性 (基準カテゴリ), 女性
	年代	20 代以下, 30 代, 40 代, 50 代, 60 代以上 (基準カテゴリ)

表 2 デモグラフィック属性

		美容・健康	食料品
性別	女性	88.9%	85.6%
	男性	11.1%	14.4%
年代	20 代以下	19.5%	17.8%
	30 代	43.5%	41.0%
	40 代	26.6%	29.0%
	50 代	8.3%	9.4%
	60 代以上	2.2%	2.8%

競合リスクモデルの時間共変量 α としては、会員のステータス (ポイント蓄積状況) を示す「会員ステージ」の『シルバー』『ゴールド以上』と、「途中訪問/来店有無」の『EC サイト訪問』『店頭来店』, 「前回購入金額の対数」と、「前回購入チャンネル」を用いる。さらに、回帰モデルの共変量 ω では競合リスクモデルで用いた変数に加え、「購買間隔 (日)」を用いる。「会員ステージ」は、購買時に蓄積されているポイント (マイル) を意味しており、これにより、どの程度のポイント保持者が購買しやすいかを把握することができる。なお、「会員ステージ」は同じ人でも購買機会によって変化する変数である。さらに、「訪問/来店有無」は当該カテゴリ以外の商品を購入した際の履歴を基に作成しており、これにより EC サイトに途中で来店した人は店頭購買をしやすくなり、また店頭に来店した人は EC サイトで購買しやすくなるといったチャンネル間の関係性を見ることができる。さらに「前回購入金額」は、前回の購買金額が多い人ほど間隔が長くなりやすいなどの影響を見ることができる。さらに、「前回購入チャンネル」をダミー変数として用いることで、チャンネル選択の依存関係を見ることが可能になる。また、デモグラフィック属性として「性別」および「年代」を用いる。データの要約統計量を表 2, 3 に示す。

4.2 解析結果

パラメータ推定の MCMC は全体で 10,000 回行い、うち最初の 5,000 回を burn-in とし、最後の 5,000 回のサンプリング値を用いて事後平均や標準偏差などを算出した。なお、文献 [21] の方法を用いてすべてのパラメータの収束を確認した。また、各商品カテゴリにおいてクラス数 1~5 まで推定し、クラス数を決定する。なお、「食料品カテゴリ」ではクラス数 4・5 の推定ができなかった。理由として、クラス数 3 (表 5) の結果を見てもわかるが、あるクラスの所属割合が非常に小さいため、クラスへの所属を説明する多項ロジットモデルのパラメータ γ の推定が非常に不安定になるためと推察される。実際、クラス数 4・5 についてはロジットモデルのパラメータ γ が推定できなかった。以後、「食料品カテゴリ」ではクラス数 1~3 の結果を検証する。

4.2.1 潜在クラス数の決定とモデル評価

まず、クラス数を決定する。今回は予測力を重視し、out-sample を用いた交差妥当性の検証によりクラス数を選択する¹。推定結果の事後平均値を用いて購買間隔と購買金額を予測し²、MSE を算出した (表 4 参照)。結果から、「美容・健康カテゴリ」および「食料品カテゴリ」どちらもクラス数 3 が最良のモデルという結果になった。以後、どちらの製品カテゴリについてもクラス数 3 について考察する。また、提案モデルのパフォーマンスを示すために、対立モデルとして、EC サイトと店頭のどちらのチャンネルで購買したかを考慮しない単一イベントモデルを推定した。提案モデルの潜在ク

¹ 実際は in-sample を用いた周辺尤度および DIC などによるモデル評価も実施しているが、予測力と解釈のしやすさという観点から out-sample による結果を優先した。

² 競合リスクモデルの予測では、イベント別の予測値を発生させ、予測値の最小のイベントを実際の予測値とする。

表3 要約統計量

		美容・健康			食料品		
		全体	EC 購買	店頭購買	全体	EC 購買	店頭購買
購買数	購買機会数	36175	1749	34426	29110	1097	28013
	購買割合	100.0%	4.8%	95.2%	100.0%	3.8%	96.2%
目的変数	購買間隔 (日)	35.8	41.0	35.5	35.7	41.9	35.5
	購買金額 (円)	1246.7	2899.2	1162.8	713.7	2035.0	662.0
共変量	会員ステージ (なし)	57.8%	46.3%	58.4%	55.3%	46.8%	55.6%
	会員ステージ (シルバー)	25.4%	28.4%	25.3%	25.8%	33.7%	25.5%
	会員ステージ (ゴールド以上)	16.8%	25.3%	16.3%	18.9%	19.5%	18.9%
	途中訪問率 (EC)	5.0%	26.2%	4.0%	5.4%	24.6%	4.7%
	途中来店率 (店頭)	48.3%	27.4%	49.4%	53.8%	24.2%	54.9%
	前回同一チャンネル利用率	95.0%	48.2%	97.4%	96.6%	55.3%	98.2%

表4 クラス数の決定とモデル評価 (MSE)

		美容・健康	食料品
提案モデル	クラス数 1	3.660	3.548
	クラス数 2	3.627	3.497
	クラス数 3	3.622	3.495
	クラス数 4	3.667	—
	クラス数 5	3.671	—
対立モデル		3.646	4.490

ラスを仮定したモデルは、対立モデルよりもよい結果となった (表4 参照)。

4.3 結果解釈

4.3.1 クラスの内訳とデモグラフィック属性の影響

各クラスの特徴とクラスへの所属を説明する多項ロジットモデルのパラメータについて表5に示す。「美容・健康カテゴリ」において、クラスの所属割合はクラス1が63.9%と最大となっており、クラス2は32.3%、クラス3は3.8%と最小のクラスとなっている。一方でEC購買と店頭購買の割合を見ると、クラス1とクラス2は店頭における購買が約98%である。しかし、クラス3はEC購買割合が77.2%と、EC購入をするセグメントを抽出することができた。クラス1・2は店頭購買が大多数を占めるセグメント、クラス3はEC購買を中心としているセグメントと解釈できる。また、「食料品カテゴリ」では、クラス1が約78%と最大のクラスとなり、次いでクラス3が18.7%、クラス2が3.4%のクラスとなった。また、クラス1・3は店頭購買割合が高い店頭メインのクラス、クラス2はEC購買割合が高いECメインのクラスと解釈できる。

次に、潜在クラスへの所属を説明するパラメータを解釈する。なお、「*」は95%信用区間で有意となった係数である。「美容・健康カテゴリ」では、有意な変数は「定数項」と「性別 (女性)」のみとなった。クラス2は「性別 (女性)」が有意なプラスの結果となり、また

クラス3では有意なマイナスとなっている。それ以外にはクラスへの所属に関与していないと解釈できる。一方で「食料品カテゴリ」では「定数項」と「性別」に加えて「年代 (20代以下)」が有意な結果になっており、Online購買が多いクラス2は有意なマイナス、店頭購買が多いクラス3は有意なプラスとなっている。

4.3.2 競合リスクモデルの解釈

続いて、競合リスクモデルのパラメータを表6に示す。両製品カテゴリと各クラスに共通して見られたのは、「会員ステージ」がプラスに作用するという点である。これは、ポイントプログラムによって消費者の購買タイミングを早めることができることを示唆している。逆に、「途中の訪問・来店」はOnline購買が多いクラス (「美容・健康カテゴリ」のクラス3と「食料品カテゴリ」のクラス2) を中心に基本的にマイナスに作用することがわかる。一方で、「美容・健康カテゴリ」のクラス3と「食料品カテゴリ」のクラス2については途中の店舗来店が店頭購買発生にプラスに有意に影響する結果となった。また、「前回の購入金額」はEC購買については「美容・健康カテゴリ」の店頭での購買が多いクラス1のみが有意なマイナスの結果となったが、店頭購買については「美容・健康カテゴリ」ではEC購買が多いクラス3および、「食料品カテゴリ」のEC購買が多いクラス2、店頭購買が多いクラス3が有意な結果となった。前回購買時の金額は、多いほど次回来店までの時間は長くなると思われるが、店舗での購買においては必ずしもそうではないことがわかる。また、「前回購入チャンネル」の影響は、どちらのカテゴリでもEC購買間隔についてはEC購買が多いクラス (「美容・健康カテゴリ」のクラス3と「食料品カテゴリ」のクラス2) についてはプラスに働き、店頭購買が多いクラスについては逆に働くことがわかる。同様に、店頭購買間隔については逆の傾向が見られた。

表5 各クラスの特徴と所属パラメータ

クラス概要	クラスの特徴	美容・健康			食料品		
		クラス 1	クラス 2	クラス 3	クラス 1	クラス 2	クラス 3
	所属割合	63.9%	32.3%	3.8%	77.9%	3.4%	18.7%
	EC 購買割合	2.3%	2.0%	77.2%	1.6%	82.4%	1.0%
	店頭購買割合	97.7%	98.0%	22.8%	98.4%	17.6%	99.0%
デモグラ	定数項	—	-1.814*	-3.044*	—	-2.990*	-3.936*
フィック	性別 (女性)	—	0.302*	-0.169*	—	-0.452*	0.070*
係数	年代 (20 代以下)	—	1.348	-0.187	—	-1.007*	3.300*
	年代 (30 代)	—	0.904	0.609	—	0.247	2.383
	年代 (40 代)	—	0.544	0.324	—	0.261	2.016
	年代 (50 代)	—	0.494	-0.289	—	0.717	2.012

表6 競合リスクモデルのパラメータ

		美容・健康			食料品		
		クラス 1	クラス 2	クラス 3	クラス 1	クラス 2	クラス 3
EC	定数項	-4.024*	-3.285*	-4.951*	-4.381*	-5.830*	-3.140*
	会員ステージ (シルバー)	0.303*	0.720*	0.504*	0.357*	0.494*	1.252*
	会員ステージ (ゴールド以上)	0.662*	0.864*	0.913*	0.327	0.658*	1.002*
	訪問有無 (EC)	-0.343*	-0.374*	-0.509*	-0.585*	-0.626*	-0.201
	来店有無 (店頭)	-0.791*	-1.745*	-0.828*	-0.948*	-0.717*	-2.152*
	ln (前回購入金額)	-0.116*	-0.053	-0.009	-0.053	0.011	-0.130
	前回同一チャンネル	-5.057*	-5.446*	0.137	-5.724*	0.099	-5.891*
	形状パラメータ α	1.157	1.141	1.221	1.169	1.427	1.261
店頭	定数項	-5.117*	-3.706*	-1.879*	-4.899*	-3.165*	-2.830*
	会員ステージ (シルバー)	0.357*	0.403*	0.522*	0.347*	0.936*	0.491*
	会員ステージ (ゴールド以上)	0.432*	0.732*	0.935*	0.413*	1.505*	0.650*
	訪問有無 (EC)	-0.458*	-0.570*	-1.050*	-0.478*	-0.664*	-0.738*
	来店有無 (店頭)	-0.881*	-1.425*	0.290*	-0.971*	0.847*	-1.547*
	ln (前回購入金額)	0.000	0.018	-0.291*	-0.019	-0.385*	-0.069*
	前回同一チャンネル	0.611*	0.376*	-0.787*	0.641*	-0.952*	0.591*
	形状パラメータ α	1.278	1.184	0.753	1.277	1.053	1.148

このことから、ECでの購買が多いクラスについては、必ずしも前回購買時のチャンネルが同じかどうかは影響していないことがわかる。一方で、店頭での購買が多いクラスについては、前回購買時のチャンネル (=店頭) がそのまま続くことがわかる。

続いて図3にベースラインハザード関数(式(3))をプロットする³。ベースラインハザードは「美容・健康カテゴリ」の店頭チャンネルにおけるクラス3を除き、右上がりの形状となっている。これは、形状パラメータ $\alpha_{mk} > 1$ という結果が得られているためである。さらに、「美容・健康カテゴリ」のECチャンネルでは、EC購買の多いクラス3がベースラインハザードが全体的に高く、また店頭チャンネルではクラス3が初めは高く、すぐに低くなる右下がりの曲線となっていることがわ

かる。「食料品カテゴリ」では、ECチャンネルではクラス2のベースラインハザードが全体的に高く、「美容・健康カテゴリ」と同じ結果となっている。しかし、店頭チャンネルではクラス1・2のベースラインハザードは上昇しているのに対して、クラス2はほぼ一定となっている。また、「食料品カテゴリ」においては、同じ店頭購買が多いクラス1・クラス3を比較しても、ベースラインハザードの形状に大きな違いが見られており、クラス間の明確な違いとなっている。

4.3.3 購買金額の解釈

次に、購買金額モデルの結果を表7に示す。購買金額では、購買間隔とは異なり、「会員ステージ」は必ずしもプラスに働くわけではないことがわかる。特にEC購買ではそれが顕著であり、ポイントが蓄積すると購買タイミングには影響するが、購買金額へはEC購買ではそこまで影響しないことを意味している。一

³ 図3にはスケールパラメータ(本モデルにおいては定数項が該当)が含まれていないため、解釈には注意が必要である。

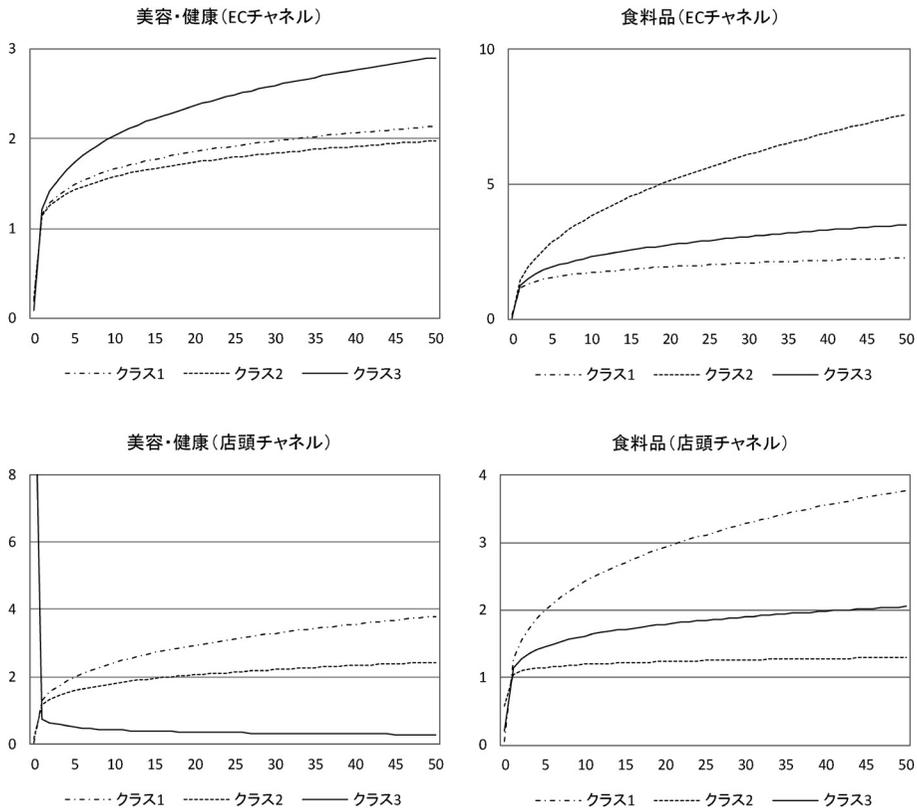


図3 ベースラインハザード関数

表7 購買金額モデルのパラメータ

	美容・健康			食料品		
	クラス 1	クラス 2	クラス 3	クラス 1	クラス 2	クラス 3
EC						
定数項	5.760*	6.370*	4.203*	3.745*	3.845*	5.829*
会員ステージ (シルバー)	-0.127	0.205	-0.056	0.061	-0.001	-0.118
会員ステージ (ゴールド以上)	-0.290*	0.278	0.092	-0.214	-0.016	-0.096
訪問有無 (EC)	0.038	-0.159	-0.243*	-0.049	-0.290*	0.013
来店有無 (店頭)	0.046	-0.181	-0.179	-0.023	-0.102	-0.159
ln (前回購入金額)	0.215*	0.082	0.426*	0.473*	0.479*	0.242
前回同一チャネル	-0.031	-0.261	0.049	-0.251	-0.093	-0.083
購買間隔	0.001	0.001	0.002	0.001	0.003*	0.004
誤差分散 σ^2	1.105	1.177	1.057	0.676	0.745	1.102
店頭						
定数項	5.008*	5.633*	5.338*	4.292*	6.026*	4.616*
会員ステージ (シルバー)	-0.040*	0.066*	0.356*	-0.012	0.149	0.034
会員ステージ (ゴールド以上)	-0.020	0.113*	0.077	0.017	-0.142	0.079*
訪問有無 (EC)	-0.158*	-0.009	-0.098	0.000	0.133	0.084*
来店有無 (店頭)	-0.105*	-0.111*	0.063	-0.102*	-0.199	-0.033
ln (前回購入金額)	0.248*	0.123*	0.164*	0.316*	0.038	0.202*
前回同一チャネル	0.106*	-0.027	0.155	0.055	0.019	0.167
購買間隔	0.001*	0.000	0.002	0.001*	0.003	0.000
誤差分散 σ^2	0.882	0.955	0.815	0.423	0.513	0.517

方、店頭購買では「美容・健康カテゴリ」ではクラス2が有意なプラスの結果となっている。また、「前回購

入金額」はEC・店頭の双方において、おおむねプラスに働くことがわかる。先ほどの購買タイミングでは

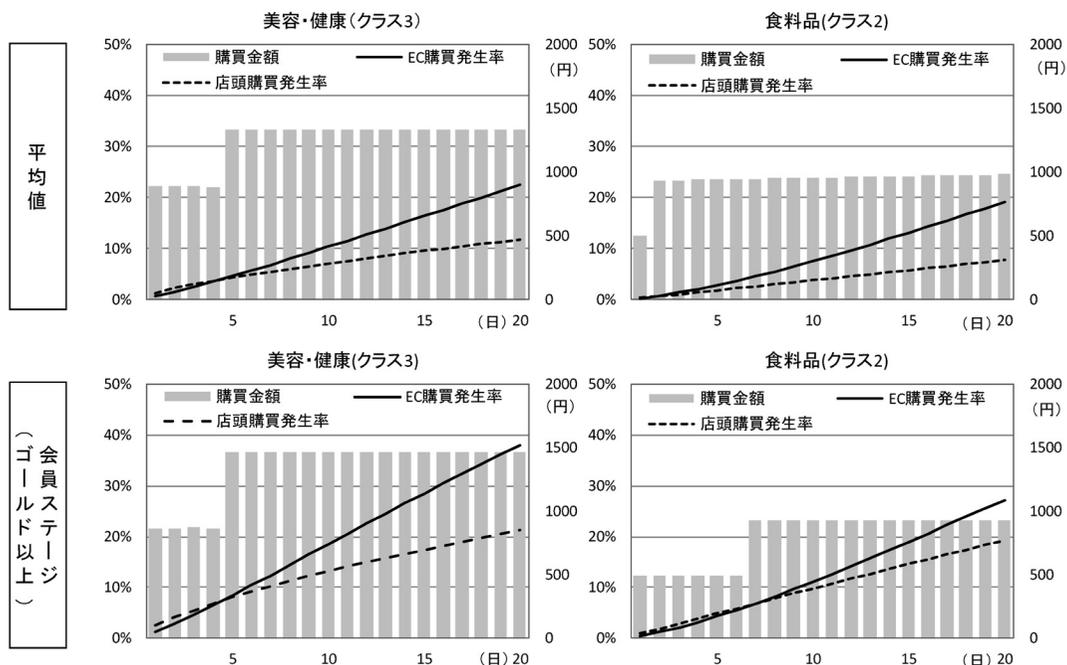


図4 イベント発生確率と購買金額の期待値

おおむねマイナスに働いていたが、金額については前回購入の状態が続くと読み取れる。また、「今回の購買間隔」は購買金額にはあまり影響を与えないという結果となった。

4.3.4 予測活用

今回の提案モデルの結果を用いた予測活用方法について示す。本研究では、競合リスクモデルを用いていることから、各競合イベントの起こりやすさを予測することができる。図4に、各カテゴリにおいてOnline中心のクラスについて、累積イベント発生確率と購買金額の予測値の期待値を示す。ここでは、累積イベント発生確率が高いほうのイベントが発生すると仮定し、購買金額は、イベント発生率の高いほうのチャネルの回帰式を用いた期待値である。上段は時間共変量がすべて平均値の場合、下段は「会員ステージ（ゴールド以上）」とし、それ以外を平均値とした場合のスコアである。「美容・健康カテゴリ」では、開始時点の確率は店頭購買のほうが高く、5日以降はEC購買のほうが高くなることわかる。それに連動して回帰式もECに変わることによって購買金額の期待値も上昇する。一方で右側の「食料品カテゴリ」では、平均値では発生確率は開始2日目からEC購買の方が高い状況が続く一方で、「会員ステージ（ゴールド以上）」では6日目までは店頭購買発生率が高い結果となっている。同様に、購買金額の期待値も7日以降は上昇している。

また、上段と下段で時間共変量の値が変化することで累積発生確率が変化していることがわかる。このほかにも、前回購買したチャネルや途中の訪問/来店などのスコアも計算することもできる。このように、競合リスクモデルを用いることで、競合イベントの発生確率の変化を予測することができ、また同時モデルを用いることで購買金額の予測も同時に行うことができる点が提案モデルの利点である。

5. まとめ

本研究では、OnlineとOfflineの双方のチャネルについての購買行動をシングルソースで捉えたプライベート・ブランドのデータを用い、ECサイトと店頭における消費者の繰り返し購買間隔と購買金額を捉える同時モデルを提案した。具体的には、OnlineとOfflineの購買間隔を捉える競合リスクモデルと、購買発生時における購買金額を捉える回帰モデルの同時モデルを提案し、消費者の異質性を説明するために、潜在クラスモデルを導入した。潜在クラスの結果から、EC購買が多いクラスと店頭購買が多いクラスを抽出することができた。また、実証分析より、競合リスクを考慮しないモデルと比較して提案モデルのほうがパフォーマンスが高いことを確認した。結果から、解析を行った二つの製品カテゴリそれぞれについてクラス数3が最もよい結果となり、いずれのカテゴリからも店頭購

買が中心の2クラスと、Online購買に特徴的な1クラスが推定された。さらに、推定結果を用いて共変量の値を変化させることで各イベントの発生確率と、それに基づく購買金額の期待値を算出することができることを示した。この結果を利用することで、CRMなどさまざまなことに応用が可能である。たとえば、個人の過去の履歴からハザード関数を算出し、購買確率が高まるタイミングに広告を配信する、といったマーケティング施策などが考えられる。また、会員ステージなどのポイントプログラムのチャンネルによる効果の違いを見ることで、購買チャンネルや購買金額に応じたポイント付与の重みを変えることなどが挙げられる。さらに、途中来店がプラスに作用するクラスに対して、特別にポイントを配布するといった方法も考えられる。

提案モデルはOnline-Offlineチャンネルにおける購買間隔のみならず、さまざまなテーマに応用可能である。たとえば、店舗を競合イベントとすれば、店舗間の競合関係を考慮した購買間隔モデルが可能になる。その場合、もしクーポンを配布すれば自社店舗に購入する可能性が高まる、といった議論が可能である。また、Webアクセスログデータを用いた研究(たとえば[18, 13])にも応用可能である。その場合、異なる検索エンジンの利用やECサイトの利用などを競合イベントとして設定すればよい。このように、競合リスクモデルを用いることで、さまざまなマーケティング課題に応用可能である。ただし、競合リスクモデルはイベントが少数に限定される寡占市場のような状況において利用すべきである。

今後の課題としてマーケティング変数の充実が挙げられる。今回はデータの制約上、広告接触やセールスプロモーションなどのマーケティング変数を盛り込むことができなかった。クーポンの配布やSNS接触、企業サイトのクチコミ情報などをシングルソースで取れているデータが存在すれば、モデルに組み込むことが可能になる。これは、このようなデータが入手できれば実施する予定である。

謝辞 本論文の改稿において2名の匿名の査読者から有益なコメントをいただきました。この場を借りて御礼申し上げます。また、本研究はJSPS科研費(JP15J03947, JP26285151)の助成を受けて実施したものです。

参考文献

- [1] J. D. Kalbfleisch and R. L. Prentice, "Competing Risks and Multistate Models," *The Statistical Analysis of Failure Time Data 2nd*, John Wiley & Sons, pp. 247-277, 2002.
- [2] 西川正子, "生存時間解析における競合リスクモデル," *計量生物学*, **29**, pp. 141-170, 2008.
- [3] E. Biyalogorsky and P. Naik, "Clicks and mortar: The effect of on-line activities on off-line sales," *Marketing Letters*, **14**, pp. 21-32, 2003.
- [4] 鶴見裕之, 増田純也, 中山厚穂, "商品に関するTwitter上のコミュニケーションと販売実績の関連性分析," *オペレーションズ・リサーチ:経営の科学*, **58**, pp. 436-441, 2013.
- [5] A. Ansari, C. F. Mela and S. A. Neslin, "Customer channel migration," *Journal of Marketing Research*, **45**, pp. 60-76, 2008.
- [6] V. Kumar and R. Venkatesan, "Who are the multi-channel shoppers and how do they perform?: Correlates of multichannel shopping behavior," *Journal of Interactive Marketing*, **19**, pp. 44-62, 2005.
- [7] R. Venkatesan, V. Kumar and N. Ravishanker, "Multichannel shopping: Causes and consequences," *Journal of Marketing*, **71**, pp. 114-132, 2007.
- [8] D. C. Jain and N. J. Vilcassim, "Investigating household purchase timing decisions: A conditional hazard function approach," *Marketing Science*, **10**, pp. 1-23, 1991.
- [9] K. Helsen and D. C. Schmittlein, "Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models," *Marketing Science*, **12**, pp. 395-414, 1993.
- [10] P. B. Seetharaman and P. K. Chintagunta, "The proportional hazard model for purchase timing: A comparison of alternative specifications," *Journal of Business & Economic Statistics*, **21**, pp. 368-382, 2003.
- [11] G. E. Bijwaard, P. H. Franses and R. Paap, "Modeling purchases as repeated events," *Journal of Business & Economic Statistics*, **24**, pp. 487-502, 2006.
- [12] P. K. Chintagunta and A. R. Prasad, "An empirical investigation of the 'Dynamic McFadden' model of purchase timing and brand choice: Implications for market structure," *Journal of Business & Economic Statistics*, **16**, pp. 2-12, 1998.
- [13] 星野崇宏, "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**, pp. 41-58, 2013.
- [14] R. M. Elashoff, G. Li and N. Li, "A joint model for longitudinal measurements and survival data in the presence of multiple failure types," *Biometrics*, **64**, pp. 762-771, 2008.
- [15] W. Hu, G. Li and N. Li, "A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data," *Statistics in Medicine*, **28**, pp. 1601-1619, 2009.
- [16] K. Patra and D. K. Dey, "A multivariate mixture of weibull distributions in reliability modeling," *Statistics & Probability Letters*, **45**, pp. 225-235, 1999.
- [17] S. Gupta and P. K. Chintagunta, "On using demographic variables to determine segment membership in logit mixture models," *Journal of Marketing Research*,

31, pp. 128–136, 1994.

- [18] 猪狩良介, 星野崇宏, “非集計 Web アクセスデータを用いたサイト普及モデル—多時点・複数サイトの階層バイズモデリング—,” マーケティング・サイエンス, **20**, pp. 43–67, 2012.
- [19] J. P. Klein and M. L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*, Springer Science & Business Media, 2003.
- [20] J. J. Gaynor, E. J. Feuer, C. C. Tan, D. H. Wu, C. R. Little, D. J. Straus, B. D. Clarkson and M. F. Brennan, “On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data,” *Journal of the American Statistical Association*, **88**, pp. 400–409, 1993.
- [21] J. Geweke, “Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments,” Oxford University Press, J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.), *Bayesian Statistics 4*, pp. 169–193, 1992.
- [22] 和合肇 (編著), 『バイズ計量経済分析—マルコフ連鎖モンテカルロ法とその応用』, 東洋経済新報社, 2005.
- [23] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*, Springer Science & Business Media, 2006.

付録. MCMC アルゴリズム

(1) 事前分布

事前分布として, $\log(\alpha_{mk}) \sim N(a_{0mk}, A_{0mk})$, $\beta_{mk} \sim MVN(\mathbf{b}_{0mk}, B_{0mk})$, $\phi_{mk} \sim MVN(\mathbf{r}_{0mk}, R_{0mk})$, $\sigma_{mk}^2 \sim IG(n_{0mk}/2, v_{0mk}/2)$, $\Gamma \sim MVN(\mathbf{g}_0, G_0)$ を設定する. なお, $\Gamma = (\gamma_2^T, \dots, \gamma_M^T)^T$ であり, 各クラスのパラメータを一つのベクトルにしたものである. また, 実質的に無情報事前分布となるように, それぞれのハイパーパラメータの値を設定した.

(2) 事後分布

α の発生

事後分布は, 以下となる.

$$p(\alpha_{mk}|\cdot) \propto \prod_{i; z_i=m}^n \prod_{j=1}^{J_i} [\alpha_{mk} T_{ij}^{\alpha_{mk}-1} \exp(\mathbf{x}_{ij}^T \beta_{mk})]^{\delta_{ijk}} \times \exp(-\exp(\mathbf{x}_{ij}^T \beta_{mk}) T_{ij}^{\alpha_{mk}}) \times \exp\left\{-\frac{1}{2} A_{0km}^{-1} (\log(\alpha_{mk}) - a_{0km})^2\right\}$$

α のサンプリングには, Random-Walk Metropolis-Hastings (以後 M-H 法) アルゴリズムを用いる. 新しい α の候補点は, $\log(\alpha_{mk}^{cand})|\cdot \sim N(\log(\alpha_{mk}^{s-1}), \tau^2)$ からサンプリングする. τ^2 は Random-Walk の分散であり, 適切な採択率となるように調節する. 新しい候補点を採択する確率は, $\min[1, p(\alpha_{mk}^{cand}|\cdot)/p(\alpha_{mk}^{s-1}|\cdot)]$

となる. ここで, s は MCMC の s 回目の iteration である.

β の発生

競合リスクモデルの係数 β_{mk} の発生には, 独立連鎖 M-H 法 [22] を用いる. 事後分布は以下となる.

$$p(\beta_{mk}|\cdot) \propto \prod_{i; z_i=m}^n \prod_{j=1}^{J_i} [\alpha_{mk} T_{ij}^{\alpha_{mk}-1} \exp(\mathbf{x}_{ij}^T \beta_{mk})]^{\delta_{ijk}} \times \exp(-\exp(\mathbf{x}_{ij}^T \beta_{mk}) T_{ij}^{\alpha_{mk}}) \times \exp\left\{-\frac{1}{2} (\beta_{mk} - \mathbf{b}_{0km})^T B_{0km}^{-1} (\beta_{mk} - \mathbf{b}_{0km})\right\}$$

β の候補点は $\beta_{mk}^{cand} \sim N(\beta_{mk}^{s-1} + V_{mk} \mathbf{g}_{mk}, V_{mk})$ からサンプリングする. なお, \mathbf{g}_{mk}, V_{mk} は以下である.

$$\mathbf{g}_{mk} = \frac{\partial \log p(\beta_{mk}|\cdot)}{\partial \beta_{mk}}, V_{mk} = -\left\{\frac{\partial^2 \log p(\beta_{mk}|\cdot)}{\partial \beta_{mk} \partial \beta_{mk}^T}\right\}^{-1}$$

ここで, 新しい候補点の採択確率は $\min\left[1, \frac{p(\beta_{mk}^{cand}|\cdot)q(\beta_{mk}^{s-1}|\cdot)}{p(\beta_{mk}^{s-1}|\cdot)q(\beta_{mk}^{cand}|\cdot)}\right]$ となる. $q(\cdot)$ は提案分布の密度関数である.

Γ の発生

多項ロジットモデルの係数 Γ の事後分布は以下となる.

$$p(\Gamma|\cdot) \propto \prod_{i=1}^n \prod_{m=1}^M \left\{ \frac{\exp(\mathbf{d}_i^T \gamma_m)}{\sum_{l=1}^M \exp(\mathbf{d}_i^T \gamma_l)} \right\}^{I(z_i=m)} \times \exp\left\{-\frac{1}{2} (\Gamma - \mathbf{g}_0)^T G_0^{-1} (\Gamma - \mathbf{g}_0)\right\}$$

提案分布および採択率の計算は, 上記の β_{mk} の発生と同じ独立連鎖 M-H 法を用いる.

z の発生

潜在クラスへの所属確率はバイズの定理を用いて,

$$p(z_i = m|\cdot) = \frac{p(z_i = m) \prod_{j=1}^{J_i} p(Y_{ij}|T_{ij}, z_i = m) p(T_{ij}|z_i = m)}{\sum_{l=1}^M p(z_i = l) \prod_{j=1}^{J_i} p(Y_{ij}|T_{ij}, z_i = l) p(T_{ij}|z_i = l)}$$

となる. 所属確率から多項乱数を用いて新しい所属インディケータ z をサンプリングする.

その他のパラメータ

ϕ_{mk}, σ_{mk}^2 は Gibbs Sampling を用いた潜在クラス回帰分析となる. 詳細は文献 [23] を参照されたい.