P21. オンラインフリーマーケット市場への生存時間分析の応用

込山 湧士[†]，松田 安昌[†]，石原 昌和[†‡]

[†]東北大学大学院 経済学研究科　[‡]New York University, Stern School of Business

komiyama.yuji.r2@dc.tohoku.ac.jp

1. 研究の背景と動機

メルカリは国内最大級のCtoCマーケットプレイスであり、その市場構造を分析することはCtoC市場の理解にとって重要である。本研究では、出品から取引完了までの時間を分析対象とする。取引が成立しない商品を除外すると分析に偏りが生じるため、生存時間分析を用いて売れ残り商品も含めて分析する。

標準的な生存時間分析では、すべての商品が最終的に取引されると仮定する。この仮定を緩めるために、取引が成立しない商品も存在することを想定した「治癒モデル」を導入する。ただし、従来の治癒モデルでは「絶対に取引が成立しない」商品を仮定する点が現実的でない。

そこで本研究では、有限の期間 c に注目し、「c 期間以内に取引が成立しない商品」を治癒群とみなす有限期間治癒モデルを提案する。

2. 生存時間分析

生存時間分析では、イベント（ここでは商品の取引完了）までの時間を確率変数 T とし、その分布を $F(t)$、生存関数を $S(t) = 1 - F(t)$ と定義する。
実データでは、観測期間内に取引が成立しない商品が存在する。これらは打ち切り（censoring）として扱われる。打ち切りデータは取引完了時点は不明だが、少なくとも観測終了時点までは成立していないことが分かる。
このような情報を考慮できる点が、生存時間分析の特徴である。標準的な枠組みでは、すべての商品が最終的に取引される（$\lim_{t\to\infty} S(t) = 0$）と仮定する。

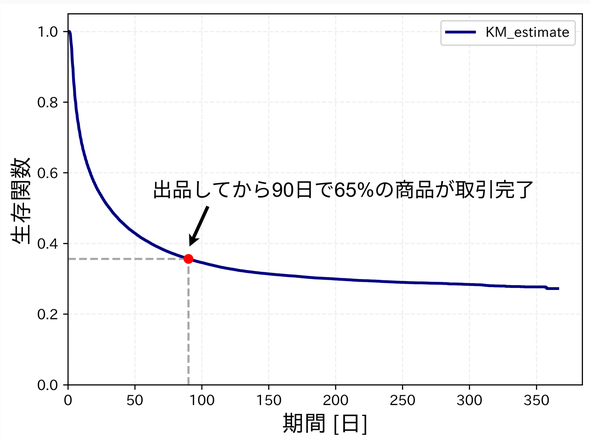

Figure 1: Tシャツ/カットソー（半袖/袖なし）のカプランマイヤー推定量 $\hat{S}(t)$

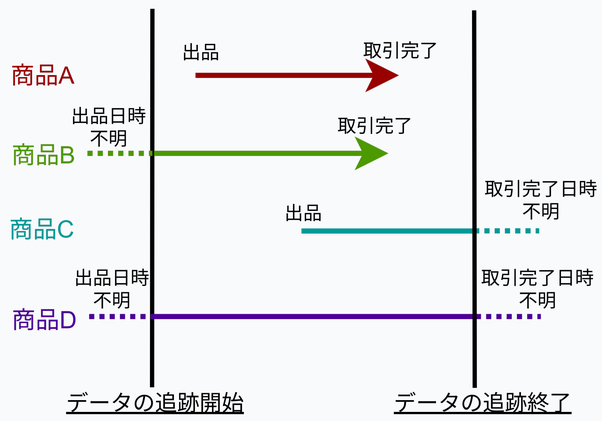

Figure 2: 観測データ（商品A）と打ち切りデータ（商品B，C，D）。本発表で分析するデータには商品Cタイプの打ち切りが含まれる。

3. 従来の Mixture Cure Model

Mixture Cure Model は治癒モデルの代表的手法であり、個体が治癒群（イベントが起こらない群）か感受群（イベントが起こる群）に属する潜在的な確率を考慮する。潜在的群への所属を表すダミー変数 Z を導入すると、$Z = 0$ が治癒群、$Z = 1$ が感受群に対応する。
生存関数は以下のように表される:

$$S(t) = (1 - \pi(x)) + \pi(x) S_p(t \mid x) \tag{1}$$

ここで $\pi(x) = P(Z = 1)$ は感受群の割合、$1 - \pi(x)$ が治癒率を表す。$S_p(t \mid x)$ は感受群の生存関数であり、従来モデルでは $\lim_{t\to\infty} S_p(t \mid x) = 0$ が仮定される。

課題:

- 無限の観測期間を前提としており、実データでの識別が困難
- 実務では有限期間での意思決定が必要
- 「永久に売れない」という概念は現実的でない

4. 提案モデル: 有限期間治癒モデル

「c 期間生存すれば治癒とみなす」という有限期間基準を導入し、式 (1) において

$$\mathrm{Supp}(S_p(t \mid x)) = [0, c], \quad S_p(c \mid x) = 0$$

とする。
$\pi(x)$ をロジスティック関数とし、感受群の確率密度関数 $f_p(t \mid x)$ を次のように表現する。

$$f_p(t \mid x; \beta) = \sum_{i=1}^{K} c_i(x; \beta) \tilde{B}_{i,3}(t),$$

ここで $\{\tilde{B}_{i,3}(t)\}_{i=1}^{K}$ は Normalized cubic B-spline の基底関数である。
係数 $c_i(x; \beta)$ は softmax 関数で表される。

$$c_i(x; \beta) = \begin{cases} \dfrac{\exp(x^\top \beta_i)}{1 + \sum_{j=1}^{K-1} \exp(x^\top \beta_j)}, & i \neq K \\ \dfrac{1}{1 + \sum_{j=1}^{K-1} \exp(x^\top \beta_j)}, & i = K. \end{cases}$$

$\pi(x)$ は「c期間以内に取引が完了する確率」を表し、$S_p(t \mid x)$ はその期間内に完了する商品の「取引完了までの時間分布」を表す。これらに対する特徴量の効果を分析することで、取引が成立しやすいかどうかに影響する要因と、取引が完了する商品において出品から完了までの時間に影響する要因を明らかにできる。さらに、この二つの効果を組み合わせることで、商品が持つ特性（狙い目、ニッチなど）を評価できる。

5. メルカリデータ (Mercari, 2023)への適用

データ概要

- 期間: 2020年1月1日〜2020年12月31日出品商品
- 抽出日: 疑似的に2021年1月1日と設定
  → 実際の抽出日は2023年6月だが、この時点では $c = 3$ か月後の全商品の治癒/非治癒が確定済みで推定の余地がなくなるため、抽出日を2021年1月1日に設定した。
- 治癒: $c = 3$ か月
  → 2020年10月1日以降の出品は右側打ち切りとなるものがある（治癒/非治癒が不明）
- 対象: Tシャツ/カットソー（半袖/袖なし）
- サンプルサイズ: 10万件

主な特徴量

商品属性（性別、ブランド、商品状態、サイズ）、配送条件（発送日数、匿名配送、送料負担）、その他（商品説明文字数、月ダミー、価格）

6. ベンチマークモデルとの比較

モデル	負の対数尤度	IBS	AUC
Cox比例ハザードモデル (Cox, 1972)	317800.95	0.21966	0.530
Cox + Logistic Mixture Cure (Sy & Taylor, 2000)	316659.85	0.21918	0.503
提案モデル	**293346.91**	**0.21817**	**0.737**

Table 1: ベンチマークモデルとの比較結果。ここでの AUC は、抽出日を疑似的にずらすことで打ち切り扱いとなった商品が3か月後に取引完了しているかどうかを判定する分類性能を評価した指標である。これら三つの指標すべてにおいて、提案モデルは既存モデルより良い性能を示した。

7. 主な知見

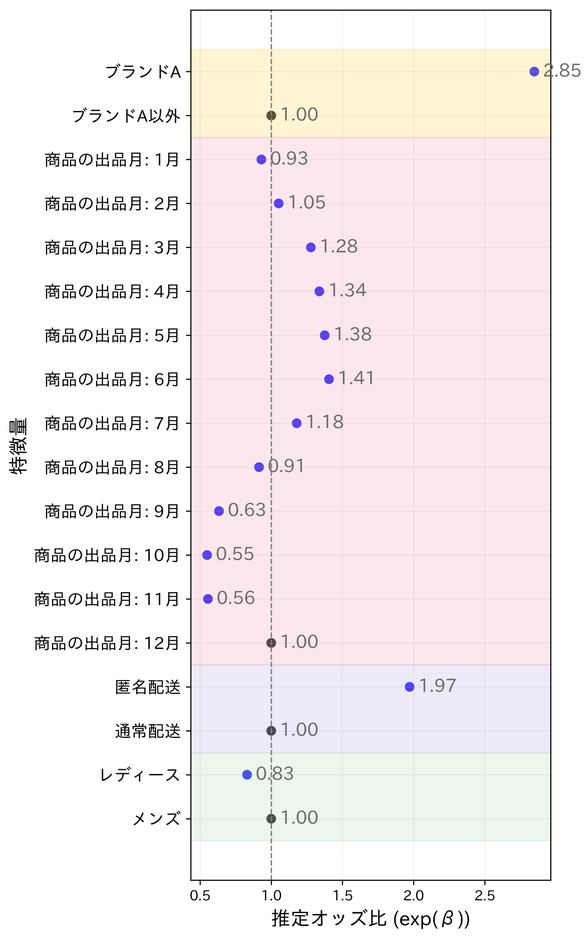

Figure 3: 3か月後までの取引完了確率に関する推定オッズ比（黒点は基準カテゴリーを示す）

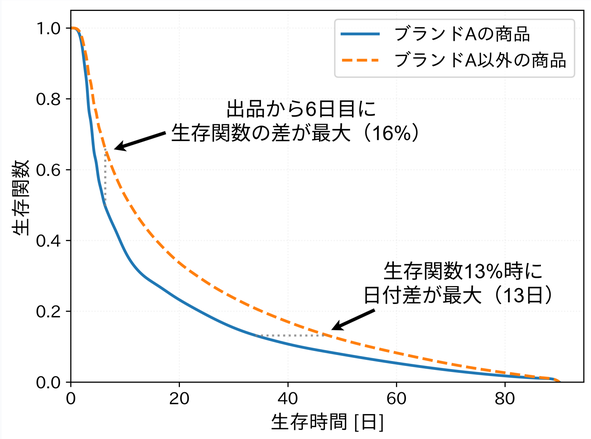

Figure 4: ブランドAと非Aの平均感受群生存関数

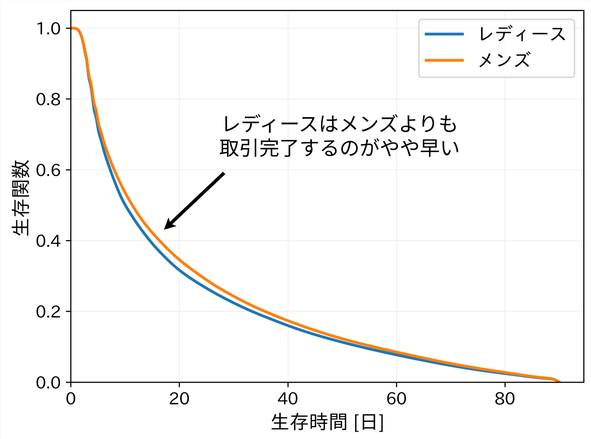

Figure 5: 性別別の平均感受群生存関数

- メンズはレディースより取引完了確率が高いが、取引完了時間はやや長い。
- ブランドAの商品は、ブランドA以外の商品よりも3か月後までに取引が完了する確率が高く、3か月以内に完了する商品の中でも取引が早い傾向がある。
- 出品から3か月後の時期が春から夏にあたる商品は、秋冬期にあたる商品よりも3か月後までに取引が完了する確率が高い。一方、従来モデル (Sy & Taylor, 2000)では出品月が年初から年末に向かうほど完了確率が単調に低下すると推定されており、提案モデルの方が説得力のある解釈を与える。

8. 結論

有限期間を考慮した治癒モデルを提案し、メルカリデータへの適用によりオンラインフリーマーケット市場における実務的な分析と予測が可能であることを示した。

謝辞

本研究では、国立情報学研究所のIDRデータセット提供サービスにより株式会社メルカリから提供を受けた「メルカリデータセット」を利用した。

参考文献

Cox, D. R. (1972).Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *34*(2), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Mercari, I. (2023). *Mercari dataset*. Informatics Research Data Repository, National Institute of Informatics. https://doi.org/10.32130/idr.17.1

Sy, J. P., & Taylor, J. M. (2000).Estimation in a Cox proportional hazards cure model. *Biometrics*, *56*(1), 227–236. https://doi.org/10.1111/j.0006-341X.2000.00227.x