

k 匿名化を用いた選択モデルの有用性と安全性の検証

2025年2月4日

高橋研究室

21M1034 佐野 優志

目 次

第 1 章 はじめに	1
1.1 背景	1
1.2 研究の目的	2
1.3 本研究の構成	3
第 2 章 k 匿名化	4
2.1 k 匿名化とは	4
2.2 k 匿名化の仕組みとその重要性	5
第 3 章 GNL モデル	6
3.1 選択確率	6
3.2 GNL モデルにおける集計問題の生成例	8
3.3 集計ルールが満たすべき条件	9
3.4 GNL モデルにおける集計ルールの導入：ネストの集計	10
3.5 GNL モデルにおける集計ルールの導入：選択肢の集計	12
第 4 章 結果	13
4.1 ブランド統合前の結果	13
4.2 ブランド統合後の結果	15
4.3 ブランド統合による影響	16
4.4 選択肢を統計した場合の確定的効用の範囲	18
第 5 章 おわりに	20
5.1 結論	20
5.2 今後の課題	20

付 錄 A GNL における二つの選択確率定式化の等価性	21
付 錄 B 例示に用いたデータの詳細	23
参考文献	26

図 目 次

3.1	k 匿名化統合前後の GNL 構造	9
4.1	ブランド統合前後 (1day) での GNL 構造	16
4.2	ブランド統合前後 (2week) での GNL 構造	17
4.3	ブランド統合前後 (1month) での GNL 構造	17

表 目 次

4.1 ブランド統合前の最尤推定されたパラメータ	14
4.2 ブランド統合後の最尤推定されたパラメータ	15
4.3 確定期的効用の範囲	19
B.1 集計問題例に用いたスキャンパネルデータ詳細	24
B.2 ブランド統合前後の選択肢	24

第1章 はじめに

1.1 背景

デジタル技術の急速な進化により、私たちの日常生活は膨大なデータとともにある時代へと突入している。特に、インターネットの普及やスマートフォンの発展に伴い、個人の行動履歴や購買履歴、さらにはオンライン上の嗜好情報などがデータとして収集・蓄積されるようになっている。これらのデータは、マーケティング、医療、公共政策など、さまざまな分野において有益な活用が進められており、データ駆動型社会の発展に寄与している。

しかしその一方で、個人情報の取り扱いに関する懸念が急速に高まっている。具体的には、個人データがどのように収集・管理され、どの程度の安全性が確保されているのかが不透明であることが問題視されている。近年、データ漏洩や不適切な利用が社会問題として頻繁に取り上げられており、プライバシー保護はますます重要な社会的課題となっている。

このような状況に対応するため、データ匿名化技術が注目を集めている。データ匿名化技術は、個人情報を保護しながらデータの利便性を維持するための手段として、多様な手法が提案されている。その中でも、 k 匿名化は特に広く利用されている技術の一つである。例として、医療分野においては、患者の病歴や診断データを匿名化することで、研究者が貴重な医療情報を活用できるようになりつつも、患者のプライバシーが保護される。また、消費者行動分析においても、匿名化されたデータを利用することで、消費者の購買傾向や嗜好を把握しながら、個人を特定するリスクを低減することが可能となる。

一方で、マーケティングや行動経済学の分野では、消費者の選択行動を理解するためにブランド選択モデルが広く利用されている。ブランド選択モデルは、消費者が商品やサービスを選択する際の意思決定プロセスを数理的に表現し、効果的な広告戦略や販売促進を行うための基盤となる。近年では、機械学習や統計解析を活用した高度なブランド選択モデルが開発され、消費者の購買行動の予測精度が向上している。

しかし、これらのモデルを高精度で運用するためには、消費者個人の詳細なデータが必要不可

欠である。この点が、プライバシー保護の観点から重要な課題となっている。具体的には、個人情報をそのまま利用することによる倫理的な問題や、法規制への抵触の可能性が懸念される。たとえば、消費者の詳細な行動履歴や購買履歴を無制限に活用することは、消費者に不安を与え、企業への信頼を損なう要因となり得る。

このように、プライバシー保護とデータ活用の両立は、現代社会において解決すべき重要な課題として浮かび上がっている。本研究は、この課題に対する具体的な解決策を模索し、 k 匿名化を活用した選択モデルの安全性と有用性のバランスを検討することを目的とする。

1.2 研究の目的

本研究の目的は、個人情報保護のための匿名化技術である k 匿名化をブランド選択モデルに適用することにより、プライバシー保護とデータの有用性をどのように両立できるかを明らかにすることである。近年、消費者データの活用が進む中で、個人のプライバシーを適切に保護しながら、データの分析精度を維持することが求められている。しかし、匿名化を施すことでデータの細部が失われ、分析の精度に影響を与える可能性がある。本研究では、このトレードオフの特性を詳細に検討し、最適なバランスを探る。

具体的には、 k 匿名化を適用することによるデータの変化が、分析結果やモデルの精度に与える影響を定量的に評価する。まず、 k 匿名化を適用したデータセットと、適用前のデータセットを比較し、匿名化がデータの統計的特性に及ぼす影響を詳細に検証する。この評価においては、異なる k の値を設定し、それぞれの設定においてプライバシー保護の強度とデータ分析の精度にどのような変化が生じるかを確認する。

さらに、ブランド選択モデルとしてGNL (Generalized Nested Logit) モデルを採用し、 k 匿名化がモデルの推定精度やパラメータの推定値に与える影響を詳細に分析する。GNL モデルは、消費者の選択行動を説明するための強力な手法であり、選択肢間の相関を考慮できる点で広く活用されている。本研究では、匿名化がGNL モデルの識別性能や予測精度にどのような影響を及ぼすかを検討する。

また、 k 匿名化の適用によってブランド選択モデルの結果がどのように変化するのかを、消費者行動のパターンや選択確率の変動を通じて詳細に評価する。この分析を通じて、匿名化がブランド選択の意思決定プロセスに及ぼす影響を解明し、データの一般化や情報の統合がモデルの精度

にどのような影響を与えるのかを明らかにする。

1.3 本研究の構成

本研究は、全5章からなる。次章以降は次のように構成されている。

第2章では、 k 匿名化の仕組みとその重要性について詳しく説明する。第3章では、GNLモデルにおける選択確率や構造について概説し、集計ルールが満たすべき条件を整理するとともに、ネストと選択肢ごとの集計ルールについて詳述する。第4章では、 k 匿名化の適用前後における最尤推定の結果を比較し、確定的効用の範囲に与える影響を評価する。最後に、第5章では本研究の結論をまとめるとともに、今後の課題について述べる。

第2章 k 匿名化

2.1 k 匿名化とは

k 匿名化は、データ共有や分析を行う際に、個人が特定されるリスクを軽減するためのデータ保護技術の一つである。この技術は、データセット内の個人情報を統計的に処理し、特定の個人が識別されにくい状態を作り出すことで、プライバシーを保護することを目的としている。具体的には、データセット内の各個人が少なくとも k 人以上のグループに含まれるようにデータを変換し、個別の識別が困難な状態にする。この結果、個人特定の可能性が大幅に低下し、安全なデータの共有や分析が可能となる。

この手法は、特にプライバシー保護が求められる場面で有用であり、医療、行政、マーケティングなど、多岐にわたる分野で広く利用されている。たとえば、医療分野では、患者の診療記録を研究目的で利用する際に、患者個人を特定できないようにするために k 匿名化が適用される。同様に、行政機関では統計データを公表する際に、個人のプライバシーを確保するためにこの手法が導入されている。また、マーケティング分野では、消費者の行動データを匿名化することで、企業が市場分析を行う際にプライバシーリスクを低減しつつ、データの活用を可能にする [1]。

k 匿名化の基本的な考え方は、個人がデータセット内で識別されるリスクを統計的に最小化することである。このために、特定のデータ項目（名前や住所などの直接識別子）を削除するとともに、間接的に個人を特定できる可能性のあるデータ（年齢、性別、居住地など）を一般化することで、データの匿名性を強化することで、個人情報の漏洩リスクを低減しつつ、データの統計的な有用性を一定程度維持することが可能となる。

しかし、 k 匿名化にはいくつかの限界がある。まず、匿名化されたデータに対して高度な分析を行うことで、特定のグループやパターンが推測されるリスクが残る点が挙げられる。たとえば、同じ属性を持つデータが少数しか存在しない場合、匿名化を施しても特定の個人やグループが識別される可能性がある。また、 k 匿名化を強化することでデータの識別性を低下させると、データの詳細性が失われ、分析の精度が低下する可能性もある。このため、プライバシー保護の強度と

データの有用性のバランスを適切に取ることが求められる [2].

2.2 k 匿名化の仕組みとその重要性

k 匿名化を実現するためには、主に「一般化」と「レコード削除」の2つの手法が用いられる [3]. これらの手法は、データの匿名性を確保するだけでなく、分析や利用のためのデータ有用性を維持することを目的としている。

一般化は、データ中の具体的な情報を抽象化することで匿名性を向上させる手法である。たとえば、特定の日付「2024年12月31日」を「2024年12月」や「2024年第52週」といったより広い範囲に変換することが挙げられる。この変換により、個人が特定されるリスクを低減しつつ、分析可能なデータとしての有用性を維持することが可能である。一般化は、特に属性データに適用されることが多く、たとえば年齢を「35歳」から「30~39歳」の範囲に変換したり、住所を「東京都渋谷区」から「東京都」まで抽象化することで、データの匿名性を高める効果がある。一方で、一般化の程度が過剰になると、データの具体性や分析の精度が低下する可能性があり、匿名性と有用性のバランスを適切に取ることが重要である。

レコード削除は、識別リスクが高いデータを完全に削除する手法である。たとえば、データセットから名前や正確な住所などの直接的な識別子を削除することで、個人特定の可能性を排除する。この方法は匿名性を大幅に向上させる一方で、データの有用性が著しく損なわれる可能性があるため、適用には慎重な判断が必要である。削除された情報が分析結果に大きな影響を与える場合、他の匿名化手法と組み合わせることが求められる。

k 匿名化の重要性は、プライバシー保護とデータ活用の両立を可能にする点にある。データセットを匿名化することで、個人情報が不正利用されるリスクを軽減し、法規制に適合した形でデータの共有や分析が可能となる。たとえば、医療分野では患者のプライバシーを守りながら、疫学研究や治療法開発のためのデータ分析を進めることができる。同様に、マーケティング分野では、消費者の個人情報を保護しつつ、購買傾向や商品需要を解析することが可能である。

本研究では、これらの仕組みをブランド選択モデルに適用し、 k 匿名化がもたらす影響を詳細に検証する。これにより、プライバシー保護とデータ活用を両立させるための具体的な手法とその限界を明らかにすることを目指す。

第3章 GNL モデル

本節では、まず GNL モデルの選択確率を示す。次に、GNL モデルに基づく集計問題の発生例を示し、集計ルールが満たすべき条件について論じる。その後、ブランド選択モデルにおけるネストと選択肢、それぞれの集計ルールを示す。

3.1 選択確率

本研究では、GNL モデルをブランド選択モデルに適用し、データ匿名化が選択モデルに与える影響を検証する。GNL モデルは、選択行動を分析するための確率モデルの一つであり、特に選択肢間の相関性を考慮した分析が可能な点で優れている。このモデルを使用し、匿名化によるデータ構造の変化が消費者行動やモデルの性能に及ぼす影響を詳細に評価することを目的とする。

GNL モデルでは、ネスティングルール（選択肢を階層的に構造化する規則）を商品の属性に基づいて定義する。また、本研究では消費者が同質であると仮定し、すべての消費者が同じ選択構造に従うものとする。選択肢の構造としては、商品 l が特定の構成要素 m を持つものとし、消費者はまず構成要素 m を選択した後、その構成要素 m を持つ商品 l を選ぶものとする。このプロセスは 2 段階で構成されており、第一段階では構成要素 m の選択が行われ、第二段階でその構成要素を持つ商品が選ばれる。このような選択プロセスは、消費者がまず商品カテゴリや属性（例えば「低価格」「大容量」など）に基づいて候補を絞り込み、その後ブランド名やその他の要素を基に具体的な商品を選択する現実の購買行動に即している。

ここで、同一の商品 l が複数の構成要素 m' に属することは許されないとする。つまり、各ブランドは一意に特定の構成要素に紐付けられる。一方で、同一の構成要素 m' には複数の商品が属することが可能である。このルールにより、商品と構成要素の関係は一対一ではなく、一対多の関係として表現される。このため、消費者が特定の構成要素 m を選択すると、その構成要素に属する商品 $l' \in \mathcal{L}_m$ のみが選択肢として残ることになる。このようにして、消費者が選択するブランドの候補が構成要素の選択によって限定される点が、本モデルの重要な特徴である。以上のモデル

構造により、GNL モデルは選択肢間の相関性や階層的な選択プロセスを考慮した精密な選択行動の解析を可能にする。

構成要素 m を通じて選択された商品 l の効用 \hat{U}_{ml} は、次のように表される：

$$\hat{U}_{ml} = V_m + \epsilon_l, \quad (3.1)$$

$$V_l = \hat{V}_m + \hat{V}_{ml} + \hat{V}_l, \quad (3.2)$$

$$\epsilon_l = \hat{\epsilon}_l + \hat{\epsilon}_{ml}. \quad (3.3)$$

ここで、 \hat{V}_m は商品構成要素（ネスト） m を選択した場合に得られる確定的効用を表し、 \hat{V}_{ml} は商品 l を構成要素 m の下で選択した場合に得られる確定的効用を示す。また、 \hat{V}_l は商品 l による確定的効用を表し、 V_l はこれらの効用の和として計算される総合的な効用である。同様に、確率的効用については、 $\hat{\epsilon}_m$ が商品構成要素 m を選択したことに対応する確率的効用を表し、 $\hat{\epsilon}_{ml}$ が商品 l の残る確率的効用を示す。これらの確率的効用は、選択行動の不確実性や選択肢間の個別性をモデル化するために用いられる。

Wen と Koppelman[4] に基づき、式 (3.4)～(3.6) の条件を適用して整理を行う。この整理により、ログサム形式を採用した GNL モデルにおけるブランド l の選択確率 \hat{P}_l は、以下の式で表される：

[GNL-LS]

$$\hat{P}_l = \sum_m \hat{P}_m \hat{P}_{l|m}, \quad (3.4)$$

$$\hat{P}_m = \frac{\exp(\hat{V}_m + V'_m)}{\sum_m \exp(\hat{V}_m + V'_m)}, \quad (3.5)$$

$$\hat{P}_{l|m} = \frac{(\gamma_{lm} \exp V_l^h)^{1/\mu_m}}{\sum_{l' \in \mathcal{N}_l} (\gamma_{l'm} \exp V_{l'}^h)^{1/\mu_m}}. \quad (3.6)$$

\hat{P}_m はネスト m を選択する確率で、 $\hat{P}_{l|m}$ はネスト m の下で商品 l を選択する確率である。ここで、

$$\hat{V}'_m := \mu_m \ln \sum_{l' \in \mathcal{N}_m} \left(\gamma_{l'm} \exp(\hat{V}_{ml'} + \hat{V}_{l'}) \right)^{1/\mu_m} \quad (3.7)$$

であり, \hat{V}'_m は通常ログサムと呼ばれる. N_m は商品構成要素集合, μ_m はネスト間の非類似度パラメータであり, 効用最大化と整合的であるためには,

$$0 < \mu_m \leq 1 \quad (3.8)$$

を満たす必要がある. 同様に γ_{lm} はアロケーションパラメータであり,

$$\gamma_{lm} \geq 0, \quad \forall m, l, \quad (3.9)$$

$$\sum_m \gamma_{lm} = 1, \quad \forall l \quad (3.10)$$

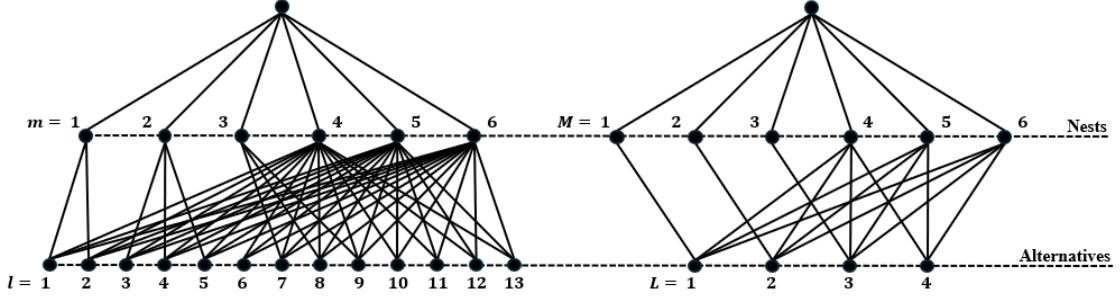
を満たす必要がある. GNL モデルにおいて, 属性分割 (Attribute Separation) というネスティングルールを採用した場合, 消費者はまずブランドの構成要素を選択し, その後に具体的なブランドを選択すると解釈される. GNL モデルにおけるパラメータの推定は, 一般的に最尤推定法によって行われる. 最尤推定は, 観測データに基づいてモデルのパラメータを推定する統計的手法であり, 尤度関数を最大化することで最も適切なパラメータを求めることができる. しかし, ネストや選択肢の数が増加すると, それに伴い非類似度パラメータやアロケーションパラメータの数も増加する. 特にアロケーションパラメータは, ネスト数や選択肢数のいずれが増加しても比例的に増加するため, 計算負荷が大きくなるという課題がある. このような問題を軽減するために, k 匿名化を適用することで選択肢やネスト数を統合し, 計算時間の短縮を図ることが可能となる.

3.2 GNL モデルにおける集計問題の生成例

本節では, GNL モデルにおいて集計問題が発生する生成例を説明するとともに, 直感的に考えられる集計ルールでは克服できないことを例 (ブランド選択モデル: 紙パック牛乳) を用いて示す. 使用するデータの詳細については付録 B を参照されたい.

選択肢数は 13 個, そしてその選択構造は図 3.1 左図に示すように, ネスト数 6 つの構造としよう. ここでのネスティングルールは, 商品が持つ属性 (値) をネストとして分類する属性分割 (Attribute Separation) の手法を用いる. ブランド l の確定的効用については, 以下に示す線形効用関数で表されるものとする:

$$V_l = \sum_{i=1}^6 \alpha_i X_{il}. \quad (3.11)$$

図 3.1: k 匿名化統合前後の GNL 構造

ここで、 X_{1l} はブランド（イオン）・ダミー（イオン=1, その他=0）、 X_{2l} はブランド（古谷）・ダミー（古谷=1, その他=0）、 X_{3l} はブランド（雪印）・ダミー（雪印=1, その他=0）、 X_{4l} は l の価格、 X_{5l} は l の容量、 X_{6l} は低温殺菌・ダミー（低温殺菌=1, 普通=0）、 α_i は i 番目の変数に対応するパラメータである。 k 匿名化を適用するにあたり、ブランド名を「イオン」「古谷」「雪印」「その他」の 4 つのカテゴリに統合し、購入日時を「2週間ごと」または「1か月ごと」に一般化する。当初、購入日時を「1週間ごと」に一般化することを検討したが、ブランド名を統合した「雪印」において、「2001 年第 6 週」の購買機会が 1 回のみとなり、一意的な個人識別が可能となるリスクが生じる。そのため、「1週間ごと」の一般化は比較対象から除外する。匿名化前後の選択構造の変化については図 3.1 に示すように、匿名化前は 13 個、匿名化後は 4 個とする。このように一般化を伴う匿名化を行うことで、選択モデルに及ぼす影響を詳細に検証する。

3.3 集計ルールが満たすべき条件

本節では、集計ルールが満たすべき条件について示す。Sweet[5] は集計が整合的に行われているために満たすべき条件として、次の 2 つの条件を示している。

条件 1 固定需要

個人 h が集計化された選択肢 L を選ぶ周辺確率 \hat{P}_L^h は、統合前の選択肢 l の効用 \hat{V}_l^h を共通の効用である \hat{V}_L^h に置き換えることによって等しい。

条件 2 選択確率の整合性

個人 h が集計化された選択肢 L を選ぶ周辺確率 \hat{P}_L^h は、統合前の選択肢 l ($l \in L$) の選択確率と等

しい：

$$\hat{P}_l = \sum_{l \in L} \hat{P}_l. \quad (3.12)$$

これ以外に、アロケーションパラメータの条件が必要となる。ここでは、次の条件提示する：

条件 3 アロケーションパラメータの整合性

$$\gamma_{lM} = \sum_{m \in M} \gamma_{lm}, \quad \forall l. \quad (3.13)$$

条件 3 は集計が式 (3.10) と整合的であるための条件である。ただし、これはネストが集計化される場合についてのみの条件であり、選択肢が集計化される場合には適用することができない [6]。

3.4 GNL モデルにおける集計ルールの導入：ネストの集計

本節では、高橋 [7] が示すネストの集計ルールについて説明する。ここでは、集計の例と同様に、まずネストとして商品構成要素 m を選択し、その後に具体的な商品 l を選択するものとする。ネスト（商品構成要素）は $m \in M$ という表現を用いる。ここで、 m はブランド統合前のネストを指し、 M は統合後のネストを意味している。本研究で求めるべき集計ルールとは、 \hat{V}_M および \hat{V}'_M に関する条件を指す。この条件をネスト m に適用し、さらに GNL の選択確率式 (3.5) に代入することで、モデル全体の計算過程を整理する。

$$\exp(\hat{V}_M + \hat{V}'_M) = \sum_{m \in M} \exp(\hat{V}_m + \hat{V}'_{m'}) \quad (3.14)$$

を得る。ここで、条件 1 より

$$\hat{V}'_M = \ln \sum_{m \in M} \exp \hat{V}'_m + \hat{V}'_{M'} \quad \forall m \quad (3.15)$$

であるため、式 (3.14) は、

$$\exp(\hat{V}_M + \hat{V}'_M) = \sum_{m \in M} \exp(\hat{V}_m + \hat{V}'_{m'}) \quad (3.16)$$

と書き直せる。ここで、式 (3.16) より \hat{V}_M に関する部分を消去し、

$$\hat{V}_M = \ln \sum_{m \in M} \exp \hat{V}'_m \quad (3.17)$$

を得る。式 (3.17) を式 (3.16) に代入し、これを整理することにより、

$$\hat{V}_M = \ln \sum_{m \in M} \exp \left(\hat{V}_m + \hat{V}_{m'} \right) - \ln \sum_{m \in M} \exp \hat{V}'_m \quad (3.18)$$

となる。これらの式のうち、式 (3.17) と (3.18) が GNL におけるブランド選択モデルでネストを集計した場合の満たすべき集計ルールとなる。

次に、ネストを集計した場合の集計ルール、ブランド統合前のネスト \hat{V}_m と統合後のネスト \hat{V}_M について示す。

$$\hat{V}_m = \sum_{m' \in N_m} \left(\gamma_{l'm} \exp \left(\hat{V}_{ml'} + \hat{V}'_{l'} \right) \right)^{1/\mu_m}, \quad (3.19)$$

$$\hat{V}_M = \ln \sum_{m \in M} \exp \hat{V}_m. \quad (3.20)$$

V_m と \hat{V}_m , \hat{V}'_m の関係は、式 (3.2) 及び付録 A を参照せよ。

また、 V_m については Sweet でいうセントラル・ログサムにあたる。従って Sweet と同様に、集計による各確定的効用項の範囲については次のとおりになる：

$$\min_{m \in M} \hat{V}_m \leq \hat{V}_M \leq \max_{m \in M} \hat{V}_m. \quad (3.21)$$

\hat{V}'_M についても同様に、

$$\min_{m \in M} \left(\hat{V}_m + \hat{V}'_m \right) - \max_{m \in M} \hat{V}'_m \leq \hat{V}_M \leq \max_{m \in M} \left(\hat{V}_m + \hat{V}'_m \right) - \min_{m \in M} \hat{V}'_m \quad (3.22)$$

を得る。 \hat{V}_{ml} については、 \hat{V}'_M は式 (3.22) に示すとおりであり、 \hat{V}_l は集計と無関係であるため定数である。従って、 \hat{V}_{ml} の上限、下限は最後の項 Ω に左右される。 Ω は、 $\hat{P}_{l|m}$ を \hat{P}_m で重み付けし、それを μ_m 乗したもの γ_{lm} で除し、対数を取ったものである：

$$\Omega = \ln \left(\frac{1}{\sum_{m \in M} \gamma_{lm}} \left(\frac{\sum_{m \in M} \hat{P}_{l|m} \hat{P}_m}{\sum_{m' \in M} \hat{P}_{m'}} \right)^{\mu_m} \right) \quad (3.23)$$

ここで、式 (3.9), (3.10) より、

$$0 \leq \sum_{m \in M} \gamma_{lm} = \gamma_{lM} \leq 1 \quad (3.24)$$

である。一般的に γ_{lM} が大きい場合、つまり多くのネストを束ねた場合、 Ω はマイナスになりやすくなる。また、式 (3.8), $0 \leq \hat{P}_m \leq 1$, $0 \leq \hat{P}_{l|m} \leq 1$ より、

$$0 \leq \left(\frac{\sum_{m \in M} \hat{P}_m \hat{P}_{l|m}}{\sum_{m' \in M} \hat{P}_{m'}} \right)^{\mu_m} \leq 1 \quad (3.25)$$

である。一般的に、 $\hat{P}_{l|m}$ が大きい場合、すなわちネスト m に属する選択肢が魅力的である場合、 Ω は正の値をとりやすくなる。最終的な Ω の正負は、式 (3.24) および式 (3.25) の大小関係によって決定される。しかし、 Ω の最大値や最小値を厳密に決定することはできず、同様に V_{MI} についてもその範囲を確定することは困難である。なお、本研究では、ブランド統合の前後においてネストの種類に変化がないことから、これらの計算は省略するものとする。

3.5 GNL モデルにおける集計ルールの導入：選択肢の集計

本節では、高橋が示す選択肢の集計ルールについて説明する。まずネストとして商品構成要素 m を選択し、次に商品 l を選択するものとする。選択肢（ブランド）を $l \in L$ という表現を用いよう。ここで l はブランド統合前の選択肢、 L は統合後の選択肢を意味している。

ここで求めるべき集計ルールとは、 \hat{V}_{mL} 、 \hat{V}_L に関する条件を指す。条件 2 を $\hat{P}_{l|m}$ 、 $\hat{P}_{L|m}$ に適用することにより、

$$\left(\gamma_{Lm} \exp \left(\hat{V}_{mL} + \hat{V}_L \right) \right)^{1/\mu_m} = \sum_{l \in L} \left(\gamma_{lm} \exp \left(\hat{V}_{ml} + \hat{V}_l \right) \right)^{1/\mu_m} \quad (3.26)$$

を得る。ここで条件 1 より、

$$\left(\gamma_{Lm} \exp \left(\hat{V}_{mL} + \hat{V}_L \right) \right)^{1/\mu_m} = \sum_{l \in L} \left(\gamma_{lm} \exp \left(\hat{V}_{ml} + \hat{V}_l \right) \right)^{1/\mu_m} \quad (3.27)$$

となる。式 (3.27) を \hat{V}_L について解くと、

$$\hat{V}_L = \mu_m \ln \left(\frac{1}{\gamma_{Lm}} \sum_{l \in L} \left(\gamma \exp \hat{V}_l \right)^{1/\mu_m} \right) \quad (3.28)$$

を得る。ちなみに式 (3.28) の条件は、 \hat{P}'_m に条件 1 を適用しても同様に得られる。次に、式 (3.28) を式 (3.26) に代入し、整理すると次の条件を得る：

$$\hat{V}_{mL} = \mu_m \ln \left(\frac{1}{\gamma_{Lm}} \sum_{l \in L} \left(\gamma_{lm} \exp \hat{V}_{ml} \right)^{1/\mu_m} \right) - \mu_m \ln \left(\frac{1}{\gamma_{Lm}} \sum_{l \in L} \left(\gamma_{lm} \exp \hat{V}_l \right)^{1/\mu_m} \right) \quad (3.29)$$

これらの式のうち、式 (3.28) と (3.29) が、GNL におけるブランド選択モデルでネストを集計した場合の満たすべき集計ルールとなる。

また、集計による各確定的効用項の範囲については次のとおりになる：

$$\min_{l \in L} \gamma_{lm} \hat{V}_l \leq \gamma_{Lm} \hat{V}_L \leq \max_{l \in L} \gamma_{lm} \hat{V}_l, \quad (3.30)$$

これらは、データのばらつきを考慮した最適化問題や確率モデルにおいて、変数が取りうる範囲を制限するために使用される。

第4章 結果

本節では、 k 匿名化統合前後での最尤推定された結果の比較と選択肢を集計した場合の確定的効用の範囲について述べる。

4.1 ブランド統合前の結果

まず、ブランド統合前の結果について、最尤推定によって得られた各パラメータの推定値を基に、その傾向を示す。

効用関数のパラメータを見ると、期間別で一貫した傾向が見られるものの、パラメータごとに顕著なばらつきが存在する。 α_2 では一貫して正の値を示し、 α_4 では一貫して負の値を示しており、期間別で見ても同じような傾向が示唆される。 α_1 と α_6 では、期間別でパラメータの符号に変化があり、ブランド選択の時間依存性や消費者の選好変化を反映している可能性を示唆している。

非類似度パラメータについては、期間ごとに大きなばらつきが見られる。特に、 μ_3 および μ_5 の 2 週間および 1 か月の推定値が 0.9999 や 0.9979 と非常に高い値を示していることから、特定のブランド間で強い相関があることが示唆される。これは、消費者が特定のブランドを選択する際に、代替選択肢としての競争関係が顕著であることを示している。一方で、 μ_1 および μ_2 は期間によって極端に低い値を示す場合があり、短期間では選択行動がより分散している可能性を示唆している。

アロケーションパラメータに関しては、期間ごとに異なるブランドの分布が観察された。例えば、 γ_{55} や γ_{56} は比較的安定しているが、 γ_{75} や γ_{76} のようなパラメータは大きく変動し特定の期間で極端な値を示している。また、0 の値を示すパラメータが存在することは、ある特定のブランドやカテゴリーが消費者の選択肢から事実上排除されている可能性を示唆する。

表 4.1: ブランド統合前の最尤推定されたパラメータ

1day		2week		1month		1day		2week		1month	
Para.	Value	Para.	Value	Para.	Value	Para.	Value	Para.	Value	Para.	Value
α_1	0.6809	α_1	-1.4033	α_1	1.8754	γ_{55}	0.1843	γ_{55}	0.3693	γ_{55}	0.3880
α_2	1.5179	α_2	0.2870	α_2	0.0018	γ_{56}	0.1830	γ_{56}	0.6201	γ_{56}	0.0348
α_3	1.8319	α_3	0.3260	α_3	-0.8181	γ_{64}	0.3997	γ_{64}	0.0549	γ_{64}	0.2994
α_4	-0.0258	α_4	-0.0389	α_4	-0.0377	γ_{65}	0.4183	γ_{65}	0.2292	γ_{65}	0.6687
α_5	0.0090	α_5	0.0084	α_5	0.0081	γ_{66}	0.1820	γ_{66}	0.7159	γ_{66}	0.0319
α_6	0.4410	α_6	0.8195	α_6	-6.9543	γ_{73}	0.0415	γ_{73}	0.0000	γ_{73}	0.0414
μ_1	0.7983	μ_1	0.0001	μ_1	0.0253	γ_{74}	0.0282	γ_{74}	0.0000	γ_{74}	0.0000
μ_2	0.5509	μ_2	0.0001	μ_2	0.0027	γ_{75}	0.1850	γ_{75}	1.0000	γ_{75}	0.9428
μ_3	0.5864	μ_3	0.9999	μ_3	0.7843	γ_{76}	0.7453	γ_{76}	0.0000	γ_{76}	0.0158
μ_4	0.0631	μ_4	0.1472	μ_4	0.8909	γ_{83}	0.1820	γ_{83}	0.9989	γ_{83}	0.3258
μ_5	0.6148	μ_5	0.9979	μ_5	0.7366	γ_{84}	0.1268	γ_{84}	0.0011	γ_{84}	0.0074
γ_{11}	0.0022	γ_{11}	0.0000	γ_{11}	0.0425	γ_{85}	0.1880	γ_{85}	0.0000	γ_{85}	0.0025
γ_{14}	0.0086	γ_{14}	0.0000	γ_{14}	0.8331	γ_{86}	0.5031	γ_{86}	0.0000	γ_{86}	0.6643
γ_{15}	0.0033	γ_{15}	0.9980	γ_{15}	0.0670	γ_{93}	0.1159	γ_{93}	0.0000	γ_{93}	0.0133
γ_{16}	0.0049	γ_{16}	0.0020	γ_{16}	0.0574	γ_{94}	0.5497	γ_{94}	0.0000	γ_{94}	0.0087
γ_{21}	0.3499	γ_{21}	0.0000	γ_{21}	0.3288	γ_{95}	0.1396	γ_{95}	1.0000	γ_{95}	0.1116
γ_{24}	0.3168	γ_{24}	0.0000	γ_{24}	0.6692	γ_{96}	0.1949	γ_{96}	0.0000	γ_{96}	0.8664
γ_{25}	0.2901	γ_{25}	0.2560	γ_{25}	0.0014	γ_{104}	0.3390	γ_{104}	0.0000	γ_{104}	0.0001
γ_{26}	0.0433	γ_{26}	0.7440	γ_{26}	0.0006	γ_{105}	0.6240	γ_{105}	0.0000	γ_{105}	0.9951
γ_{32}	0.1129	γ_{32}	0.0000	γ_{32}	0.0062	γ_{106}	0.0370	γ_{106}	1.0000	γ_{106}	0.0048
γ_{34}	0.7800	γ_{34}	0.1723	γ_{34}	0.9805	γ_{114}	0.2236	γ_{114}	0.0000	γ_{114}	0.6452
γ_{35}	0.1009	γ_{35}	0.8195	γ_{35}	0.0124	γ_{115}	0.7054	γ_{115}	1.0000	γ_{115}	0.1148
γ_{36}	0.0063	γ_{36}	0.0082	γ_{36}	0.0010	γ_{116}	0.0710	γ_{116}	0.0000	γ_{116}	0.2400
γ_{42}	0.0713	γ_{42}	0.0000	γ_{42}	0.1225	γ_{124}	0.5980	γ_{124}	0.0000	γ_{124}	0.0128
γ_{44}	0.4635	γ_{44}	0.0014	γ_{44}	0.0264	γ_{125}	0.2121	γ_{125}	0.0020	γ_{125}	0.9630
γ_{45}	0.3807	γ_{45}	0.9986	γ_{45}	0.4901	γ_{126}	0.1899	γ_{126}	0.9980	γ_{126}	0.0243
γ_{46}	0.0845	γ_{46}	0.0000	γ_{46}	0.3611	γ_{134}	0.1917	γ_{134}	1.0000	γ_{134}	0.0124
γ_{52}	0.0533	γ_{52}	0.0000	γ_{52}	0.5338	γ_{135}	0.8083	γ_{135}	0.0000	γ_{135}	0.9876
γ_{54}	0.5794	γ_{54}	0.0106	γ_{54}	0.0434						

表 4.2: ブランド統合後の最尤推定されたパラメータ

1day		2week		1month		1day		2week		1month	
Para.	Value	Para.	Value	Para.	Value	Para.	Value	Para.	Value	Para.	Value
α_1	8.2714	α_1	4.3831	α_1	3.8316	γ_{15}	0.0000	γ_{15}	0.0000	γ_{15}	0.0001
α_2	8.5423	α_2	0.3986	α_2	-15.6204	γ_{16}	0.0061	γ_{16}	0.0000	γ_{16}	0.9999
α_3	2.2577	α_3	19.7751	α_3	19.2397	γ_{22}	0.7100	γ_{22}	0.0003	γ_{22}	1.0000
α_4	-1.0814	α_4	7.3926	α_4	-6.8555	γ_{24}	0.0001	γ_{24}	0.0000	γ_{24}	0.0000
α_5	0.6440	α_5	15.1747	α_5	14.0522	γ_{25}	0.2899	γ_{25}	0.0000	γ_{25}	0.0000
α_6	-12.3034	α_6	3.7961	α_6	17.1754	γ_{26}	0.0000	γ_{26}	0.9997	γ_{26}	0.0000
μ_1	1.0000	μ_1	0.0001	μ_1	0.0001	γ_{33}	0.0002	γ_{33}	0.2505	γ_{33}	0.0000
μ_2	0.5274	μ_2	0.0001	μ_2	0.1520	γ_{34}	0.0014	γ_{34}	0.5569	γ_{34}	0.0000
μ_3	0.9998	μ_3	1.0000	μ_3	0.0001	γ_{35}	0.9983	γ_{35}	0.0001	γ_{35}	0.6972
μ_4	0.4338	μ_4	0.9996	μ_4	0.8737	γ_{36}	0.0000	γ_{36}	0.1925	γ_{36}	0.3028
μ_5	0.0001	μ_5	1.0000	μ_5	0.0039	γ_{44}	0.9873	γ_{44}	1.0000	γ_{44}	1.0000
γ_{11}	0.9939	γ_{11}	0.9193	γ_{11}	0.0000	γ_{45}	0.0127	γ_{45}	0.0000	γ_{45}	0.0000
γ_{14}	0.0000	γ_{14}	0.0807	γ_{14}	0.0000						

4.2 ブランド統合後の結果

次に、ブランド統合後の結果について、最尤推定によって得られた各パラメータの推定値を基に、その傾向を示す。ブランド統合前後の結果を比較すると、統合が消費者のブランド選択に及ぼす影響が明確に示唆される。統合前の段階では、効用関数のパラメータに一定の傾向が見られたものの、パラメータごとにばらつきが存在し、消費者の選好が時間とともに変動していることが示唆される。特定のブランドに対する選好は一貫して正または負の傾向を示す一方で、一部のブランドでは時間経過とともに選択のされ方が変化していることが確認できる。ブランド統合後は、効用関数のパラメータが大きく変動し、消費者のブランド選好が再編されたことが確認できる。統合されたブランドの影響力が増す一方で、一部のブランドでは選択が大幅に減少しており、ブランド統合が消費者の意思決定に直接的な影響を与えたことが示唆される。また、統合によって特定の選択肢の効用が顕著に変化し、ブランド間の競争関係が大きく変容したことが確認できる。

非類似度パラメータの変化を考察すると、ブランド統合前後でブランド間の競争構造が大きく変化していることが分かる。統合前は、特定のブランド間で強い競争関係が見られたが、統合後

効用関数のパラメータに関しては、統合前と比較して大きな変動が見られる。特に、 α_1 から α_6 の値が、期間（1日、2週間、1か月）ごとに大きく異なり、選択肢の統合により消費者のブランド選好が変化していることが示唆される。例えば、 α_1 の推定値は統合後に 8.2714 (1日), 4.3831 (2週間), 3.8316 (1か月) と全体的に高い正の値を示しており、統合されたブランドに対する効用が一貫して高いことがわかる。一方で、 α_2 は1か月後に -15.6204 と極端な負の値を示し、特定の統合ブランドに対する選好が大きく低下していることが確認される。さらに、 α_6 の推定値も期間ごとに変動が激しく、1日には -12.3034, 1か月には 17.1754 と大きく変化していることから、匿名化とともに消費者の選択に大きな影響を与えていていることが示唆される。

次に、非類似度パラメータを考察すると、ブランド統合後の市場におけるブランド間の競争関係の変化が明確に示されている。統合前に比べて、 μ_1 および μ_5 の値が 1.0000 (1日) または 0.0001 (2週間, 1か月) と極端な値を示しており、ブランド間の類似性が大きく変化していることが確認できる。

アロケーションパラメータについては、ブランド統合後の選択肢の分布が変化していることが確認できる。特に、 $\gamma_{15}, \gamma_{16}, \gamma_{24}, \gamma_{25}$ の値が一部の期間で 0.0000 を示しており、特定の選択肢が完全に排除されるか、非常に低い選択確率を持つようになったことが示されている。一方で、 γ_{22} や γ_{26} などは期間によって大きな変動を示しており。また、 γ_{44} が 1.0000 を示していることから、一部の選択肢は完全に市場での優位性を持つようになったと考えられる。

4.3 ブランド統合による影響

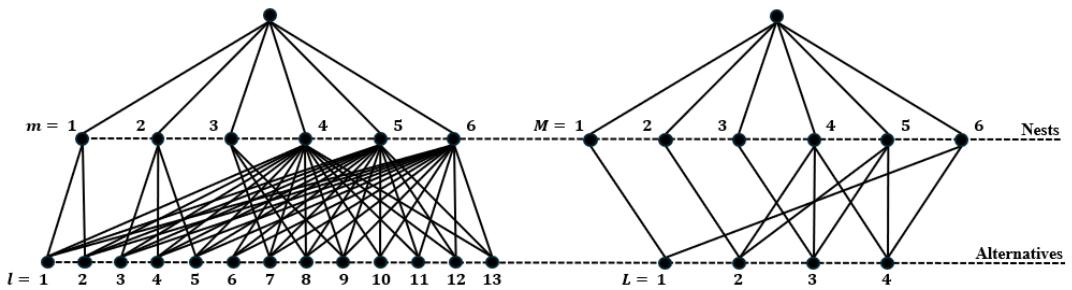


図 4.1: ブランド統合前後 (1day) での GNL 構造

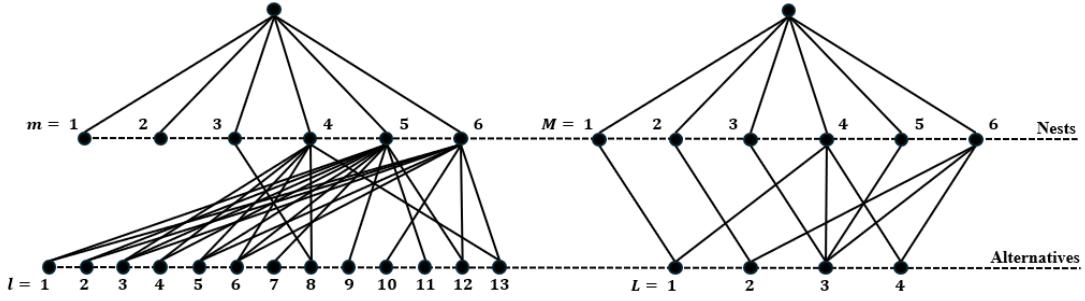


図 4.2: ブランド統合前後 (2week) での GNL 構造

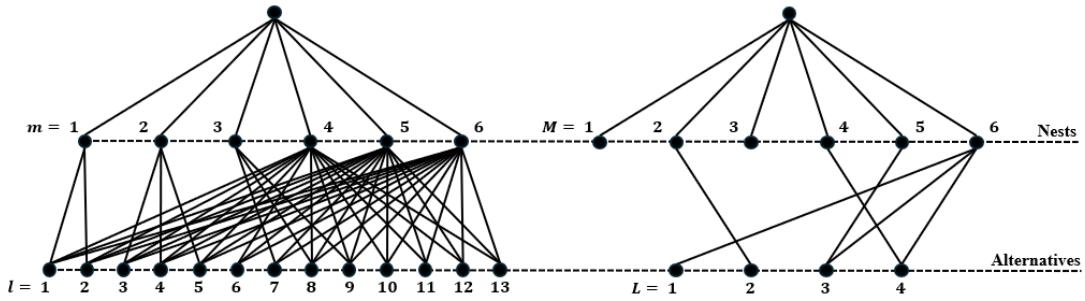


図 4.3: ブランド統合前後 (1month) での GNL 構造

は一部のブランド間の類似性が高まり、消費者の選択行動にも影響を与えていた。特に、統合後は特定のブランドが市場での競争優位性を確立する一方で、一部の選択肢間の競争が極端に弱まる場面も確認できる。これにより、統合がブランド間の競争力の再配置に寄与していることが示唆される。

アロケーションパラメータの変化についても、ブランド統合が選択肢の分布に与える影響が顕著であることが確認できる。統合前はブランドごとの分布が比較的安定していたが、統合後は特定のブランドが市場での優位性を高める一方で、完全に選択肢から排除されるブランドも存在することが示唆される。このことから、ブランド統合は単なる選択肢の統合にとどまらず、消費者の行動や競争環境の力学を大きく変化させる要因となっていることが分かる。それぞれのブランド統合前後の GNL モデル構造については、図 4.1 から図 4.3 に示している。

ブランド統合前後の比較結果から、統合が消費者の選択行動、ブランド間の競争関係、選択肢の分布に大きな影響を与えることが示唆される。統合によって特定のブランドの影響力が増し、消費者の選択に対する影響力が高まる一方で、他のブランドの選択確率が大幅に低下し、消費者の選択行動が変容する傾向にある。

4.4 選択肢を統計した場合の確定的効用の範囲

次に、選択肢を統計した場合の確定的効用の範囲を示す。ブランド統合前後のブランド選択モデルにおける確定的効用の範囲を比較することで、統合が消費者の選択行動に与える影響が明確に示唆される。統合前は、各ブランド選択肢の確定的効用に一定のばらつきが見られるものの、消費者の選好は比較的安定しており、大きな変動は生じていない。特定のブランドに対する選好が一貫して高い水準を維持している一方で、その他のブランド間では競争が均等に分散している傾向が確認できる。これは、統合前の市場環境において、ブランド選択が安定していたことを示唆している。

一方で、ブランド統合後には特定のブランドの確定的効用が顕著に変化し、消費者のブランド選好の再編が進んだことが確認できる。特に、ブランド統合直後の短期間では、消費者の選択行動に大きな変化は見られず、統合の影響は限定的である。統合後の中期においては、ブランド間の競争関係が変化し、消費者の選択肢の分布が再編成される傾向が見られ、特定のブランドへの選好が集中し始める傾向にある。さらに、長期間の分析では、統合の影響がより顕著になり、一部のブランドは消費者の選好において強く支持されるようになった一方で、競争力を失ったブランドの選択確率が大幅に低下し、事実上市場から淘汰されるブランドも見られる傾向にある。

このように、ブランド統合が消費者の選択行動に与える影響は、短期間では限定的であるが、中長期的には選択肢の競争力や市場の構造を大きく変化させることが示唆される。特定のブランドの確定的効用が上昇することにより、消費者の選択肢がより限定的になり、一部のブランドが市場での支配力を強める結果となる。確定的効用の値を安定させるための手法として、初期値を変更することや効用関数のパラメータの値を固定することなどが挙げられる。また、アロケーションパラメータの値の偏りが2週間よりも1ヶ月の方が安定しており、セール期間などによる購買機会に大きな変動が生じる可能性があるため、購入日時の統合を「3週間」とすることで、より安定した分析が可能になるかもしれない。

表 4.3: 確定的効用の範囲

DU.	1day			2week			1month		
	min	$\gamma_{Kj}\bar{V}_k$	max	min	$\gamma_{Kj}\bar{V}_k$	max	min	$\gamma_{Kj}\bar{V}_k$	max
$\gamma_{11}\bar{V}_1$	0.0106	1.5718	1.3073	0.0000	—	0.0000	0.1227	—	1.2541
$\gamma_{14}\bar{V}_1$	0.0415	—	1.1836	0.0000	—	0.0000	0.9883	—	5.5152
$\gamma_{15}\bar{V}_1$	0.0159	—	1.0839	-0.4541	—	3.5674	0.0021	—	0.4435
$\gamma_{16}\bar{V}_1$	0.0237	—	0.1618	-1.3195	—	0.4749	0.0009	0.0336	0.3800
$\gamma_{22}\bar{V}_2$	0.2067	5.2924	0.9635	0.0000	—	0.0000	0.0363	1.4272	4.6642
$\gamma_{24}\bar{V}_2$	1.3435	—	6.6563	0.0041	—	1.4704	0.0765	—	8.3673
$\gamma_{25}\bar{V}_2$	0.5903	6.0340	1.8205	2.1978	—	6.9934	0.0725	—	3.3902
$\gamma_{26}\bar{V}_2$	0.0369	—	1.5990	0.0000	1.3905	5.4182	0.0059	—	1.7267
$\gamma_{33}\bar{V}_3$	0.0779	—	1.1736	0.0000	1.3834	8.4067	0.0078	—	0.9382
$\gamma_{34}\bar{V}_3$	0.0529	—	2.3808	0.0000	-0.9163	5.0784	0.0000	—	7.2836
$\gamma_{35}\bar{V}_3$	0.3473	4.5316	0.5192	1.8775	—	2.8063	1.7701	0.3026	2.6458
$\gamma_{36}\bar{V}_3$	1.3993	—	2.0915	0.000	—	0.0000	0.0297	1.1220	0.0443
$\gamma_{44}\bar{V}_4$	0.0129	4.8405	2.7517	0.0000	2.0060	4.1496	0.0003	2.2320	4.5750
$\gamma_{45}\bar{V}_4$	0.0046	—	5.0019	0.0000	—	7.0908	0.0208	—	5.0535
$\gamma_{46}\bar{V}_4$	0.0046	—	1.2530	0.0000	—	5.0784	0.0162	—	4.3439

第5章 おわりに

5.1 結論

本研究の結果として, k 匿名化を適用することでプライバシー保護の強度が向上する一方で, データの一般化による情報損失がブランド選択モデルの精度に影響を及ぼすことが明らかとなった. 具体的には, k の値が大きくなるにつれて, データ内の個人識別性が低下し, プライバシー保護の強度が増すが, 選択確率の推定精度が低下する傾向が見られた. この影響は特に, 高次元の特徴量を含むデータセットにおいて顕著であり, 匿名化による情報の欠損がブランド選択の予測精度や識別性能の低下につながる可能性が示唆された.

また, GNL モデルにおけるパラメータ推定への影響を分析した結果, 匿名化の適用によって一部の選択構造が変化し, ブランド間の相対的な選好関係に変動が生じる可能性が確認された. 特に, 選択肢間の相関構造を考慮する GNL モデルにおいては, データの一般化による影響が単なる精度低下にとどまらず, 選択行動の解釈にも影響を与えることが示された.

5.2 今後の課題

今後の課題として, まず他の匿名化手法との比較が挙げられる. 本研究では k 匿名化を中心に検討したが, l 多様性や t 近接性など他の手法と比較し, プライバシー保護とデータ有用性のトレードオフを評価することが重要である.

また, ネスト構造の妥当性についての検討が挙げられる. ブランド属性や選択肢の類似性を基にしたネスト設計がモデルの精度に与える影響を評価し, 異なる基準でのネスト設計がモデルの適合性や予測精度をどう変化させるかを検討し, 選択モデル全体の適合性向上の模索を今後の課題とする.

付録A GNLにおける二つの選択確率定式化の等価性

ここでは、GNLにおける二つの選択確率の定式化が等価であることを示す。一般的なGNLの定式化には2つ（スケールとパラメータの導入を含めるとそれぞれ2つで計4つ）の表現がある。一方が本研究で用いているログサム表記 [GNL-LS] であり、もう一方が非類似度パラメータがネスト選択式にも現われる [GNL-O] である。これら以外にログサム表記において \hat{V}_m を省略した表記も見られるが、これは [GNL-LS] の特殊形と見なせる。ネスト選択時にログサム表記を用いない場合の定式化 [GNL-O] は次に示すとおりである：

[GNL-O]

$$P_l = \sum_m P_m P_{l|m}, \quad (\text{A.1})$$

$$P_m = \frac{\left(\sum_{l' \in \mathcal{N}_m} (\gamma_{l'm} \exp V_{l'})^{1/\mu_m} \right)^{\mu_m}}{\sum_m \left(\sum_{l' \in \mathcal{N}_m} (\gamma_{l'm} \exp V_{l'})^{1/\mu_m} \right)^{\mu_m}} \quad (\text{A.2})$$

$$P_{l|m} = \frac{(\gamma_{lm} \exp V_l)^{1/\mu_m}}{\sum_{l' \in \mathcal{N}_m} (\gamma_{l'm} \exp V_{l'})^{1/\mu_m}} \quad (\text{A.3})$$

ここで、 V_l は式 (1) ブランド h の全ての属性を含めた確定的効用である。また非類似度パラメータ、アロケーションパラメータの効用最大化と整合的であるための条件は [GNL-LS] と全て同じ（式 (8) - (10)）である。

式変形により、実際に両定式化が等価であることを示そう。基本的な導入の流れは、NL (Nested Logit Model) における場合を示している Train [6] と同様である。式 (60) より、

$$P_l = \sum_m \left(\frac{(\gamma_{lm} \exp V_l)^{1/\mu_m}}{\sum_{l' \in \mathcal{N}_m} (\gamma_{l'm} \exp V_{l'})^{1/\mu_m}} \cdot \frac{\left(\sum_{l' \in \mathcal{N}_m} (\gamma_{l'm} \exp V_{l'})^{1/\mu_m} \right)^{\mu_m}}{\sum_m \left(\sum_{l' \in \mathcal{N}_m} (\gamma_{l'm} \exp V_{l'})^{1/\mu_m} \right)^{\mu_m}} \right) \quad (\text{A.4})$$

である。ここで、式(63)の V_m に式(2)を代入し整理すると、

$$\begin{aligned}
P_l &= \sum_m \left(\frac{\left(\gamma_{lm} \exp(\hat{V}_m + \hat{V}_{ml} + \hat{V}_l) \right)^{1/\mu_m}}{\sum_{l' \in \mathcal{N}_m} \left(\gamma_{l'm} \exp(\hat{V}_m + \hat{V}_{ml'} + \hat{V}_{l'}) \right)^{1/\mu_m}} \right. \\
&\quad \cdot \left. \frac{\left(\sum_{l' \in \mathcal{N}_m} \left(\gamma_{l'm} \exp(\hat{V}_m + \hat{V}_{ml'} + \hat{V}_{l'}) \right)^{1/\mu_m} \right)^{\mu_m}}{\sum_m \left(\sum_{l' \in \mathcal{N}_m} \left(\gamma_{l'm} \exp(\hat{V}_m + \hat{V}_{ml'} + \hat{V}_{l'}) \right)^{1/\mu_m} \right)^{\mu_m}} \right) \\
&= \sum_m \left(\frac{\gamma_{lm}^{1/\mu_m} \left(\exp \hat{V}_m / \mu_m + \exp \hat{V}_l' / \mu_m \right)}{\sum_{l' \in \mathcal{N}_m} \gamma_{l'm}^{1/\mu_m} \left(\exp \hat{V}_m / \mu_m + \exp \hat{V}_{l'}' / \mu_m \right)} \right. \\
&\quad \cdot \left. \frac{\left(\sum_{l' \in \mathcal{N}_m} \gamma_{l'm}^{1/\mu_m} \left(\exp \hat{V}_m / \mu_m + \exp \hat{V}_{l'}' / \mu_m \right) \right)^{\mu_m}}{\sum_m \left(\sum_{l' \in \mathcal{N}_m} \gamma_{l'm}^{1/\mu_m} \left(\exp \hat{V}_m / \mu_m + \exp \hat{V}_{l'}' / \mu_m \right) \right)^{\mu_m}} \right) \\
&= \sum_m \left(\frac{\gamma_{lm}^{1/\mu_m} \exp \hat{V}_m / \mu_m \exp \hat{V}_{l'}' / \mu_m}{\exp \hat{V}_m / \mu_m \sum_{l' \in \mathcal{N}_m} \left(\gamma_{l'm}^{1/\mu_m} \exp \hat{V}_{l'}' / \mu_m \right)} \right. \\
&\quad \cdot \left. \frac{\exp \hat{V}_m \left(\sum_{l' \in \mathcal{N}_m} \left(\gamma_{l'm}^{1/\mu_m} \exp \hat{V}_{l'}' / \mu_m \right) \right)^{\mu_m}}{\sum_m \exp \hat{V}_m \left(\sum_{l' \in \mathcal{N}_m} \left(\gamma_{l'm}^{1/\mu_m} \exp \hat{V}_{l'}' / \mu_m \right) \right)^{\mu_m}} \right) \\
&= \sum_m \left(\frac{\gamma_{lm} \exp(\hat{V}_l')^{1/\mu_m}}{\sum_{l' \in \mathcal{N}_m} \left(\gamma_{l'm} \exp(\hat{V}_{l'}') \right)^{1/\mu_m}} \cdot \frac{\hat{V}_m + \hat{V}_m'}{\sum_m (\hat{V}_m + \hat{V}_m')} \right) \\
&= \sum_m \hat{P}_{l|m} \hat{P}_m = \hat{P}_l
\end{aligned}$$

となり、両者は等価である。最初から2番目の等号の箇所で、

$$\hat{V}_l' = \hat{V}_{ml} + \hat{V}_l \tag{A.5}$$

を用いる。また、最後から2番目の等号の箇所では、式(7)及び $\exp(x)b^c = \exp(x + clnb)$ という関係を用いている。本研究で前者の定式化を用いたのは、式展開上、明示的にネストに関連する効用と選択肢に関連する効用を区別可能な[GNL-LS]の表記が都合が良いためである。

付録B 例示に用いたデータの詳細

本研究で用いたデータは、2001年の1年間にわたり収集されたものであり、調査対象は首都圏にある同一チェーンの2店舗である。データには、合計1,801世帯の購買履歴が含まれており、全購買機会の総数は31,582回にのぼる。各購買データには、購入された商品の詳細な属性情報が含まれており、購買機会の分布については表B.1に示す。また、ブランド統合の影響を分析するため、統合前および統合後の各選択肢の詳細を表B.2に整理している。

これらのデータを基に、GNLモデルにおける集計問題の発生メカニズムを検討し、購買行動への影響を定量的に分析する。本研究で用いたデータは、2001年の1年間にわたり収集されたものであり、調査対象は首都圏にある同一チェーンの2店舗である。データには、合計1,801世帯の購買履歴が含まれており、全購買機会の総数は31,582回にのぼる。各購買データには、購入された商品の詳細な属性情報が含まれており、購買機会の分布については表B.1に示す。また、ブランド統合の影響を分析するため、統合前および統合後の各選択肢の詳細を表B.2に整理している。これらのデータを基に、GNLモデルにおける集計問題の発生メカニズムを検討し、購買行動への影響を定量的に分析する。

表 B.1: 集計問題例に用いたスキャンパネルデータ詳細

l	Brand	Volume (ml)	pasteurization or not	Purchasing opportunities
1	aeon	1,000		1,703
2	aeon	500		861
3	furuya	1,000	pasteurization	1,664
4	furuya	500	pasteurization	47
5	furuya	1,000		19,625
6	morinaga	1,000		1,458
7	yukizirushi	200		350
8	yukizirushi	500		193
9	yukizirushi	1,000		356
10	hellokitty	1,000		249
11	satsuraku	1,000		4,558
12	haruna	1,000		187
13	koiwai	1,000		331

表 B.2: ブランド統合前後の選択肢

l	Brand	L	Brand
1	aeon	1	AEON
2	aeon	1	AEON
3	furuya	2	FURUYA
4	furuya	2	FURUYA
5	furuya	2	FURUYA
6	morinaga	4	OTHER
7	yukizirushi	3	YUKIZIRUSHI
8	yukizirushi	3	YUKIZIRUSHI
9	yukizirushi	3	YUKIZIRUSHI
10	hellokitty	4	OTHER
11	satsuraku	4	OTHER
12	haruna	4	OTHER
13	koiwai	4	OTHER

謝辞

本卒業研究を無事に完了できたのは、多くの方々のご指導とご支援のおかげであり、心より感謝申し上げます。

まず、日々ご指導いただいた高橋先生に深く感謝いたします。研究テーマの選定から論文作成に至るまで、親身にご指導いただきました。特に初期段階では、適切な助言をいただき研究の方向性を定めることができました。また、専門的な助言だけでなく、私の成長を促す課題を与えてくださいり、それを乗り越える中で知識やスキルを向上させることができました。先生のご指導なくして本研究の成果は得られなかつたと確信しております。

また、研究室の仲間にも感謝いたします。日々の意見交換や協力を通じて、多くの刺激を受け、視野を広げることができました。互いに研究の進捗を共有し、課題を乗り越える中で得た経験は、大学生活における貴重な財産となりました。研究活動以外でも、共に過ごした時間はかけがえのない思い出となりました。

さらに、家族の支えなしには研究に専念することはできませんでした。特に両親には、学費や生活面での支援だけでなく、精神的な支えをいただきました。日々の何気ない会話や励ましの言葉が心の支えとなり、困難に直面した際にも家族の存在が大きな励みとなりました。改めて、家族の支えがどれほど大きな力となっていたかを実感しております。

最後に、大学生活を通じて関わったすべての方々に心より感謝申し上げます。先生方や職員の皆様、そして友人たちの支えがあったからこそ、充実した学生生活を送り、多くの学びと経験を得ることができました。これからも学び続け、支えてくださった皆様への感謝を忘れずに精進してまいります。

改めて、この場を借りて深く御礼申し上げます。

参考文献

- [1] Kristen, L., David J. D. and Raghu R.: “Mondrian Multidimensional K-Anonymity,” *22nd International Conference on Data Engineering*, (2006)
- [2] Mehmet, E. N. and Christopher C.: “Multirelational k -Anonymity,” *Transactions on Knowledge and Data Engineering*, Vol.21, No.8, pp. 1104–1117 (2009)
- [3] Latanya, S.: “ k -Anonymity: A Model for Protecting Privacy,” *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, Vol.10, pp. 571–588 (2002)
- [4] Wen, C. and Koppelman, F. S.: “The Generalized Nested Logit Model,” *Transp. Res. Part B*, Vol.35, No. 7, pp. 627–641 (2001)
- [5] Sweet, R. J.: “An Aggregate Measure of Travel Utilit,” *Transp. Res. Part B*, Vol.31, No. 5, pp. 403–416 (1997)
- [6] Ivanova, O.: “A Note on the Consistent Aggregation of Nested Logit Demand Functions,” *Transp. Res. Part B*, Vol.39, No. 10, pp.890–895 (2005)
- [7] 高橋啓: “GNLにおける集計問題を回避するための集計ルールの導出,” 日本経営工学科会論文誌, Vol.10, pp. 63–75 (2012)
- [8] Train, K. E.: Discrete Choice Methods with Simulation Second Edition, CambTidge University Press, NY (2009)