離散選択モデル 補足資料1: 理論編

講師: 遠山祐太

最終更新: 2025-01-06

イントロダクション

補足資料

- 本スライドでは離散選択モデルに関する理論的補足をカバーする。
- 内容的に、学部上級ー大学院レベルのものも含むため、必ずしも理解する必要はない。
- 上級の内容に興味がある受講生向け。

構造推定アプローチとは？

構造推定(Structural Estimation)とは？

- 構造推定: 経済主体の意思決定モデルを活用した実証分析のアプローチ
- 手順
  - Step 1: 分析したい経済現象に関する経済モデルを構築する。
  - Step 2: 経済モデル内のパラメタを、データを用いて推定する。
  - Step 3: 推定したモデルを用いて、**反実仮想 (counterfactual)**シミュレーションを行う。

構造推定の例：価格付け

- 問：企業が自社製品の価格付けを検討している。どうすればよいか？
- Step 1：以下の利潤最大化問題を考えよう。

$$\max_{p} \ (p - mc)D(p)$$

  - p：価格、 mc: 限界費用(＝単位費用), $D(p)$: 需要関数
- Step 2: 需要関数 $D(p)$ と 限界費用 mc を知りたい。
  - 線形な需要関数を考えて、(操作変数法などで)推定する

$$Q_t = \alpha_0 + \alpha_1 \cdot P_t + \beta \cdot income_t + \epsilon_t$$

  - 限界費用は(1) 企業の内部情報、(2) 計量経済学的に推定する (詳細は割愛)
- Step 3: 利潤を最大にするような価格を探す！（数値計算）

価格付け分析の実例

- 例１：Ziprecruiter.com の価格付け実験
  - 米国のオンライン求人サイトが、求人企業の月会費の改定を検討。
  - シカゴ大学の研究者と共同して需要測定のためのA/Bテストを実行。
  - 分析から、価格の最適化により収入を約６０%改善できることが判明。
  - 実際に、価格を９９ドルから２４９ドルへ改定。
- 例２：バッファローのnasneの価格付け
  - ネットワークレコーダーの新製品に関する価格付けの検討。
  - UTEconが、販売予測モデルの構築に携わる。
  - 販売価格29800円が「スイートスポット」にはまり、売れ行きは好調だった。
- 両事例の詳細についてはこちらを参照。

構造推定の例：企業合併分析

- 問：自動車市場において、トヨタとホンダが合併を検討している。この合併を許すべきか？
- 論点：
  - (-) 寡占化の進行。競争がなくなって、価格が上がるかも。
  - (+) 新企業の効率化。規模の経済、ノウハウの共有など。
- 重要性：大型の企業合併は公正取引委員会により事前審査が求められる。

合併シミュレーション

- 構造推定による分析ステップ：
  1. 企業間の競争モデルを構築する（例：ベルトラン競争）
  2. モデルの要素（需要関数・費用関数）をデータから推定する。
  3. 仮に合併が起きた場合にどのような価格付けになるかをシミュレーションする。

- **合併シミュレーション**と呼ばれ、独禁法実務においても活用されている。
  - 例：公取委「平成２８年度における主要な企業結合事例」 の「出光興産㈱による昭和シェル石油㈱の株式取得及びＪＸホールディングス㈱による 東燃ゼネラル石油㈱の株式取得 」

- 学術研究の事例はこちらの補足資料
  - Fan (2013)
  - Wollmann (2018)

その他の例 (私の研究関連)

- 環境政策：排出権取引の政策ルールを変更したら、どのような均衡が実現するか？
  - 参考：こちらにFowlie, Reguant, and Ryan (2016)の解説記事
- 政治経済学: 「もし全員が投票」したら、選挙結果はどのように変化するか？
  - 投票行動モデルを構築し、選好と投票コストを推定。そのうえでシミュレーション
  - Kawai, Toyama, and Watanabe (2021) を参照。

構造推定の強み・弱み

- 強み：経済モデルを用いたシミュレーションができる。
  - 例：ある政策の代わりに別の政策が行われたらどうなっていたか？
  - 例：意思決定におけるあるメカニズム・チャネルを捨象した場合、どうなるか？
  - 因果効果・政策効果の推定の一手段
  - 経済的に関心があるアウトカム（特に経済厚生）を見ることができる。
- 弱み：経済モデルに強く依拠している。
  - 仮定が強い。
  - 分析が複雑になることが多い（プログラミングなど）
- 次節にて構造推定アプローチと誘導形アプローチの議論についてもう少し詳しく説明。

いわゆる「構造推定 VS 誘導形」について

いわゆる「誘導系」対「構造推定」の議論

- 「誘導系」アプローチ：ミクロ実証研究における主流
  - データの良いVariationを見つける （自然実験、操作変数法、回帰不連続デザイン）
  - 回帰分析フレームワークで、因果効果(トリートメント効果)を推定
  - 統計学でのCausal inference methodに依拠
  - 自然実験アプローチとも呼ばれる。

- 「誘導系」・「構造推定」アプローチは1990年代半ばくらいから広がっていた。

- それと同時に、これら手法の間の対立も発生。

対立の概略 (個人的見解を多分に含む)

- 「誘導系」から「構造推定」への批判
  - 経済モデルを明示的に定式化している。仮定がとても強い。
  - モデルパラメタの推定を、どのようにやっているか不明瞭
  - シミュレーションによる結果がブラックボックス。
- 「構造推定」から「誘導系」への批判
  - 背後のモデルがないので、推定している因果効果パラメタの解釈が不明瞭
  - 均衡効果を考慮できない (いわゆるSUTVAの仮定)
  - 「自然実験」がある「理想的」なデータの状況の分析ばかりになるのではないか？

個人的見解：これからの時代は両方必要！！

- 「誘導系」と「構造推定」は代替的ではなく、**補完的**
- 「誘導系」でも、単に因果効果を推定するだけではなく、**背後のメカニズム**の説明が必要
  - 経済モデルの重要性
- 「構造推定」では「自然実験」をうまく活用してパラメタを推定する。
  - RCTと構造推定を合わせる研究も増えている。
  - 例１：Todd and Wolpin (2006)
  - 例２：Ito, Ida, and Tanaka (2016)
  - 例３：Kawaguchi, Uetake, and Watanabe (2021)

手法論争に関する関連文献

- "SYMPOSIUM: CON OUT OF ECONOMICS" appeared in Spring 2010 volume of Journal of Economic Perspectives
  - Angrist and Pischke による、いわゆる構造推定アプローチ（特にIOやマクロ）への批判。
  - それに対する反論 (Keane, Sims, Nevo, Whinston)
- "Structural vs. atheoretic approaches to econometrics"
  - Michael Keane (労働経済学者) による、「自然実験アプローチ」への批判
- "Structural vs. Reduced Form:" Language and Models in Empirical Economics"
  - Phil Haile (IO)による論考。
  - そもそも「誘導系 VS 構造推定」という呼び方や分け方が不適切ではないか？という議論。
  - 多少読みづらいかもしれないが、**個人的に強くオススメ**

パラメタの識別

離散選択モデルにおけるパラメタの識別

- 識別: モデルパラメタがデータから一意に特定できること。
- 離散選択モデルにおいて選択確率は**効用の差**のみに依存している。

$$P(d_i = j) = \Pr\left(\{\epsilon_{ij}\}_{j=1}^{J} : V_{ij} + \epsilon_{ij} \geq V_{ik} + \epsilon_{ik}\ \forall k \neq j\right)$$

- 効用の定式化において2種類の**正規化(normalization)**が必要となる。
  - 1: location : 定数を足しても同じ (i.e., $U_{ij} + \alpha$ for all j)
  - 2: scale : U_{ij} について定数倍しても同じ (i.e., αU_{ij} for all j for some $\alpha > 0$)

Location Normalization

- 新しい定式化として $\tilde{U}_{ij} = U_{ij} + \alpha$ を考える。
- すると、

$$
\begin{aligned}
Prob(\tilde{U}_{ij} > \tilde{U}_{ik}, \forall k) &= Prob(U_{ij} + \alpha > U_{ik} + \alpha, \forall k) \\
&= Prob(U_{ij} > U_{ik}, \forall k)
\end{aligned}
$$

- 含意: ある選択肢から得られる効用の水準を正規化する必要がある。
- 実証分析では**アウトサイドオプション($j = 0$)**の効用をゼロと置くことが多い。

Scale Normalization

- 新しい定式化として $\tilde{U}_{ij} = \lambda U_{ij}$ を考える。ただし $\lambda > 0$.
- すると

$$\begin{aligned} Prob(\tilde{U}_{ij} > \tilde{U}_{ik}, \forall k) = & \quad Prob(\lambda U_{ij} > \lambda U_{ik}, \forall k) \\ = & \quad Prob(U_{ij} > U_{ik}, \forall k) \end{aligned}$$

- 含意: 選考ショックの分散について基準化する必要がある。
  - ロジットモデルでは $Var(\epsilon_{ij}) = \pi^2/6$ としている。
- 別の例：二項プロビットモデル $U_{i1} = \beta x_i + \epsilon_{i1}, U_{i0} = 0$ with $\epsilon_{i1} \sim N(0, \sigma^2)$.

$$Prob(d_i = 1) = \Phi\left(\frac{-\beta x_i}{\sigma}\right)$$

σ と β は別々に識別されない。

ランダム係数ロジットモデルの詳細

ランダム係数ロジットモデル

- 前のモデルにおいて、消費者間の異質性は選考ショック $\epsilon_{i,j}$ のみに反映。
- 消費者の異質性の入れ方
  - 観察される属性(社会経済属性) を入れる。
  - 製品属性の選好に関して、観察されない異質性を考える -> **ランダム係数**
- 定式化

$$u_{ij} = \beta_i X_j + \epsilon_{ij}$$

  - ランダム係数 $\beta_i \sim f(\beta)$.
  - X_j: 製品属性のベクトル
  - ϵ_{ij} は以前と同様。

選択確率

- ランダム係数 β_i を所与とすると、選択確率は

$$Pr(d_i = j|\beta_i) = \frac{\exp\left(\beta_i X_j\right)}{\sum_{j=1}^{J} \exp\left(\beta_i X_j\right)}$$

- しかし分析者は β_i を観察できない。従って、 β_i について平均をとったものを考える。

$$Pr(d_i = j) = \int \frac{\exp\left(\beta_i X_j\right)}{\sum_{j=1}^{J} \exp\left(\beta_i X_j\right)} f(\beta_i) d\beta_i$$

推定における問題点

- 単純化のために一変数を考える。ランダム係数の分布を以下と考える。 $\beta \sim N(\mu, \sigma^2)$.
- 目標：ランダム係数の分布パラメタ (μ, σ^2).
- 個人の選択確率は

$$Pr(d_i = j) = \int \underbrace{\frac{\exp\big((\mu + \sigma\nu_i)X_{ij}\big)}{\sum_{j=1}^{J} \exp\big((\mu + \sigma\nu_i)X_{ij}\big)}}_{Pr(d_i=j|\nu_i,\theta)} dG(\nu_i)$$

ここで $G(\nu_i)$ は標準正規分布 -> $(\mu + \sigma\nu_i) \sim N(\mu, \sigma^2)$

- 問題点：積分があるため、この選択確率を解析的な形で求めることができない。

解決策：シミュレーションによる近似

- 選択確率における積分をモンテカルロ積分する。
- Step 1: 乱数 ν を $G(\nu)$ から生成する。 ν^r とラベル付する。.
- Step 2: 乱数 ν^r に基づいてロジット確率を計算する。

$$Pr(d_i = j|\nu^r) = \frac{\exp\big((\mu + \sigma\nu^r)X_j\big)}{\sum_{j=1}^{J} \exp\big((\mu + \sigma\nu^r)X_j\big)}$$

- Step 3: この手順を R 回繰り返し、その平均を取る。

$$\hat{Pr}(d_i = j|\theta) = \frac{1}{R}\sum_{r=1}^{R} Pr(d_i = j|\nu^r, \theta)$$

Simulated Maximum Likelihood Estimator

- シミュレーションした対数尤度の最適化を行う。

$$SLL(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log \hat{P}r(d_i = j)$$

- Simulated MLEと呼ばれる。詳細はTrain Chater 10 を参照。
- シミュレーションの方法として、モンテカルロ積分以外にもいくつか方法がある。
  - Importance sampling
  - Halton draw
  - Gaussian quadrature

きのこ・たけのこ事例の尤度関数

- 参考：Train Chapter 6.7 "Panel Data"
- パラメタ θ をもつ個人 i が、選択 y_{ik} を行う確率を $P_{i,k}\left(j = y_{i,k} | \theta\right)$ とする。
- 各個人は 5 回選択を行うが、その選択を通じて選好パラメタ θ は共通とする。
- パラメタ θ をもつ個人 i がデータで観察される一連の選択 $\{y_{i,k}\}_{k=1}^{5}$ を行う確率は

$$L_i(\theta | \{y_{i,k}\}_{k=1}^{5}) = \prod_{k=1}^{5} P_{i,k}\left(j = y_{i,k} | \theta\right)$$

尤度関数続き

- パラメタ自体は観察されないので、分布で積分を取る必要がある。

$$P_i(\Omega) = \int L_i(\theta | \{y_{i,k}\}_{k=1}^{5}) f(\theta | \Omega) d\theta$$

  - ここで $f(\theta|\Omega)$ はパラメタの分布。
  - Ω は分布のパラメタ。 $\Omega = (\theta_{Kino}, \sigma^2_{Kino}, \theta_{Take}, \sigma^2_{Take}, \theta_\beta, \sigma^2_\beta)$
  - モンテカルロ積分などで近似を行う必要がある。

- 尤度関数は

$$L(\Omega) = \prod_{i=1}^{N} P_i(\Omega)$$