大学院講義　2025年度前期　交通経済学

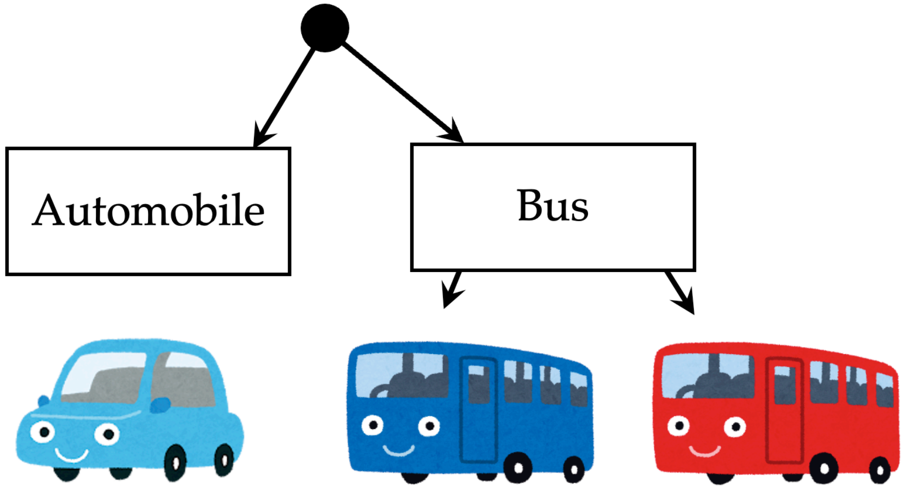
推定・評価・予測・応用

モデルをデータに当てはめる技法

大澤 実（経済研究所）

前回の振り返り

- G 関数にもとづく **MEV** モデルが MNL 系モデルを統合
- **Nested Logit** (NL) モデルは **選択の階層構造** を明示的に表現
    - 多項ロジット (MNL) における赤バス・青バス問題に対処
    - **包括的価値**：下位の選択の「価値」を上位の選択に伝達



目標

再び MNL を題材にして：

1. 離散選択モデルの推定技法を理解する
2. 推定結果の解釈・政策介入前後の厚生評価を理解する
3. データを用いて実際にモデルの推定と予測・評価を実行する（次回）

何もないところにモデルという計測装置を導入する営みを体験しよう.

離散選択モデルの推定と評価

MNLを題材に基本概念を学ぶ

多項ロジットモデル (MNL)

- ランダム効用：各選択肢の特徴ベクトルを x_i として，線形効用では

$$\begin{aligned} U_{ni} = V_{ni}(\beta) + \varepsilon_{ni} &= \alpha_i + \beta \cdot x_{ni} + \varepsilon_{ni} \\ &= \alpha_i + \beta_1 x_{ni,1} + \beta_2 x_{ni,2} + \cdots + \beta_K x_{ni,K} + \varepsilon_{ni} \end{aligned}$$

  ただし x_{ni} は説明変数ベクトル，　$\varepsilon_{ni} \sim$ i.i.d. Gumbel

- 選択確率：$P_{ni} = \dfrac{\exp(\mu V_{ni}(\beta))}{\sum_{j \in A} \exp(\mu V_{nj}(\beta))}$
- **良さ**：closed-form の選択確率による計算の容易さ
- **限界**：IIA特性，個人の異質性表現の限界
  ※ 計測可能な異質性は導入可能（e.g., 所得）

※ 統計学の基本用語の一般的定義のおさらい

- **母集団** (population)：観察・分析対象全体（e.g., ある都市の通勤者）
- **標本** (sample)：母集団から抽出されたデータ（e.g., PT調査結果）
- **抽出** (sampling)：標本を母集団から選び出す手続き．理論上の基本的前提は **無作為抽出** (random sampling).
- **統計量** (statistic)：標本から計算される量（e.g., 平均通勤時間などの記述統計量，推定されたモデルパラメタなどの推定量）
- **推定** (estimation)：母集団に対して仮定された構造のもとで，観測データからパラメータの情報を抽出する作業

交通需要分析における推定 (Estimation)

- 離散選択モデルのパラメータをデータに基づいて数値的に求める手続き
  - 前提の例：「母集団は MNL という確率モデルに従っている」
- 推定のための基準には原理的には色々ある.
  - 最小二乗法，Bayes 推定，一般化モーメント法 (GMM)，……
- MNL では通常 **最尤推定** を用いる
  - 良さ：一致性 (consistency)・漸近正規性 (asymptotic normality)
  - 標準誤差 (standard error) が推定される
    - 推定されたパラメータの「不確かさ」を表す指標.

非集計選択モデルの最尤推定 1/2

- 観測された **選択データ** (choice data)： $y = (y_n)_{n \in N} = (y_{ni})_{n \in N, i \in A}$
  - y_n は個人 $n \in N$ の選択を表現：$y_{ni} \in \{0, 1\}$, $\sum_{i \in A} y_{ni} = 1$
- 個人の選択は独立で，$i \in A$ を選ぶ確率は $P_{ni}(\theta)$ とする．
  - θ はモデルのパラメータ・推定すべき対象（例：$\theta = (\alpha, \beta, \mu)$）
- このモデルのもとで θ **の値を固定したとき** y が起こる確率は

$$L(\theta \mid y) \equiv \Pr(y \mid \theta) = \prod_{n \in N} \prod_{i \in A} P_{ni}(\theta)^{y_{ni}}$$

- L を **尤度** (likelihood) と呼ぶ．y はいま固定なので L は θ **の関数**

非集計選択モデルの最尤推定 2/2

- L の対数をとると

$$\ell(\theta \mid y) \equiv \log L(\theta \mid y) = \sum_{i \in N} \sum_{i \in A} y_{ni} \log P_{ni}(\theta)$$

- ℓ を **対数尤度** (log likelihood) と呼ぶ
- **最尤推定 (MLE)**：ℓ を最大化する $\theta \in \Theta$ を選ぶ：$\theta^* = \arg\max_{\theta \in \Theta} \ell(\theta)$
  - MLE = Maximum Likelihood Estimation
- 推定されたモデル $P_{ni}(\theta^*)$ を予測に用いる

識別 (Identification)

- **識別の問題**：理論的にパラメータが **一意に** 定まるか？　参考記事
- **定義**：パラメータ θ の統計モデル $\mathcal{P} \equiv \{P_\theta \mid \theta \in \Theta\}$ について，以下の条件を満足するとき **識別可能** (identifiable) という：

$$P_{\theta_1} = P_{\theta_2} \quad \Rightarrow \quad \theta_1 = \theta_2 \quad \forall \theta_1, \theta_2 \in \Theta.$$

※ θ はパラメータ全てを並べたベクトル．Θ は取りうる全ての値．

  - 満足しないとき，識別不能 (unidentifiable / non-identifiable) という

- 識別不能なモデルでは **データがいくらあっても** 原理的に推定不可能
$\rightarrow$ 識別可能性の検討は必須

例：MNL における識別可能性

- **定数項** α_i の識別：相対値のみ識別可能
  - 効用の相対値が選択確率 $\{P_{ni}\}$ を決定 $\Rightarrow$ 尤度 ℓ も決定
  - 1 つを基準にすれば識別可能
- **スケール** μ の識別：不可能

$$\mu V_i(\beta) = \mu\alpha_i + \mu\beta \cdot x_{ni}$$

  - 必ず μ と α, β が掛け合わさった形
    $\Rightarrow$ μ を 2倍して α, β を 1/2 倍しても選択確率は変わらない
  - $\mu = 1$ **に基準化すれば識別可能**

例：MNL における識別可能性と対数尤度関数

- $\theta = (\mu, \beta)$とし，θ についての定数部分を無視すると

$$\begin{aligned}\ell(\theta) &= \sum_{i \in A} y_i \log \frac{\exp(\mu V_i(\beta))}{\sum_{j \in A} \exp(\mu V_j(\beta))} \\ &= \mu \sum_{i \in A} y_i V_i(\beta) - N \log \sum_{j \in A} \exp(\mu V_j(\beta))\end{aligned}$$

- μ 固定 $\Rightarrow$ $\ell(\theta)$ は狭義凹関数 $\Rightarrow$ $\max_{\theta \in \Theta} \ell(\theta)$ の解が存在すれば一意
- μ を固定しないと $\ell(\theta)$ は **狭義凹関数** ではない ✍️
  - μ によるスケーリングに対しては解が一意に定まらない

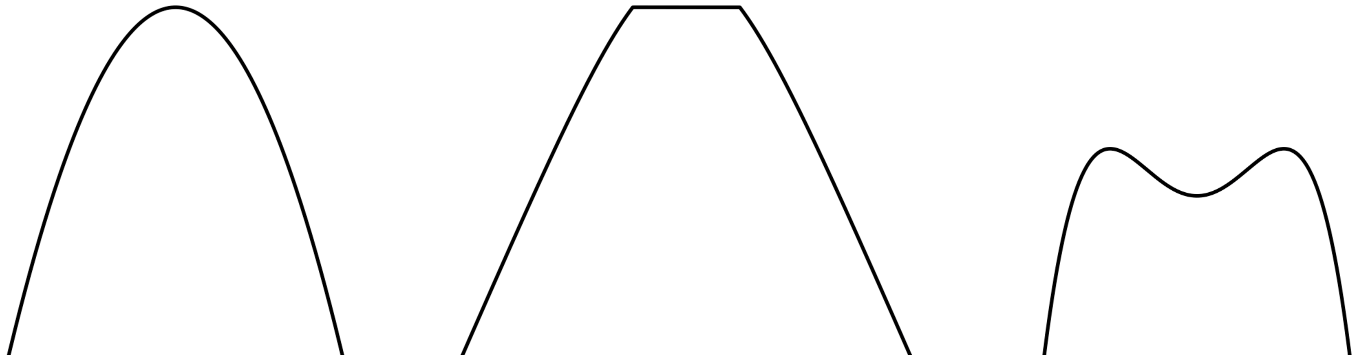
※ 関数の形状と最大化問題の解の一意性

- **定義**：関数 f が定義域 S で **凸 (convex)** であるとは

$$\alpha f(x) + (1-\alpha) f(y) \geq f(\alpha x + (1-\alpha) y) \qquad \forall \alpha \in [0,1], \forall x, y \in S$$

  - 狭義凸 (strictly convex) は等号を外す． **凹 (concave)** は逆の符号．

- 1 変数の狭義凹関数・凹関数・非凸関数の例：



- これらの関数をなるべく大きく（**最大化**）するような点は？🤔

モデル推定結果の評価

パラメータの統計的有意性 1/2

- 推定されたパラメータ $\hat{\theta}$ をどの程度信頼できるのか？
  $\rightarrow$ **統計的有意性** (statistical significance) の問題
- データ $y =$ 抽出によって得られる確率変数（と捉える）
  $\Rightarrow$ MLE によって推定されるパラメータ $\hat{\theta}(y)$ もまた確率変数！
- β_{fare} の推定値が -0.5 で，推定ごとのバラツキが 1.0 だったとする．
  - 場合によっては β_{fare} が正になるかもしれない😲　そもそも $\beta_{\text{fare}} = 0$ と区別できない．このような推定値は信頼不可（t **検定** ✍️）
    $\Rightarrow$ 推定値のバラツキ = **標準誤差** (standard error) を評価する必要
- $\hat{\theta}$ が従う確率分布は何か？

パラメータの統計的有意性 2/2

- **定理（MLEの漸近正規性）**：適切な正則性 (regularity) 条件のもとで，サンプルサイズ $|N|$ が大きいとき，MLE 推定量は正規分布する：

$$\hat{\theta} \dot{\sim} \mathcal{N}\left(\theta, \frac{1}{|N|}\mathcal{I}_\theta^{-1}\right) \quad \text{ただし } \mathcal{I}_\theta = \mathbb{E}_\theta\left[\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^\top\right].$$

  - この $\mathcal{I}_\theta$ を **Fisher 情報行列** (Fisher information matrix) と呼ぶ
  - 数値的には，推定値 $\hat{\theta}$ における対数尤度関数のヘッセ行列 $=$ **観測情報量行列** (observed information matrix) で近似する
- すなわち，$\hat{\theta}$ の共分散行列の推定値は $\widehat{\mathrm{Var}}(\hat{\theta}) = \frac{1}{|N|}\left[-\frac{\partial^2}{\partial \theta^2}\log L(\hat{\theta})\right]^{-1}$
- パッケージが勝手にやってくれるので心配しないでよい😊

コメント：定数項について

- 選択肢 $i \in A$ のランダム効用

$$U_{ni} = V_{ni}(\beta) + \varepsilon_{ni} = \alpha_i + \beta \cdot x_{ni} + \varepsilon_{ni} \qquad (\alpha_1 = 0)$$

- 定数 α_i は選択肢ごとのランダム効用の期待値とみてもよい．
  - 慣習的に ASC (Alternative Specific Constant) と呼ばれる
  - 🤢 データ固有の事情の影響？（ある選択肢が偶然多く選ばれた）
  - 🤢 新しい選択肢の追加には対応しづらい（α の値はどう決める？）
  - 🤢 「現況再現性」(in-sample fit) を「向上」 $\rightarrow$ 汎化性能 ❌
- 定数項の影響があまりに大きいようなら移転性が低い可能性

適合度の検討：①尤度比検定 1/3

- **帰無モデル** (null model) と提案するモデルとを比較する.
  - 複雑なフルモデルが当てはまりを改善しているかを判定
- 帰無モデルの **例**：定数項 $\alpha = (\alpha_i)_{i \in A}$ のみの MNL ($\beta = 0$)
- **最大化された** 尤度の比が十分大きくなるか？

$$\frac{L(\theta^*)}{L(\theta^*_{\text{null}})} > c$$

  - $c > 1$ は大きければ大きいほど提案モデルの当てはまりがよさそう.
- このアイデアに基づくのが **尤度比検定** (likelihood ratio test)

適合度の検討：①尤度比検定 2/3

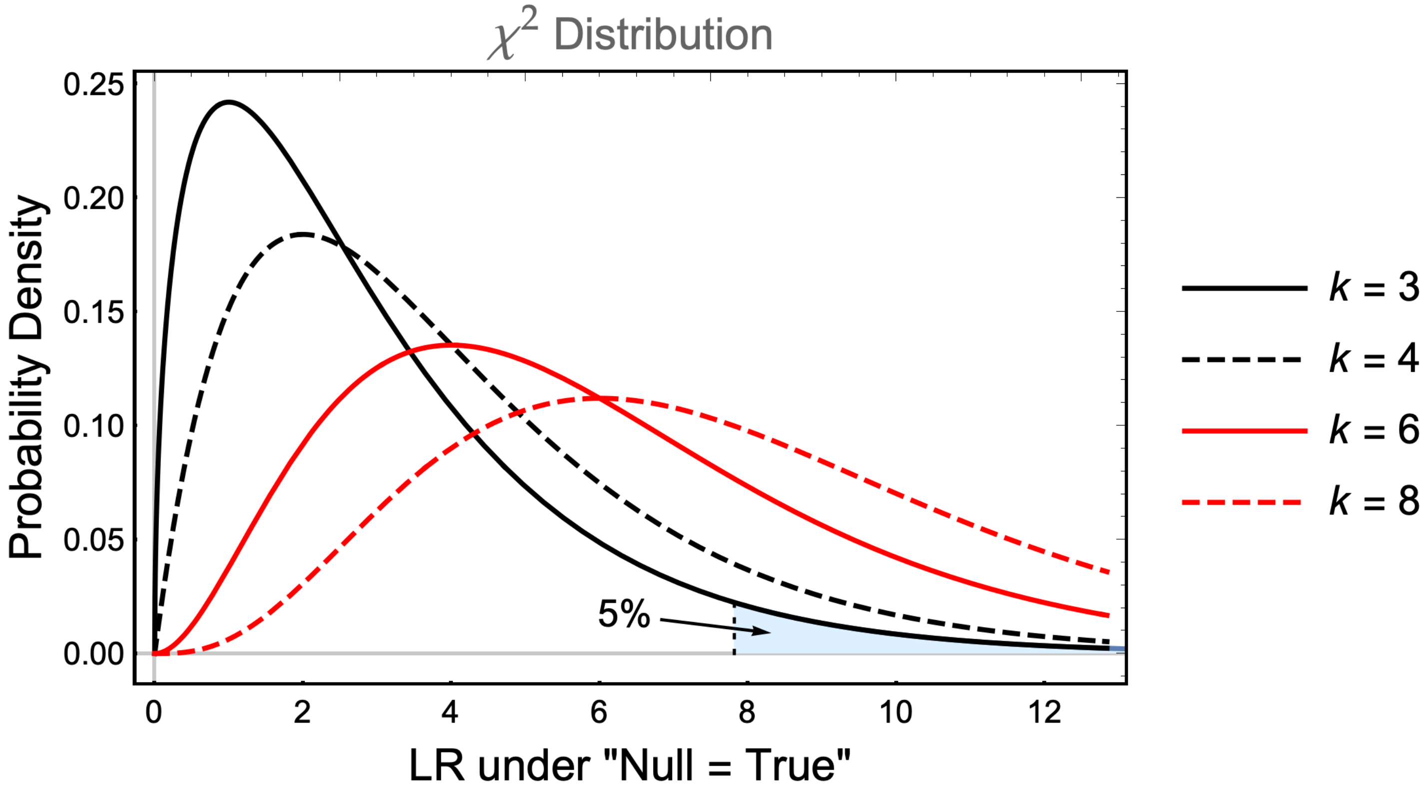
- k を帰無モデルから追加された **自由度** とする．
  - 定数のみ MNL の場合，k は β の次元 $=$ 説明変数 x_{ni} の次元
- **Wilks の定理**：帰無モデルが正しいなら，標本サイズが大きいとき

$$\mathrm{LR} \equiv -2\log\frac{L(\theta^*_{\mathrm{null}})}{L(\theta^*)} = 2\log\frac{L(\theta^*)}{L(\theta^*_{\mathrm{null}})} = 2\big(\ell(\theta^*) - \ell(\theta^*_{\mathrm{null}})\big)$$

  は自由度 k のカイ二乗分布 χ^2_k に従う．LR を **尤度比統計量** と呼ぶ．
- χ^2_k で起こりづらい LR が実現 $\Rightarrow$ 帰無モデルを **棄却** (reject)

適合度の検討：①尤度比検定 3/3

- χ^2_k の確率密度の形状



- $k = 3$ なら，例えば $\mathrm{LR} \approx 8.7$ はたかだか確率 5% の事象の発生

適合度の検討：②McFaddenの擬似 (pseudo-) R^2

- 決定係数 R^2：$[0, 1]$ に値をとり 1 に近いほど相関が強い

  👴 .｡oO（似たような指標がほしいなぁ……）(McFadden, 1974)

- 尤度比に基づく適合度指標 (Likelihood Ratio Index)

$$\rho^2 \equiv 1 - \frac{\ell(\theta^*)}{\ell(\theta^*_{\mathrm{null}})} \in [0, 1]$$

  - $\rho^2 = 0$, $\rho^2 = 1$ となるのはどのような時か？🤔

- 純粋理論的な上限は1だが実際にはない．0.2〜0.4 程度で十分に高い

汎化性能 (Generalization Performance)

- 推定したモデルが **新規データ** y' **でも有効か？** $\leftrightarrow$ **過学習** (overfitting)
  - 「よく当たるが当たるのは推定に使用したデータだけ」では困る
- モデルの妥当性評価の切り口：
  in-sample, out-of-sample, out-of-domain
- ① **再現性** (reproducibility)：**同じ** 母集団からの新たな標本 (sample) に対してモデルの性能が維持されるか
- ② **移転性** (transferability)：**異なる** が類似性のある母集団（e.g., 他都市，他時点）や **異なる** 取得手段のデータでも予測能力があるか

汎化性能の評価：①再現性

手法	内容	評価指標の例
交差検証 (k-fold cross-validation)	データを k 個に分割して k 回訓練・評価	平均対数尤度, 平均誤差
Bootstrap 評価	データから擬似的な標本抽出を繰り返し推定結果の分布を構築 $\rightarrow$ 頑健性評価	標準誤差, 信頼区間, バイアス
情報量規準 (information criterion)	対数尤度とモデルの複雑さから汎化誤差 (generalization error) を近似	AIC, BIC, etc.

- 複数のモデルを比較する前提
- AIC = Akaike Information Criterion（赤池情報量規準）

適合度の検討：②'自由度調整済み擬似 R^2

- 尤度比指数 ρ^2 は説明変数を増やすと増加しやすい（過学習の問題）
- モデルの複雑化に対するペナルティを付与：

$$\bar{\rho}^2 \equiv 1 - \frac{\ell(\theta^*) - \boldsymbol{k}}{\ell(\theta^*_{\mathrm{null}})} \in [0, 1]$$

  - AIC とのハイブリッド指標 (Ben-Akiba & Lerman, 1985, p.167)

※ 背景テーマは **Bias–Variance tradeoff**：単純なモデルほどバイアスが大きい（誤る）が，複雑なモデルほど新しいデータに対して不安定.

汎化性能の評価：②移転性

手法	例
ドメイン間検証	都市Aで推定 → 都市Bで検証
RP/SP間交差予測	RPで推定 → SPで検証
時系列比較	年度Aで推定 → 年度Bを検証

- 指標例：対数尤度，予測精度，弾力性，選択確率，変化率誤差, etc.

※ **顕示選好** (Revealed Preference; RP) データは実際に行われた選択，**表明選好** (Stated Preference; SP) データはアンケート・実験等により申告された選択．様々な理由により人々はアンケート通りに動かない．

モデル推定結果の活用

限界代替率 (Marginal Rate of Substitution)

- 推定されたパラメータ同士の **比** が重要
- 効用を変化させることなく説明変数の値を変えられる **限界代替率**

$$V_i = -\beta^*_{\text{time}}\text{Time}_i - \beta^*_{\text{cost}}\text{Cost}_i + \beta^*_{\text{safe}}\text{Safety}_i + \beta^*_{\text{conf}}\text{Comfort}_i$$

とすると，効用を維持するような変化は例えば

$$0 = -\beta^*_{\text{cost}}\Delta\text{Cost}_i - \beta^*_{\text{time}}\Delta\text{Time}_i \quad \Rightarrow \quad \frac{\Delta\text{Time}_i}{\Delta\text{Cost}_i} = -\frac{\beta^*_{\text{cost}}}{\beta^*_{\text{time}}}$$

$$0 = \beta^*_{\text{safe}}\Delta\text{Safety}_i + \beta^*_{\text{comf}}\Delta\text{Comfort}_i \quad \Rightarrow \quad \frac{\Delta\text{Comfort}_i}{\Delta\text{Safety}_i} = -\frac{\beta^*_{\text{safe}}}{\beta^*_{\text{comf}}}$$

旅行時間短縮価値 (Value of Travel Time Saving)

- あるモードの旅行時間の短縮にどれだけの **金銭的価値** を置くか
  - **交通時間価値** (Value-of-Time; VoT) とも呼ぶ
- 交通プロジェクト評価で用いる最も重要な基礎パラメータの一つ
- 例：以下のような交通モード $i \in A$ の効用関数を考える：

$$V_{ni} = \beta_1 \frac{c_{in}}{w_n} + \beta_2 T_{ni} + \beta_3 T_{ni}^2$$

  - w_n：個人 $n \in N$ の所得，c_{in}：料金，T_{in}：所要時間
- ✍️ 個人 n の選択肢 i に対する VoT v_{ni} を求めよ．

旅行時間短縮価値：例

- VoT は **個人・モード**に依存：

$$v_{ni} = -\left(\frac{\mathrm{d}c_{ni}}{\mathrm{d}T_{ni}}\right)_{\text{at } V_{ni}} = \frac{\partial V_{ni}/\partial T_{ni}}{\partial V_{ni}/\partial c_{ni}} = \frac{\beta_2 + 2\beta_3 T_{ni}}{\beta_1} w_n$$

  - 所得 w_n が大きいほど，所要時間 T_{ni} が大きいほど大きい
  - あくまで (w_n, T_{in})という値の近傍における時間価値

- $T_{ni} = T^0 \to T_{ni} = T^1$ のような変化の価値は

$$\int_{T^0}^{T^1} v_{ni}(T_{ni})\mathrm{d}T_{ni} = \frac{\beta_2 + 2\beta_3 \bar{T}}{\beta_1} w_n \cdot \Delta T \qquad \textbf{with} \begin{cases} \bar{T} = (T^0 + T^1)/2 \\ \Delta T = T^1 - T^0 \end{cases}$$

選択確率 P_{ni} の弾力性 (Elasticity)

- **自己弾力性** (direct elasticity)：選択肢 i の特性 k の変化率に対して

$$e_{ik}^{ni} \equiv \frac{x_{ni,k}}{P_{ni}} \cdot \frac{\partial P_{ni}}{\partial x_{ni,k}} = \beta_k (1 - P_{ni}) x_{ni,k}$$

- **交差弾力性** (cross elasticity)：選択肢 j の特性 k の変化率に対して

$$e_{jk}^{ni} \equiv \frac{x_{nj,k}}{P_{ni}} \cdot \frac{\partial P_{ni}}{\partial x_{nj,k}} = -\beta_k P_{nj} x_{nj,k}$$

- ✍ 確認せよ.
- これらの式に推定された β^* を代入すれば推計値が得られる.

集計的な選択確率 P_i の弾力性

- 全員分を集計した自己弾力性・交差弾力性はそれぞれ

$$E^i_{ik} = \sum_{n \in N} s_{ni} e^{ni}_{ik}, \quad E^i_{jk} = \sum_{n \in N} s_{ni} e^{ni}_{jk}$$

  各個人の選択肢 i における選択シェア s_{ni} で加重平均：

$$s_{ni} \equiv \frac{P_{ni}}{\sum_{m \in N} P_{mi}}$$

- 以上はいずれも **点弾力性** (point elasticity) ⇒ 小さな変化が前提
- 大きな変化には **弧弾力性** (arc elasticity) を用いる

限界効果 (Marginal Effects)：変化量そのもの

- **自己限界効果** (direct ME)：選択肢 i の特性 k の変化率に対して

$$m_{ik}^{ni} \equiv \frac{\partial P_{ni}}{\partial x_{ni,k}} = \beta_k P_{ni}(1 - P_{ni})$$

- **交差限界効果** (cross ME)：選択肢 j の特性 k の変化率に対して

$$m_{jk}^{ni} \equiv \frac{\partial P_{ni}}{\partial x_{nj,k}} = -\beta_k P_{ni} P_{nj}$$

- 集計量についても弾力性と同様に定義する.

厚生変化の評価：MNL の場合

- 選択肢集合 A, 効用ベクトル V_n のとき, 個人 n の **消費者余剰**（円）は

$$\mathbb{E}\left[\mathrm{CS}_n\right] = \frac{1}{\lambda_n}\mathbb{E}\left[\max_{i\in A} U_{ni}\right] = \frac{1}{\lambda_n} S_A(V_n) = \frac{1}{\lambda_n}\log\sum_{i\in A}\exp(V_{ni}) + C$$

  ただし λ_n は個人 n の **所得の限界効用** (marginal utility of income)

- $(A, V) \mapsto (A', V')$ という介入があった場合の消費者余剰の変化は

$$\mathbb{E}[\Delta\mathrm{CS}] = \sum_{n\in N}\mathbb{E}[\Delta\mathrm{CS}_n], \qquad \mathbb{E}\left[\Delta\mathrm{CS}_n\right] = \frac{1}{\lambda_n}\big(S_{A'}(V'_n) - S_A(V_n)\big)$$

- **仮定**：介入前後で λ_n は不変 ← どんなとき妥当か？🤔

所得の限界効用

- 所得の限界効用 λ_n を ARUM で直接推定することはできない
- しかし，もし金銭コスト変数 c_{nj} が効用の定義 V に含まれているなら **Roy の恒等式** (Roy's identity) により推計可能：

$$\lambda_n = -\frac{1}{Q_{nj}} \frac{\partial V_{nj}}{\partial c_{nj}}$$

  - ここで Q_{nj} は例えばモード j による（分析単位期間の）通勤回数
  - この評価の妥当性の条件：j に（ほぼ）依存しないこと

本日のまとめ

- MNLを題材に，推定・評価・政策応用まで基本的分析フローを確認
- パラメータの識別・統計的有意性・汎化性能などの基礎用語の紹介
- VoT や弾力性，消費者余剰など，実務や政策評価に関する話題の紹介

プログラムを利用した演習

次回予告

やってみよう！*Discrete Choice Analysis*

離散選択分析の主流無償パッケージ

- Biogeme （Python 言語）
  - 研究や大規模データの分析に向く.
- Apollo （R 言語）
  - 多様な構造化モデルを簡潔な記述で実装.

※ パッケージ (Package)：ひとまとまりの機能を提供するために記述されたソフトウェアの集合．Maintainer Community が投入した血と汗の量に応じて同じパッケージが多言語対応していることもある.

次回講義までの準備

- Biogeme を手持ちのノート PC で使用できる状態にし，講義に持ってきてください.
- Google Colab などを利用するのでも構いませんが，Wi-Fi へのアクセスは事前に確認してください.
- 使い慣れているソフトウェアがある場合はそれでも構いません.

※ ChatGPT など LLM の活用は学習の趣旨を損なわない範囲で推奨します（なお，正確性の担保は使用者の責任です）.

参考文献

[1] 土木計画学研究委員会 (編) (1996). **非集計行動モデルの理論と実際**. 土木学会.

[2] Small, K. A., & Verhoef, E. T. (2007). **The Economics of Urban Transportation** (2nd Eds.). Routledge.

[3] Train, K. E. (2009). **Discrete Choice Methods with Simulation**. Cambridge.

[4] Ben-Akiva, M. E., & Lerman, S. R. (1985). **Discrete Choice Analysis: Theory and Application to Travel Demand**. MIT Press.

[5] Parady, G., & Axhausen, K. W. (2024). Size matters: The use and misuse of statistical significance in discrete choice models in the transportation academic literature. *Transportation*, 51(6), 2393-2425.