



行動モデルの基礎

と昨日の話題

山梨大学名誉教授 佐々木邦明

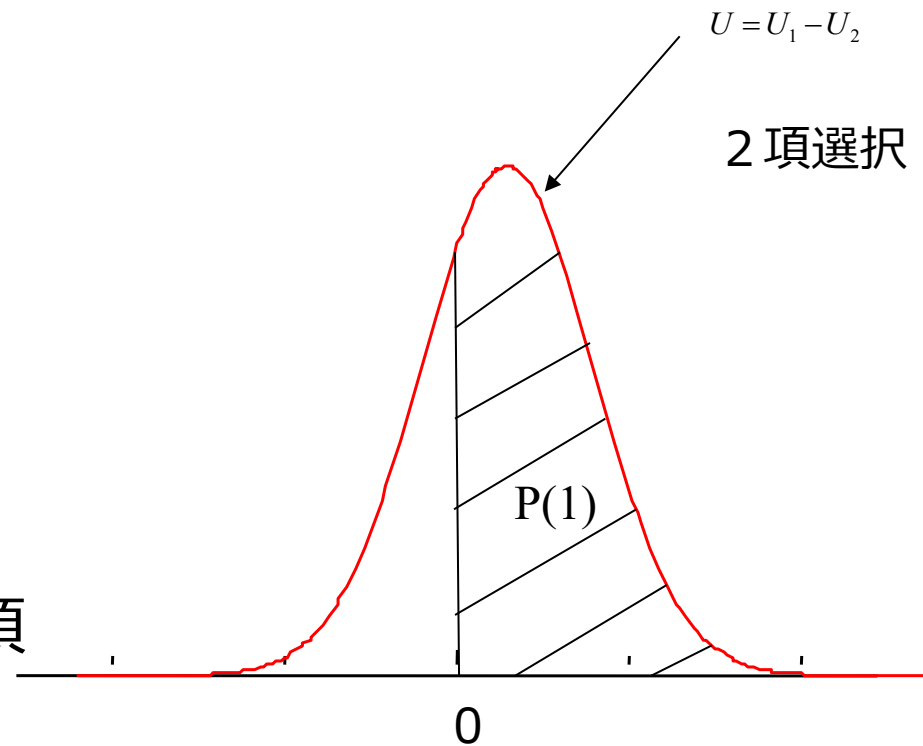
行動モデルの基礎

- 効用関数

- $u_{ij} = \mathbf{X}_{ij}'\boldsymbol{\beta} + \varepsilon_{ij}$
- $\boldsymbol{\beta}$: 未知パラメータベクトル
- ε_{ij} : 個人 j の選択肢 i の効用に関する誤差項

無関係な選択肢からの独立 (IIA)

$$\begin{aligned} P_A/P_B &= \frac{\exp(\mathbf{X}_A'\boldsymbol{\beta}/\mu)}{\sum_k \exp(\mathbf{X}_k'\boldsymbol{\beta}/\mu)} / \frac{\exp(\mathbf{X}_B'\boldsymbol{\beta}/\mu)}{\sum_k \exp(\mathbf{X}_k'\boldsymbol{\beta}/\mu)} \\ &= \frac{\exp(\mathbf{X}_A'\boldsymbol{\beta}/\mu)}{\exp(\mathbf{X}_B'\boldsymbol{\beta}/\mu)} = \exp \mu (\mathbf{X}_A'\boldsymbol{\beta} - \mathbf{X}_B'\boldsymbol{\beta}) \end{aligned}$$



$$\begin{aligned} P_{ij} &= \Pr(u_{ij} \geq u_{ik} \forall k) = \Pr(\mathbf{X}_{ij}'\boldsymbol{\beta} + \varepsilon_{ij} > \mathbf{X}_{ik}'\boldsymbol{\beta} + \varepsilon_{ik}) \\ &= \Pr(\varepsilon_{ij} - \varepsilon_{ik} > \mathbf{X}_{ik}'\boldsymbol{\beta} - \mathbf{X}_{ij}'\boldsymbol{\beta}) = \int_{\mathbf{X}_k'\boldsymbol{\beta} - \mathbf{X}_i'\boldsymbol{\beta}}^{\infty} f(\varepsilon_{ij} - \varepsilon_{ik}) d(\varepsilon_j - \varepsilon_k) \\ &= \frac{\exp(\mathbf{X}_{ij}'\boldsymbol{\beta}/\mu)}{\sum_k \exp(\mathbf{X}_{ik}'\boldsymbol{\beta}/\mu)} \end{aligned}$$

選択肢集合の問題

- ロジットモデル
 - IIA特性により，選択肢集合が誤っていても，無関係な選択肢からの独立によって，ある2選択肢間の選択比率に影響はない
- 選択肢集合の特定を誤ると，パラメータのバイアスが発生する
 - 選択肢集合はどうやって決めるのか？
 - 時空間制約を満たす選択肢はすべて選択肢集合に導入
 - そもそも選択していない

確率的な選択肢集合の取り扱い

- $P_n(i) = \sum_k P(i|C_k)P(C_k)$ Manski型確率的選択肢集合モデル
- C : 選択肢集合
- 具体例 (定食 G , カレー H , 丼 U)
 - $$P_n(G) = P(G|C(G, H, U))P(C(G, H, U)) + P(G|C(G, U))P(C(G, U)) + P(G|C(G, H))P(C(G, H)) + P(G|C(H, U))P(C(H, U)) + P(G|C(G))P(C(G)) + P(G|C(H))P(C(H)) + P(G|C(U))P(C(U))$$
 - 選択肢集合形成モデルを別途作る必要がある

選択肢集合形成モデル

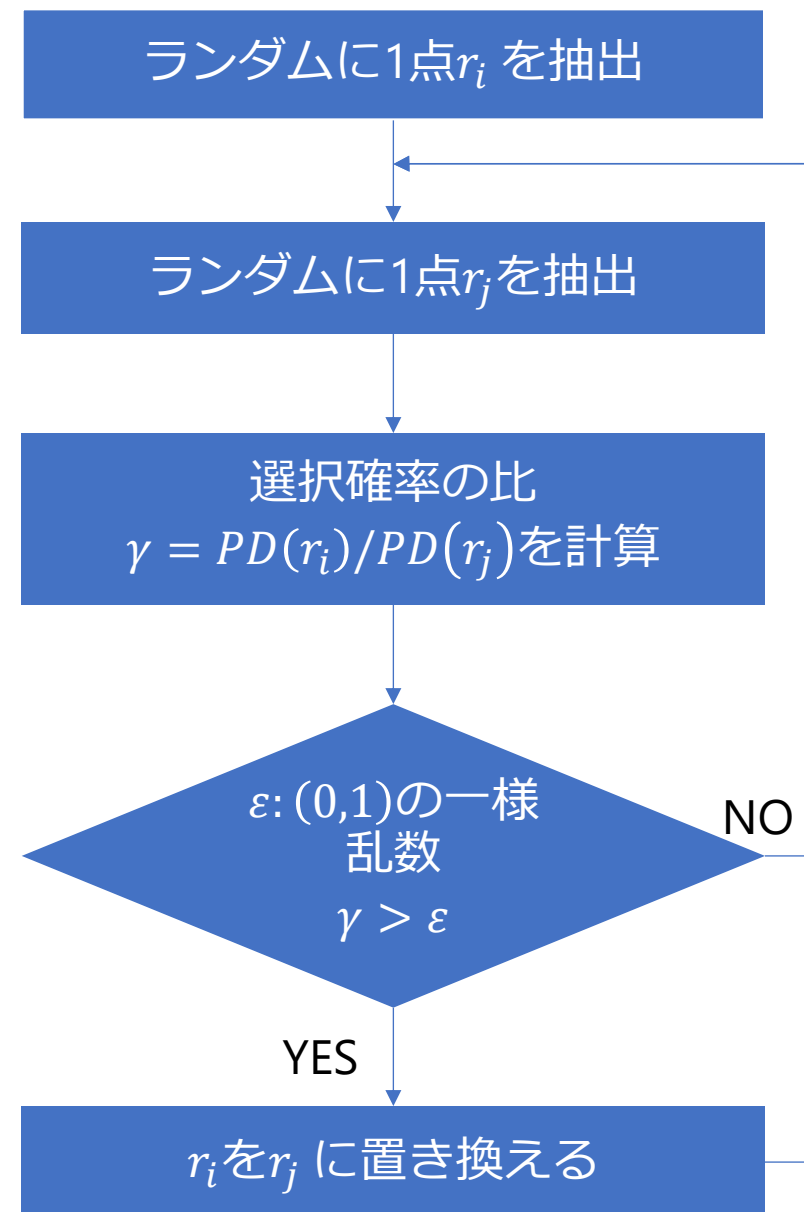
- 選択肢集合に入る(選択対象となる)かどうかは、選択肢ごとに独立に決まっている
 - 丼を選択肢に入れるかどうかは、学食を選択肢にするかどうかと関係がない
- (定食, カレー) の選択肢集合は
 - $P(\text{定食IN}) * P(\text{カレーIN}) * P(\text{丼OUT})$ で決まる. 個別の確率をロジット形式で表現すると
 - $$P(G_{in}) = \frac{1}{1 + \exp\{-f(x_G)\}}$$
- より簡便な形も使われる
$$P(i) = \frac{\exp(V_i - \ln P(i \in C))}{\sum_j \exp(V_j - \ln P(j \in C))}$$

選択肢を絞り込む方法

- 効用最大化以外の意思決定プロセスを考える
 - 連結型：選択肢の属性ごとに必要条件を設定
 - 代替案を規定する各属性の値に関して最低でもこれだけの属性値を持たねばならないという基準が決まっており、全ての属性に関して基準が満たされた場合に限り、その代替案が選別される
 - 分離型：選択肢の属性ごとに十分条件を設定
 - 代替案を選別するために用いられる各属性値のうち、いずれか一つでもその基準値を満たすものがあればその代替案を選別する
 - 辞書編纂型：最も重要な属性から比較していく
 - 最も重要な属性を全ての代替案で比較して最良の代替案を選択する。もしその属性について最良の代替案が複数存在すれば、重要度が高い順に、2 番目、3 番目と比較を続ける

選択肢集合

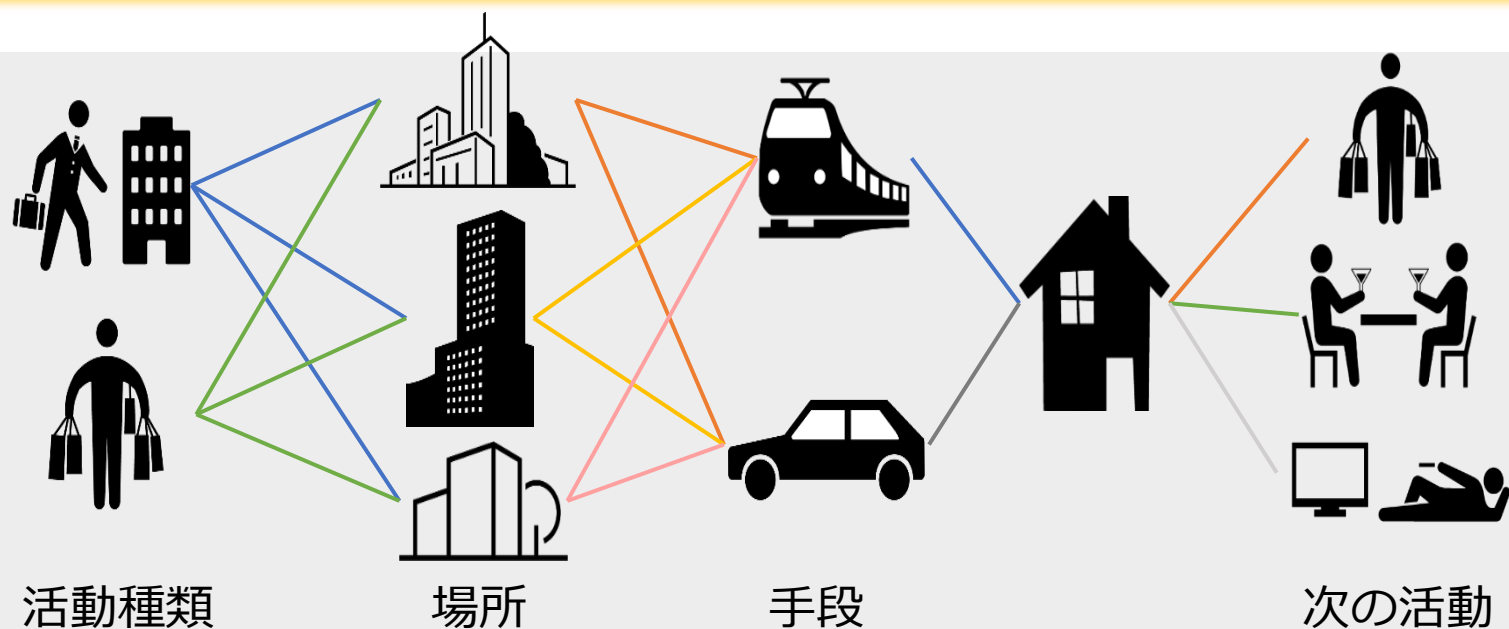
- MCMC的に選択肢集合を近似
 - ランダムに選択肢集合を生成し、得られた選択肢集合での選択確率を計算して、それが一定数より大きい場合には選択肢集合として採択していく(菊池ら, 2001)



多数回繰り返しサンプリング

各種モデルへの展開

選択モデルの計画問題への適用



- 一日の活動パターンを選択モデルの組み合わせで表現する
- 選択結果を合計すると各種施設の需要になる
- 条件を与えると、一日の行動を再現できる

組み合わせを選択するモデル

- 場所と手段の組み合わせ選択
 - 買い物の事例
 - ターミナル駅・電車
 - ターミナル駅・自動車
 - ショッピングモール・電車
 - ショッピングモール・自動車
 - からの選択
- この場合の問題点は何か？

誤差構造

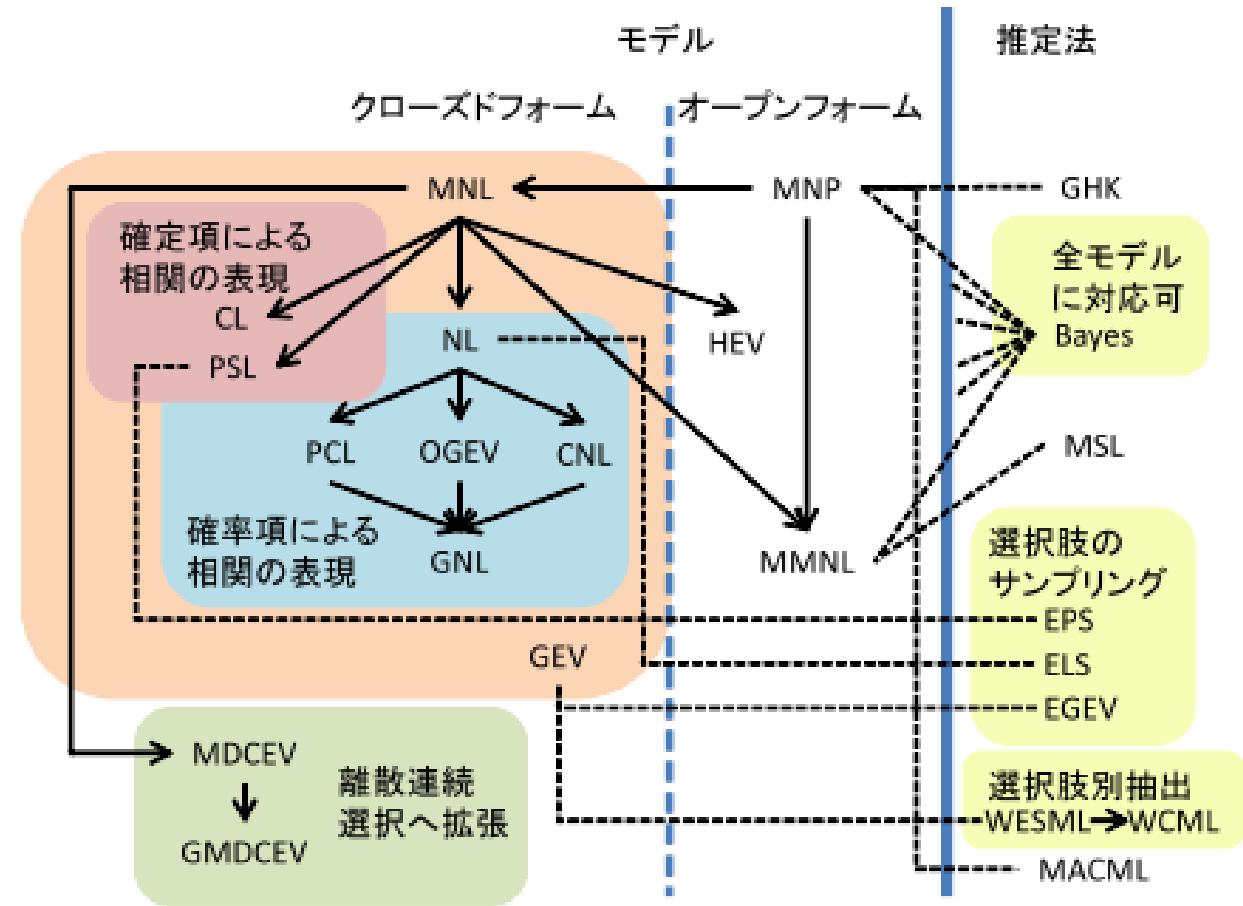
- 
- $u_{dept*tra} = X'_{dept}\boldsymbol{\beta} + X'_{tra}\boldsymbol{\beta} + X'_{dept*tra}\boldsymbol{\beta} + \varepsilon_{dept} + \varepsilon_{tra} + \varepsilon_{dept\times tra}$
 - $u_{sub*tra} = X'_{sub}\boldsymbol{\beta} + X'_{tra}\boldsymbol{\beta} + X'_{sub*tra}\boldsymbol{\beta} + \varepsilon_{sub} + \varepsilon_{tra} + \varepsilon_{sub\times tra}$

となつて, 誤差に相関が生まれる

- $Var(\varepsilon_{dept} \times \varepsilon_{sub}) = 0$
- $Var(\varepsilon_{tra} \times \varepsilon_{tra}) \neq 0$
- $Var(\varepsilon_{dept\times tra} \times \varepsilon_{sub\times tra}) \cong 0$

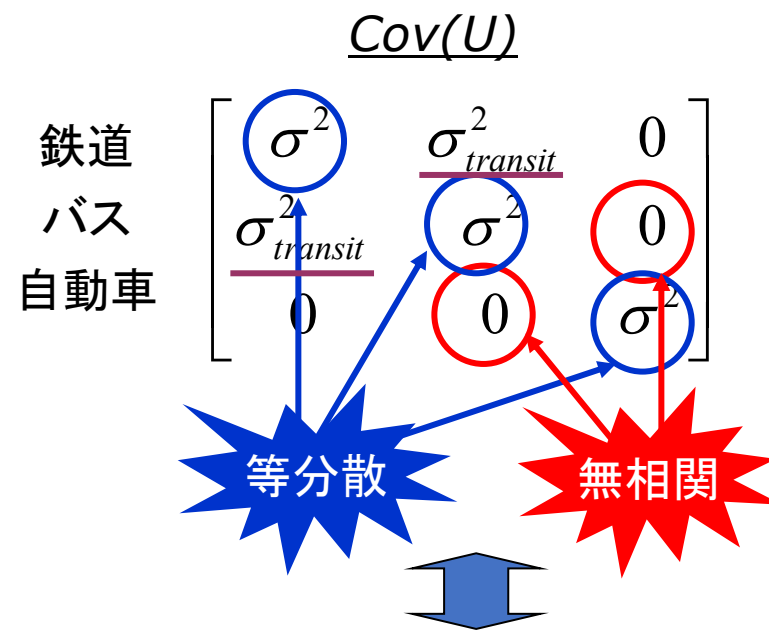
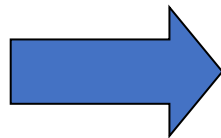
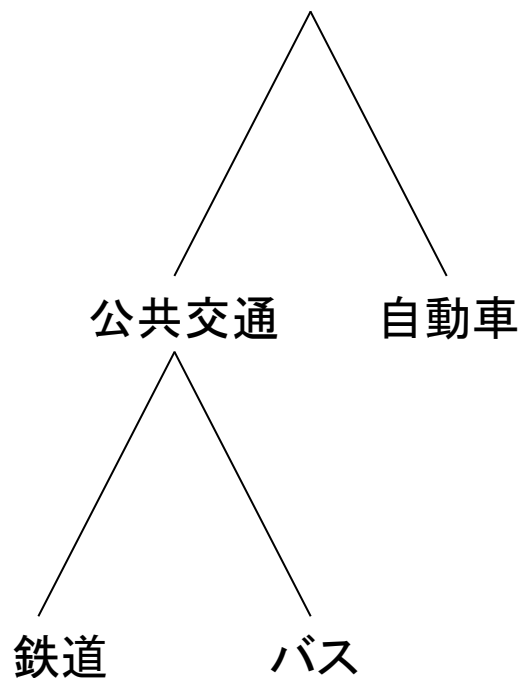
IIAを緩和するために

- 積分なし：クローズドフォーム
- 積分あり：オープンフォーム
- 推定法：単純な最尤推定よりも効率的な方法



ネスティッドロジットモデル

NLモデルの誤差構造



一般化した誤差構造(MNPモデル)

鉄道
バス
自動車

$$\begin{bmatrix} \sigma_{rail}^2 & \sigma_{rail,bus} & \sigma_{rail,car} \\ \sigma_{rail,bus} & \sigma_{bus}^2 & \sigma_{bus,car} \\ \sigma_{rail,car} & \sigma_{bus,car} & \sigma_{car}^2 \end{bmatrix}$$

ミックストロジット (MXL) モデル

プロビット
モデルの
柔軟な誤
差構造

プロビットタイプ
の誤差項

$$\begin{aligned} U_{car} &= \beta X_{car} + \eta_{car} + v_{car} \\ U_{bus} &= \beta X_{bus} + \eta_{bus} + v_{bus} \\ U_{rail} &= \beta X_{rail} + \eta_{rail} + v_{rail} \end{aligned}$$

ロジットモ
デルの操
作性

IIDガンベル分布

$$L(car|\eta) = \frac{e^{\beta X_{car} + \eta_{car}}}{e^{\beta X_{car} + \eta_{car}} + e^{\beta X_{bus} + \eta_{bus}} + e^{\beta X_{rail} + \eta_{rail}}}$$

$$P(car) = \iiint_{\eta} \Lambda(car|\eta) f(\eta) d\eta$$

$$\varepsilon = \begin{matrix} & \eta & & \\ \begin{bmatrix} \sigma_{car}^2 & \sigma_{car,bus} & \sigma_{car,rail} \\ \sigma_{car,bus} & \sigma_{bus}^2 & \sigma_{bus,rail} \\ \sigma_{car,rail} & \sigma_{bus,rail} & \sigma_{rail}^2 \end{bmatrix} & + & \begin{matrix} v \\ \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \end{matrix} \end{matrix}$$

離散連続モデル

- 離散系の選択と連続系の量
- 自動車の車種選択と高速走行距離
- 活動選択と活動時間
- 効用関数によって、離散選択と連続選択が同時に決まる

モデルの体系（福田・力石 2013）

- 誘導型
 - 連続量に条件を付けて離散選択を内包（Tobitモデル等）
- 構造型
 - 資源制約下の効用最大化からの導出
 - キューンタッカー条件
 - ロワの恒等式(間接効用関数から需要関数)

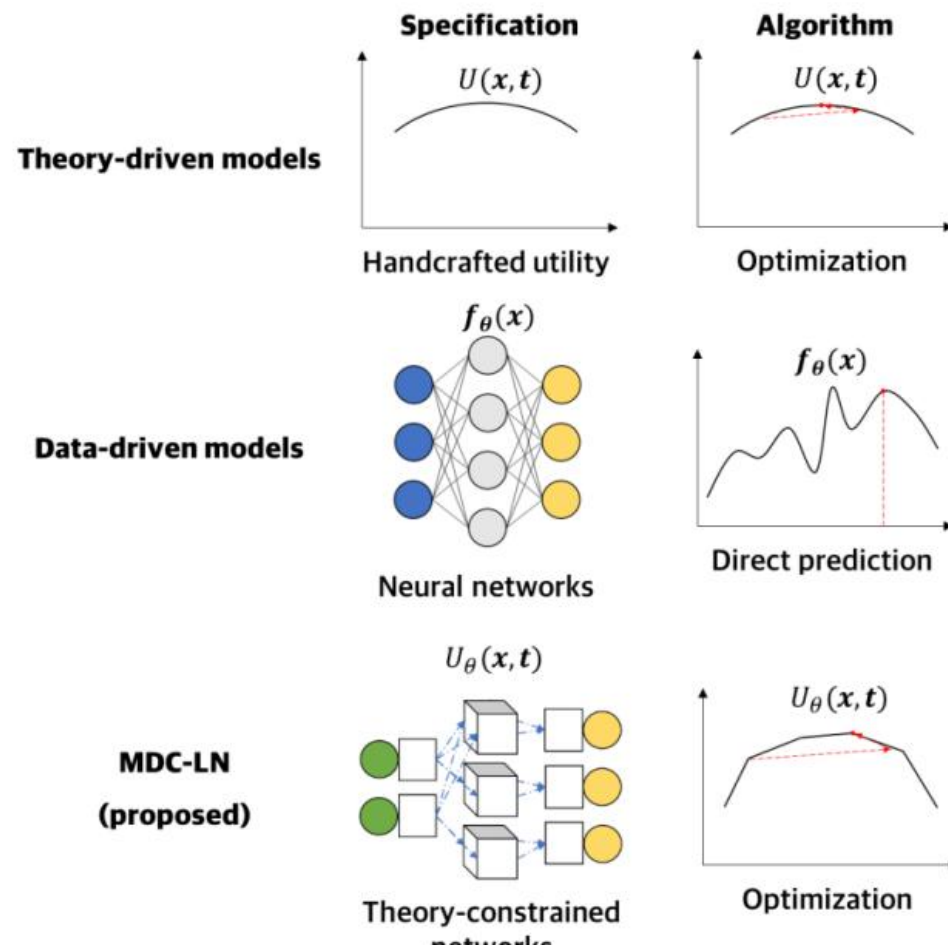
Multiple Discrete-Continuous Extreme Value (MDCEV)モデル

- 非観測要因がガンベル分布に従うクローズドフォームの尤度関数を持つ
- 複数財の選択と量を記述

$$U(x) = \sum_k \frac{\gamma_k}{\alpha_k} [\exp(\beta' z_k + \varepsilon_k)] \cdot \left\{ \left(\frac{x_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right\}$$

$$P(x_1^*, x_2^*, x_3^*, \dots, x_M^*, 0, 0, \dots, 0) = \frac{1}{p_1} \cdot \frac{1}{\sigma^{M-1}} \left[\prod_{i=1}^M f_i \right] \left[\sum_{i=1}^M \frac{p_i}{f_i} \right] \left[\frac{\prod_{i=1}^M e^{V_i/\sigma}}{(\sum_{k=1}^K e^{V_k/\sigma})^M} \right] (M-1)!$$

$$f_i = \left(\frac{1 - \alpha_i}{x_i^* + \gamma_i} \right)$$



パラメータの推定に

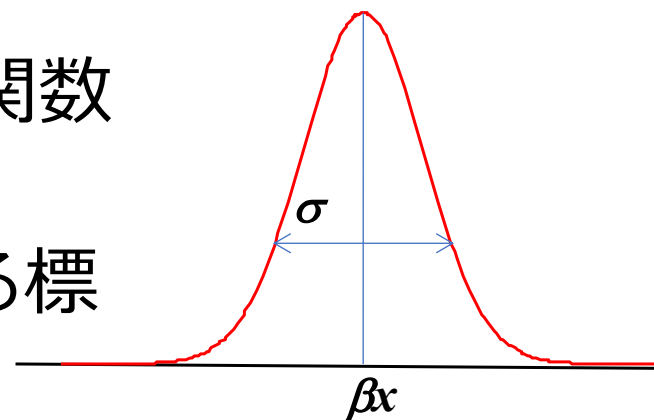
パラメータの決定方針

- データに最も適合するように β を決めたい
- データ
 - 選択データ{ {Car}, {Train}, {Train}, {Car}, {Car}.....}
 - 属性データ x ({CostCar1, CostTrain1, TimeCar1, TimeTrain1},
{CostCar2, CostTrain2, TimeCar2, TimeTrain2},
{CostCar3, CostTrain3, TimeCar3, TimeTrain3},
{CostCar4, CostTrain4, TimeCar4, TimeTrain4},
{CostCar5, CostTrain5, TimeCar5, TimeTrain5}.....)
- ロジットモデルで計算される選択確率が、実際の選択データに最も適合する β とする
- 最も適合する確率とは？

行動モデルの推定と最尤推定

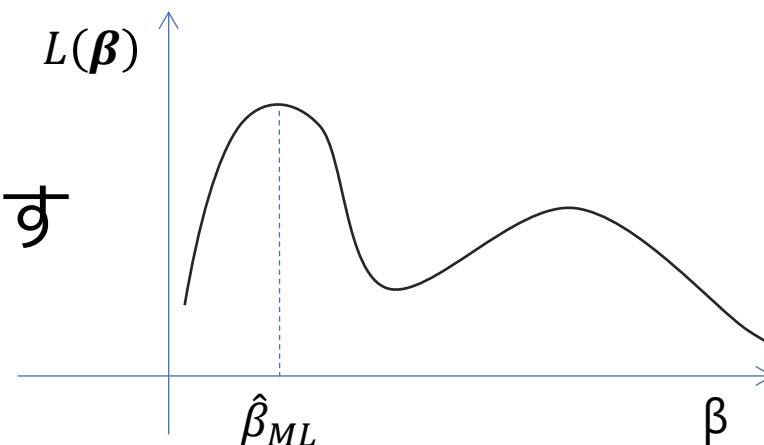
- 有限個のパラメータで記述される確率密度関数の推定
- パラメータベクトル β の下で, モデル f による標本の生起確率を尤度とする

- $L(\beta) = \prod_{i=1}^n f(y_i | \beta)$



- (対数)尤度関数が最大になる β を最尤推定値とする

- $\hat{\beta}_{ML} = \operatorname{argmax}_{\beta} \log L(\beta)$



最尤推定法

- 点推定量を求める一般的な方法
- 右上の式を β の関数とみなしたものが尤度関数
- 尤度関数を最大化する β の値を最尤推定量とするのが最尤推定法

$$(\beta|y) = \prod_{i=1}^n f(y_i|\beta)$$

平均値の推定を例にすると

データ(**x : 3,5,4**)が得られたとき,
平均をいくつとするのがよいか？

⇒平均がいくつの分布だったら

データ(**x : 3,5,4**)がもっとも得られやすいか？

ロジットモデルの最尤推定

- $L(\boldsymbol{\beta}) = \prod_{i=1}^n f(y_i | \boldsymbol{\beta})$

- $f(y_i | \boldsymbol{\beta}) = \prod_{j=1}^J \left\{ \frac{\exp(V_{ij})}{\sum_{j=1}^J \exp(V_{ij})} \right\}^{y_{ij}}$

選ばれた選択肢の選択確率

β は未知数, x は既知のデータ

- $V_{ij} = \boldsymbol{\beta} \mathbf{x}_{ij} = \beta_1 + \beta_1 x_{1i} + \beta_2 x_{2i} \cdots + \beta_K x_{Ki}$

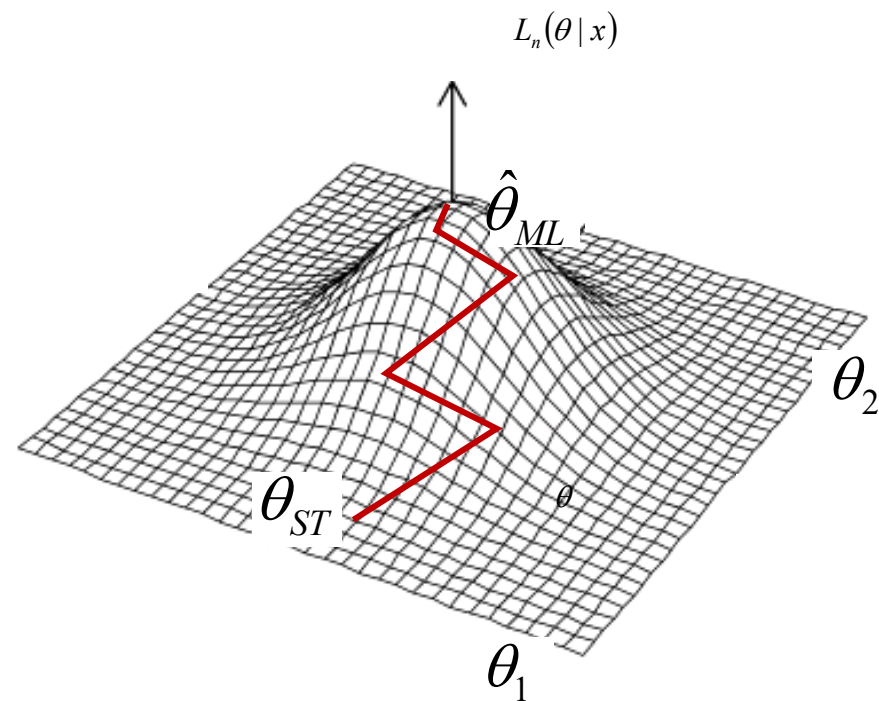
データ($\mathbf{y}$: 車,車,鉄道,鉄道,鉄道,車,鉄道, ...)が得られたとき, $\boldsymbol{\beta}$ をいくつかとすると, 再現性が高いのか?

⇒ $\boldsymbol{\beta}$ がいくつかだったらデータ($\mathbf{y}$)が得られやすいのか?
 $\boldsymbol{\beta}$ を色々と変えてみて一番Lが高くなる $\boldsymbol{\beta}$ を探す

最大化アルゴリズムの考え方

周りが見えない中で、近傍の情報から頂点を目指す

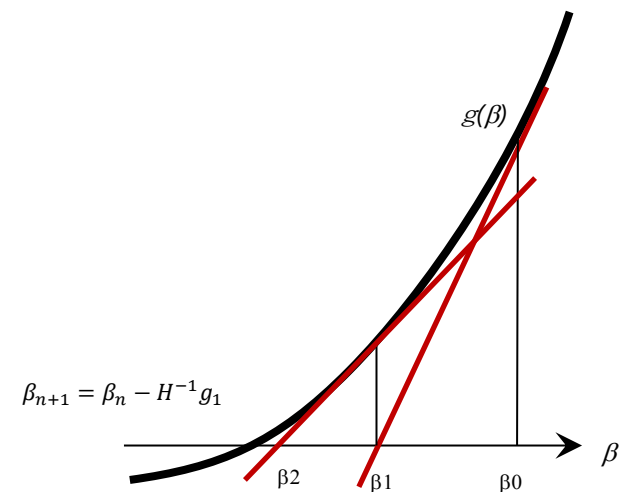
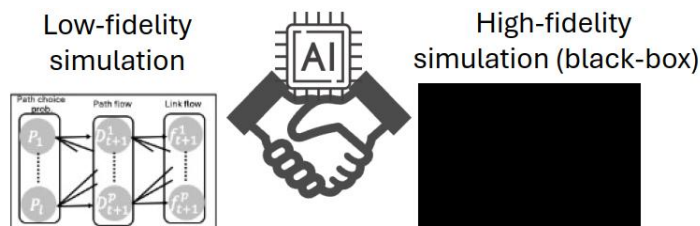
- 対数尤度関数の段階的な最大化
 - 初期値を与える
 - 初期値周りで勾配(1次微分)等を用いて次の推定値の方向を決める
 - 初期値付近で1次微分, 2次微分を用いて適切に次の点を決めて推定値を得る
 - 収束基準(一次微分ベクトル)で判定し, 収束していない場合は, 現在の値から次の推定値に移る



代表的な繰り返し計算法

尤度関数を最大化 尤度関数の一階微分 = 0 を解く

- Newton-Raphson法
 - テイラー展開の1次近似を利用して進める
- 準Newton法 (BFGS, L-BFGS法)
 - ヘッセ行列を, パラメータの差分と一階微分の差分を用いて逐次近似する.
 - L-BFGSはヘッセ行列の更新式を展開して, 初期値と差分の関数形で表す.
- 確率的勾配降下法
 - サンプルと勾配降下法を合わせて実施



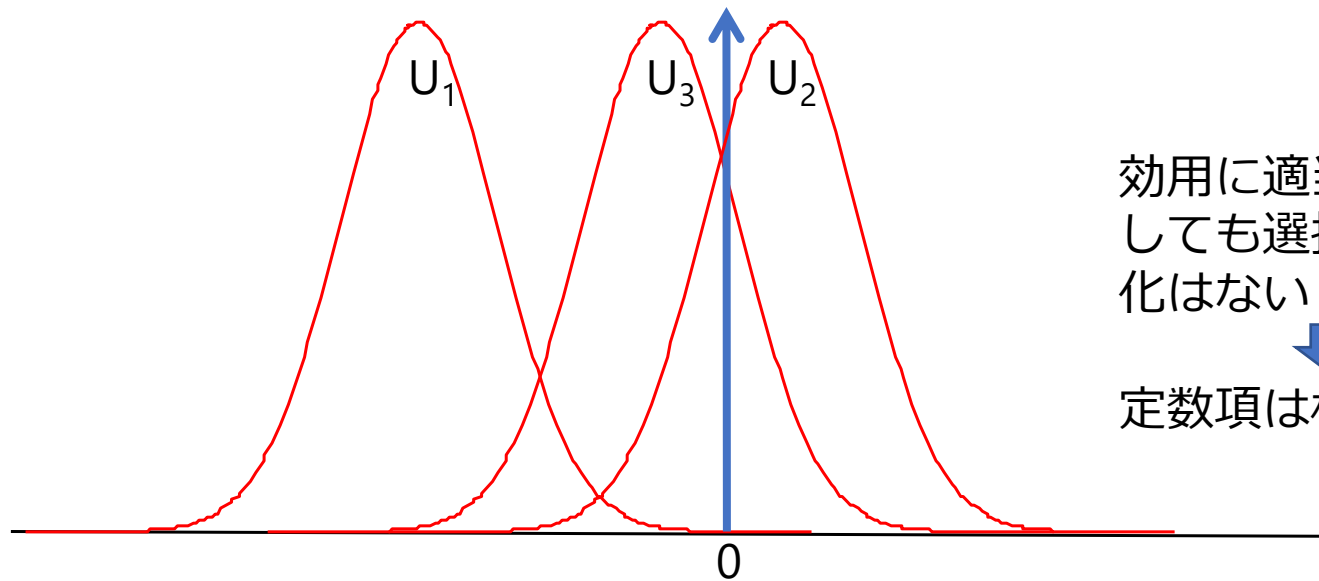
H: 尤度関数の二階微分 ヘッセ行列
g: 尤度関数の一階微分

パラメータ推定の課題

- 序数的効用と基数的効用
- 数が規定できるができるのが基数的効用
 - 測定可能で個人間比較が可能
- 大小関係のみが序数的効用
 - 直接測定不可能で個人間比較が不可能

注意

個人属性
選択肢特有の変数
選択肢ごとに異なるが、個人間で変化しない



効用に適当な数を足しても選択確率に変化はない



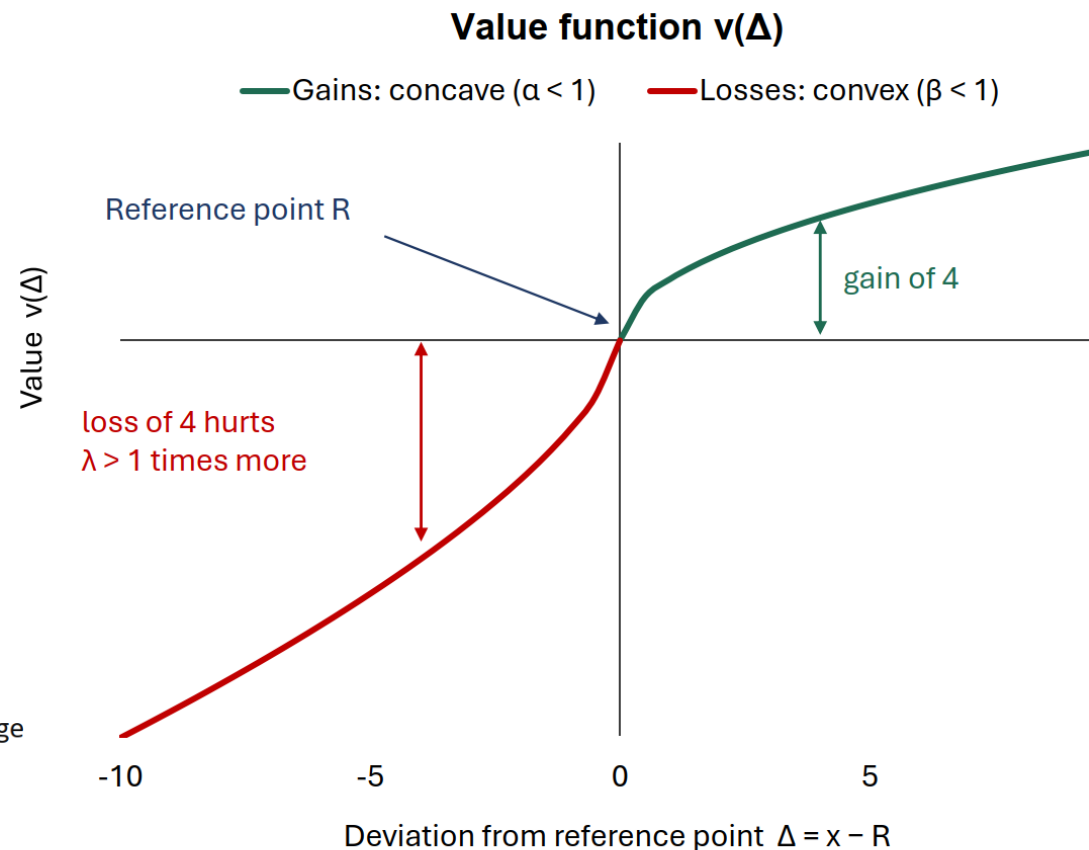
定数項は相対的数値

プロスペクト理論

- 参照点からの差分が価値関数になっている
- Kahneman and Tversky (1979) Prospect Theory: An Analysis of Decision under Risk
- 基数効用と参照点を想定

$$\bar{\tau}_{0,n} = \bar{\tau}_n + \tanh\left(f_{\text{ref}}(X_{\text{ind},n}; \varphi_{\text{ref}})\right) \times \sigma(v_{\text{raw}}) \times (\tau_{\text{max},n} - \tau_{\text{min},n})$$

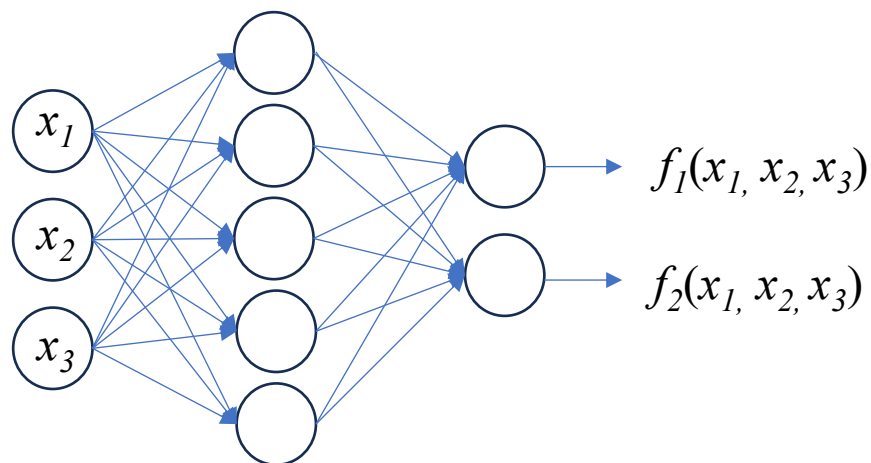
$\bar{\tau}_n$ anchor $x_{k,n} \rightarrow$ reported average commute time; $\text{RNG}_{k,n} \rightarrow$ individual's experienced commute range



ニューラルネットワーク (NN) は なにがいいのか？

$$Y = NN(R)$$

- 普遍性定理
 - 中間層を 1 層以上持つNNは任意の連続関数を近似できる



Utility

$$V_{ni} = \beta^\top x_{ni}$$

$$V_n = NN(x_n, z_n)$$

$$V_{ni} = NN_i(x_{ni}, z_n)$$

$$V_{ni} = \beta^\top x_{ni}^K + NN_i(x_n^L)$$

$$V_{ni} = (1 - \delta) V_{T,i} + \delta V_{DNN,i}$$

$$V_{ni} = ASC_i + \sum_k NN_{ik}(x_{nik})$$

$$V_n = NN(x_n, z_n) + \text{Gradient penalty during training}$$

$$V_{ni} = NN_i^{mono}(x_{ni})$$

効用関数の立場は？

nature machine intelligence

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾ [Subscribe](#)

[nature](#) > [nature machine intelligence](#) > [perspectives](#) > article

Perspective | Published: 13 May 2019

Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

[Cynthia Rudin](#) 

- 解釈可能性を残したい
 - 政策評価等々
 - 合意形成？
- 解釈可能性をどう入れ込むのか？
 - ある意味直観に整合する制約を入れ込む
 - 最大効用を与える選択枝を選択する
 - 効用は属性の関数で表現できる
- 制約
 - データ構造に入れこむ事前知識
 - スパース性

そもそも

- 効用あるいは効用関数

- $u_j(car) = \alpha_{car} + X'_{carj}\beta + \varepsilon_{carj} = \alpha_{car} + cost'_{carj}\beta + time'_{carj}\beta + \varepsilon_{carj}$
- $u_j(rail) = \alpha_{rail} + X'_{railj}\beta + \varepsilon_{railj} = \alpha_{rail} + cost'_{railj}\beta + time'_{railj}\beta + \varepsilon_{railj}$

- 制約条件付きの最大化

$$\max_{X_1, X_2} U(X_1, X_2)$$

$$s.t. \quad p_1 X_1 + p_2 X_2 = Y$$

ラグランジュの未定乗数法を用いて解くと

$$U(X_1^*, X_2^*) = U\left(X_1(p_1, p_2, Y), X_2(p_1, p_2, Y)\right) = V(p_1, p_2, Y)$$

※予算制約を例として

なんで個人属性が効用に？

- 個人属性が効用関数にある理由

- 年齢によって効用が違うとは？
- 職業によって異なる効用

- 役割

- 観測できない変数の代理変数とする
 - 例：間接効用関数に所得を入れたいが、難しければ年齢や職業で近似する
- パラメータの違い（異質性）を補完する
 - 通勤では、所要時間の感度が大きい（パラメータの絶対値が大きくなる）としたら、所要時間のパラメータを目的の関数にすればよい
- 観測されない変数の代理変数としたら、観測されない変数の変化による影響を受ける

1st time: I'm in a hurry today, so **travel time** matters most. I'll choose the fastest option.

2nd time: **Comfort** matters most to me now. I'll choose the least crowded option.

...

intra-individual heterogeneity