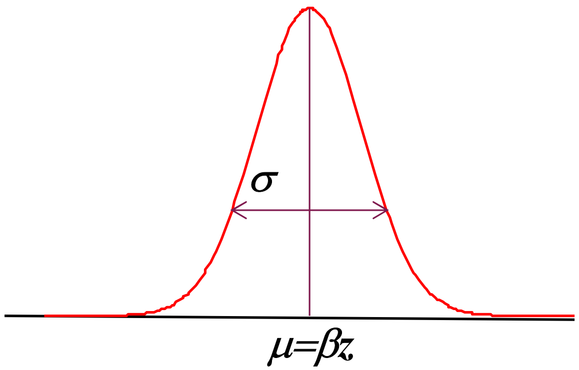
行動モデルのパラメータ推定

早稲田大学　佐々木邦明

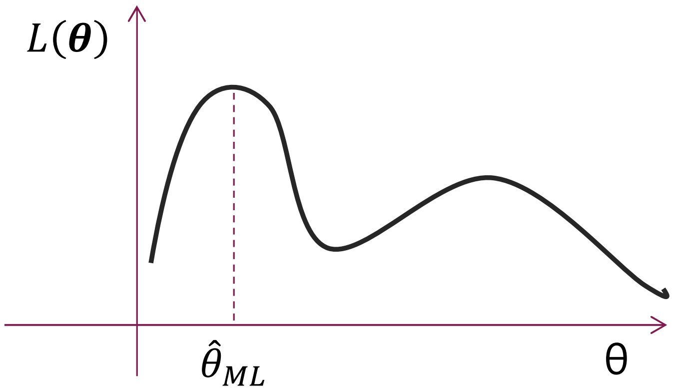
最尤推定

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θの下で，モデルfによる標本の生起確率を尤度とする
  - $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{y}_i|\boldsymbol{\theta})$



- (対数)尤度関数が最大になるθ を最尤推定値とする
  - $\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \log L(\theta)$



最尤推定法

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

- 点推定量を求める一般的な方法
- 右上の式をθ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 θの値を最尤推定量とするのが最尤推定法

平均値の推定を例にすると

データ(**y：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(**y：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

- $L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(\boldsymbol{y}_i|\boldsymbol{\beta})$
- $f(\boldsymbol{y}_i|\boldsymbol{\beta}) = \prod_{j=1}^{J} \left\{ \frac{exp(V_{ij})}{\sum_{j=1}^{J} exp(V_{ij})} \right\}^{y_{ij}}$
- $V_{ij} = \boldsymbol{\beta x}_{ij} = \beta_1 + \beta_1 x_{1i} + \beta_2 x_{2i} \cdots + \beta_K x_{Ki}$

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

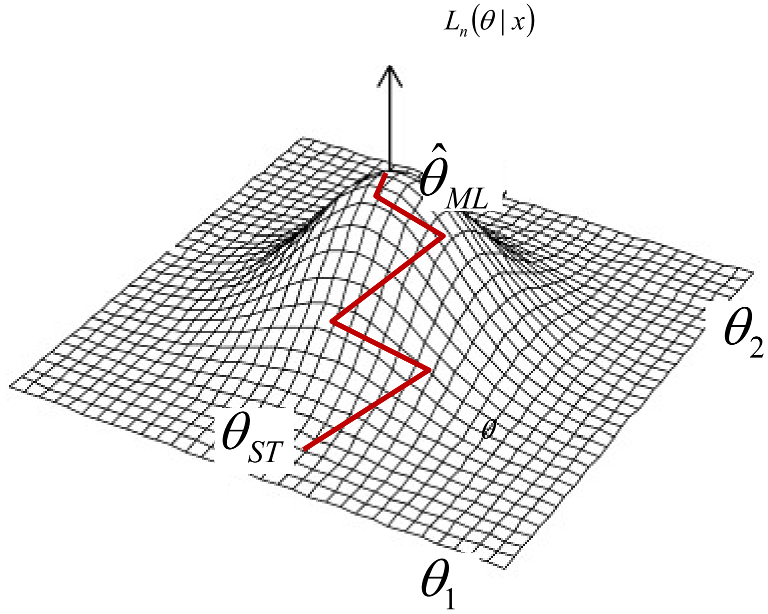
データ(**y**：馬,馬,鉄道,鉄道,鉄道,馬,鉄道,・・)が
得られたとき，**β**をいくつとすると，再現性が高いのか？

⇒ **β**がいくつだったらデータ(**y**)が得られやすいのか？
βを色々と変えてみて一番**L**が高くなる**β**を探す

最大化アルゴリズムの考え方

周りがあまり見えない中で，近傍の情報から頂点を目指す

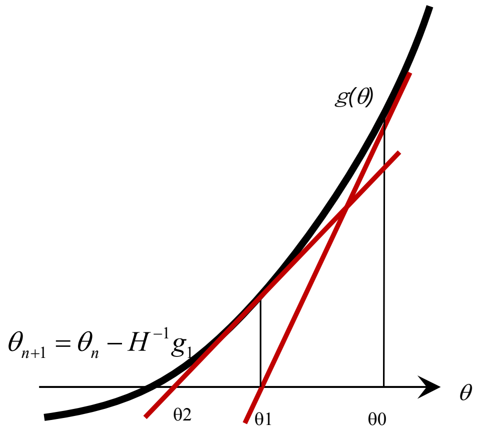
- 対数尤度関数の段階的な最大化
  - 初期値を与える
  - 初期値周りで勾配(1次微分）等を用いて次の推定値の方向を決める
  - 初期値付近で１次微分，２次微分を用いて適切に次の点を決めて推定値を得る
  - 収束基準(一時微分ベクトル)で判定し，収束していない場合は，現在の値を初期値として2に進む



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和で表す.
- 焼きなまし法（SANN）
  - 適当にパラメータを遷移しつつ，関数が大きくなる遷移を選ぶ.
  - 悪化する遷移もある確率で許容し，局所解を避ける

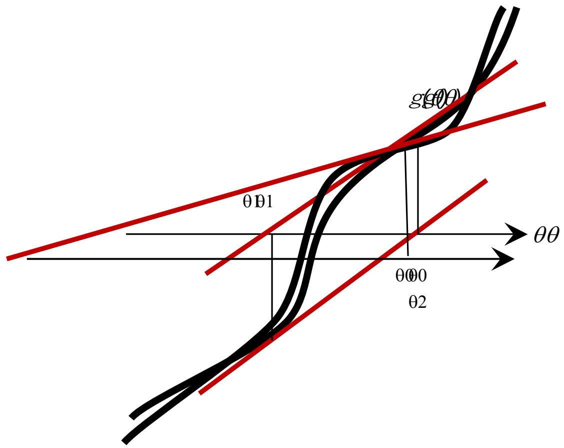


H:尤度関数の二階微分　ヘッセ行列
g: 尤度関数の一階微分

パラメータ推定がうまくいかない

- 収束するとはθ_{n+1}とθ_nが同じになる
  - g′が0になる
- 収束しない
  - 無限に繰り返す
  - θ_2が計算不能
- 局所最適解
  - 見かけ上の最大化

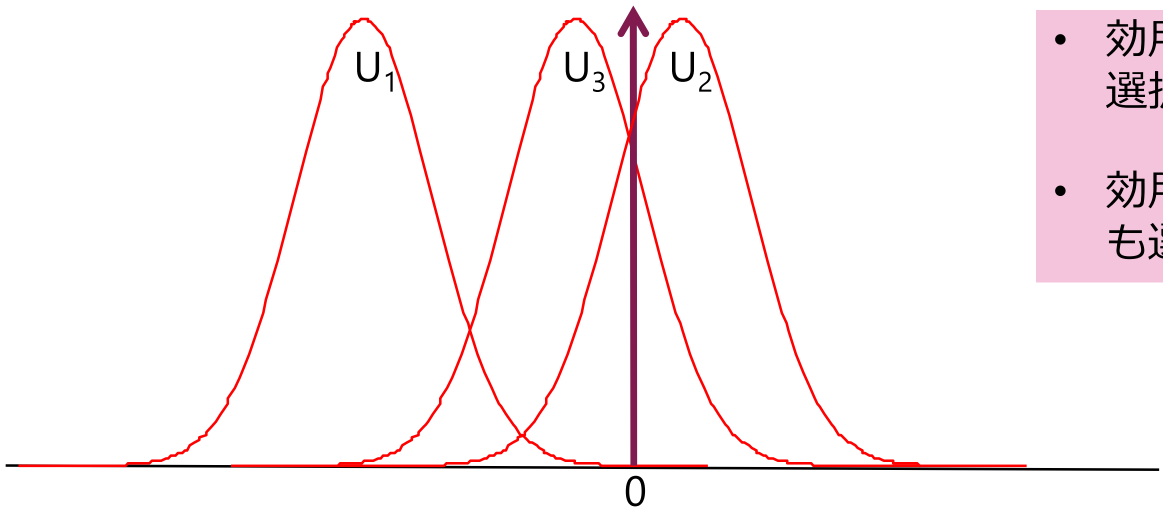
- H^{-1}ヘッセ行列の逆行列が早々に死亡
  - 変数が完全相関
  - 変数が効用関数に影響しない式形
- 関数の近似状況
  - 初期値の問題
- **推定プログラムに誤り**



そもそも識別不可

- 最大値において唯一解が求まらない可能性がある（最大値となるパラメータベクトルが無数にある）
  - 意思決定者間では異なるが，選択肢間では異ならない変数
  - 選択肢間では異なるが，意思決定者間で異ならない変数
  - 異なるモデル間でのパラメータの直接比較は無意味

例



- 効用に適当な数を足しても選択確率に変化はない
- 効用に適当な倍率をかけても選択確率に変化がない

最尤推定法の理論的性質

- 一致性
- 漸近不偏性
- 漸近有効性
- 漸近正規性

最尤推定法におけるモデル選択

- 真の確率密度関数を近似するものが含まれる必要がある
- ⇒フレキシブルなモデルを選ぶ

- 最尤推定は自由度の高さ前提
- ⇒自由度が低すぎるモデルは不適切

$$\begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix}$$

- 平均対数尤度の比較（KL情報量）
  - 例えば，共分散行列を考える
  - (非)制約モデル（A対称行列，B対角行列，C対角行列で分散同一）を考えるとCはBに含まれ，BはAに含まれるので，平均対数尤度L*は必ず
  - L*(A)≧L*(B)≧L*(C)になる．

赤池の情報量基準（AIC）

- KL情報量が適切でない理由：平均対数尤度が期待対数尤度のよい近似になっていない.

負のエントロピー　期待対数尤度

- $\mathrm{KL}(p||\hat{p}) = E\{\log p(\boldsymbol{x})\} - E(\log \hat{p}(\boldsymbol{x}))$

- AIC(赤池の情報量基準)
  - $\mathrm{AIC} = -\sum_{i=1}^{n} \log f\left(x_i, \hat{\theta}_{ML}\right) + t$
  - 負の対数尤度に補正項 t (パラメータ数）を足す

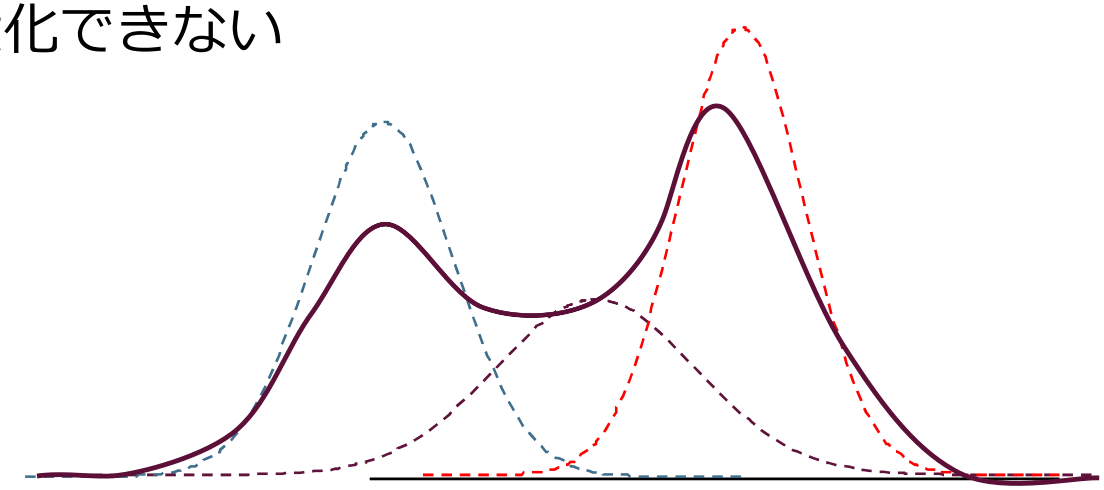
混合モデル

混合モデルの推定技法

- 混合モデル？
  - $f(\boldsymbol{x}_i|\theta) = \sum_{i=1}^{m} w_i \emptyset(\mu, \sigma^2)$ $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{x}_i|\boldsymbol{\theta})$
  - $\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \log L(\theta) \quad subject\ to \begin{cases} w_{1,} \cdots w_m \geq 0 \\ \prod_{i=1}^{m} w_i = 1 \end{cases}$

媒介変数

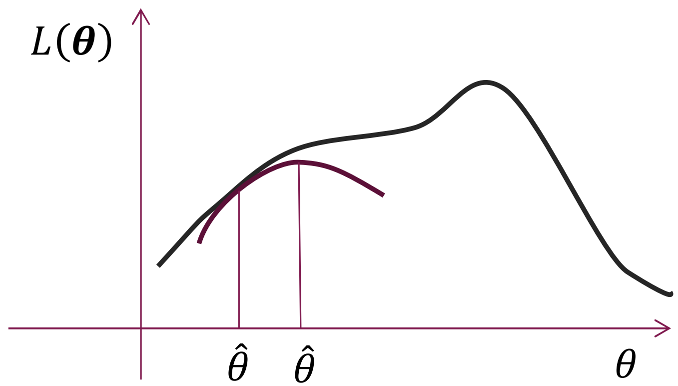
- 媒介変数を用いて尤度関数を表現できる． $w_l = \frac{exp(\gamma_l)}{\sum_{l'=1}^{L} exp(\gamma_{l'})}$
- ただし簡単に最大化できない



混合モデルの推定技法

- EM法
  - 不完全データの最適化法
  - 混合モデルは不完全データからの学習法

- 適当な初期値を定める
- 初期値に応じて媒介変数を求める（E）
- 求めた媒介変数から解を計算する（M）
- 対数尤度関数は減少せず，局所最適解に収束する



EMアルゴリズムと潜在クラスモデル

- Eステップ（Expectation-step）：欠測値として扱う変数の期待値の推定
  - 個人iがクラスsに属する場合に1となるyisの期待値yis*を求める

- Mステップ(Maximization-step)：その期待値を用いたパラメータ推定
  - yis*を用いてパラメータ（πsとPi|s(j)）を推定する
  - 推定されたπs とPi|s(j)を用いてyis*を更新する

- ２つのプロセスを繰り返して最尤推定

おわりに

行動モデルを用いた政策分析

- **行動モデルを推定し，そのパラメータを用いて，変数の変化による選択の変化を見る．**

- 政策：変数の変化
  - 例えば：**所要時間**を短縮，**駐車場の料金**を割り引く等，政策に対応した変数が必要

- 政策評価
  - 数え上げ法：個人の選択確率を予測して，積み上げる　$S(j) = \frac{1}{N}\sum_{i=1}^{N} P_i(j)$

シミュレーションによる政策分析

- 行動モデルを推定し，そのパラメータを用いて，変数の変化による選択の変化を見る.
- 回遊行動の分析
  - マイクロシミュレーションを用いることで，複雑なモデルの組み合わせを政策評価可能
  - シミュレーションなので，複数回実施して平均的な評価を行う

Step 1	サンプルのデータ，個人属性や発ゾーン，LOSデータなどを各段階のモデルに個人nの個人属性やLOS，各種ダミー等といった説明変数データを代入し，全ての選択肢ごとの選択確率P_{in}を求める．求めた選択確率から確率分布F_{in}を作成する.
Step 2	[0,1]の一様乱数γ_nを発生させ，γ_nの値が，$F_{(i-1)n} \leq \gamma_n < F_{in}$を満たす選択肢$i$を選択するものとする.