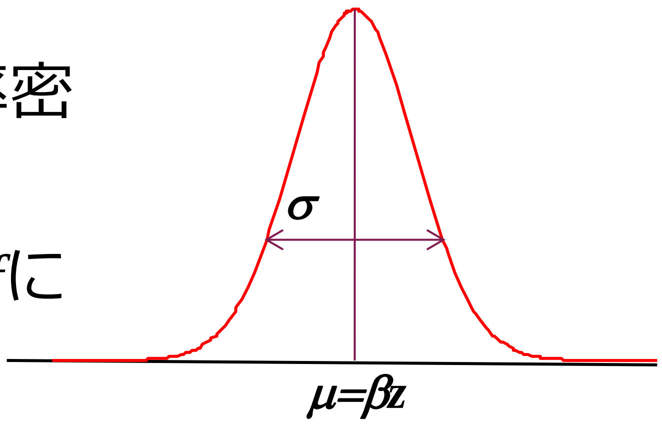
行動モデルの推定

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

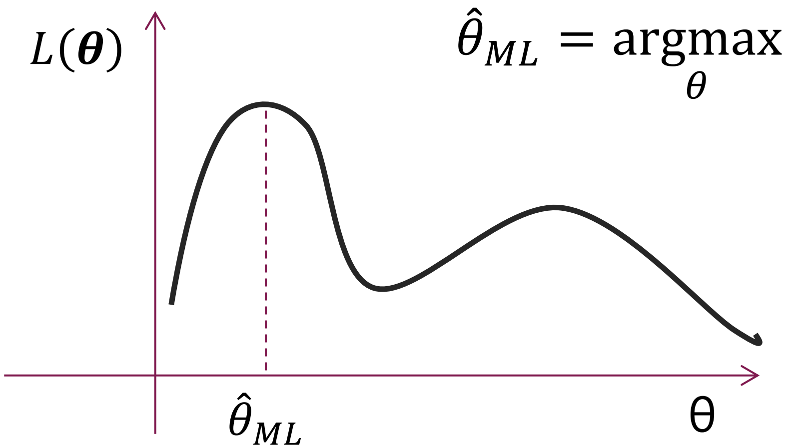
- 有限個のパラメータで記述される確率密度関数の推定
- パラメータベクトル$\boldsymbol{\theta}$の下で，モデルfによる標本の生起確率を尤度とする



$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{y}_i|\boldsymbol{\theta})$$

- (対数)尤度関数が最大になる$\boldsymbol{\theta}$ を最尤推定値とする

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \log L(\theta)$$



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

- 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

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

  - (非)制約モデル（A対称行列，B対角行列，C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.

$$\mathrm{KL}(p||\hat{p}) = \underbrace{E\{\log p(\boldsymbol{x})\}}_{\text{負のエントロピー}} - \underbrace{E(\log \hat{p}(\boldsymbol{x}))}_{\text{期待対数尤度}}$$

- 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

- オッカムさん曰く「同じ程度に現象を説明できる仮説があるならば，単純なものを選ぶべき」

混合モデル

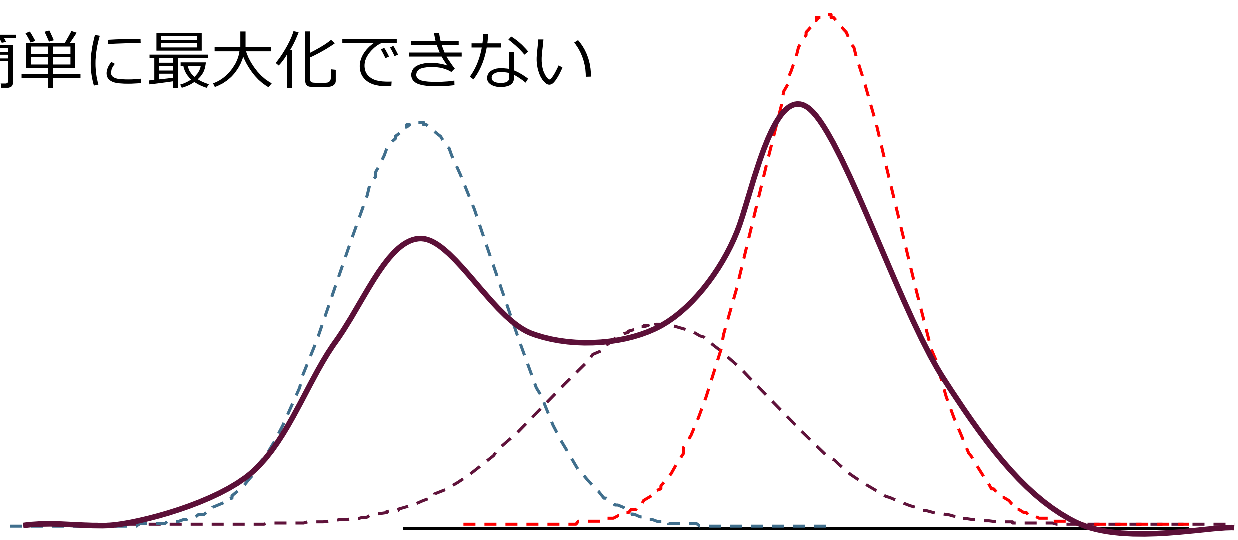
混合モデルの推定技法

- 混合モデル？

$$f(\boldsymbol{x}_i|\theta) = \sum_{i=1}^{m} w_i \emptyset(\mu, \sigma^2) \quad L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\boldsymbol{x}_i|\boldsymbol{\theta})$$

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \log L(\theta) \quad subject\ to \begin{cases} w_{1,} \cdots w_m \geq 0 \\ \prod_{i=1}^{m} w_i = 1 \end{cases}$$

- 媒介変数を用いて尤度関数を表現できる.
- ただし簡単に最大化できない



混合モデルの推定技法

- 勾配法
  - 適当な初期値を決める
  - 初期値周りの勾配を求め，適当なεを用いて勾配方向に上る
  - 上ったところを新たなパラメータとして繰り返す

- εの決め方によって収束速度が異なる

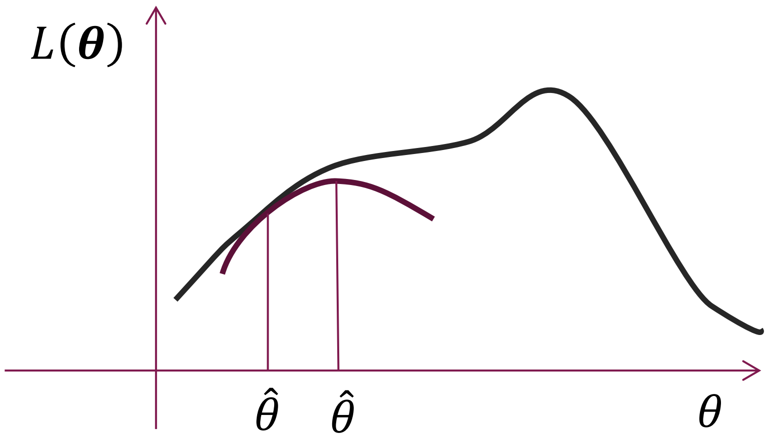
- 局所最適解を求めるので，初期値を多様化する多点探索を行う

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

1. 適当な初期値を定める
2. 初期値に応じて媒介変数を求める（E）
3. 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y_{is}の期待値y_{is}^*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
  - y_{is}^*を用いてパラメータ（π_sと$P_{i|s(j)}$）を推定する
  - 推定されたπ_s と$P_{i|s(j)}$をもちいてy_{is}^*を更新する

２つのプロセスを繰り返して最尤推定

ベイズ推定

ベイズ予測分布

- パラメータ$\boldsymbol{\theta}$を確率的な変数として確率密度関数を推定する.

事前確率	事後確率	尤度	同時確率
$p(\theta)$,	$p(\theta \vert X)$,	$p(X \vert \theta)$,	$p(X, \theta)$

  などを定義する

- 無数のパラメータを考えて，モデルをパラメータの事後確率に応じて平均化する

$$\hat{p}_{bayes}(x) = \int f(x|\theta) p(\theta|X) d\theta$$

ベイズ予測分布

ただし

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)}$$

$$p(X|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$p(X) = \int \overbrace{p(X,\theta)}^{\text{同時分布}}d\theta = \int \overbrace{p(X|\theta)}^{\text{尤度}}p(\theta)d\theta = \int \prod_{i=1}^{n} f(x_i|\theta)\,p(\theta)d\theta$$

なので

$$\hat{p}_{bayes}(x) = \frac{\int f(x|\theta)\prod_{i=1}^{n} f(x_i|\theta)\,p(\theta)d\theta}{\int \prod_{i=1}^{n} f(x_i|\theta)\,p(\theta)d\theta}$$

最大事後確率推定法　MAP

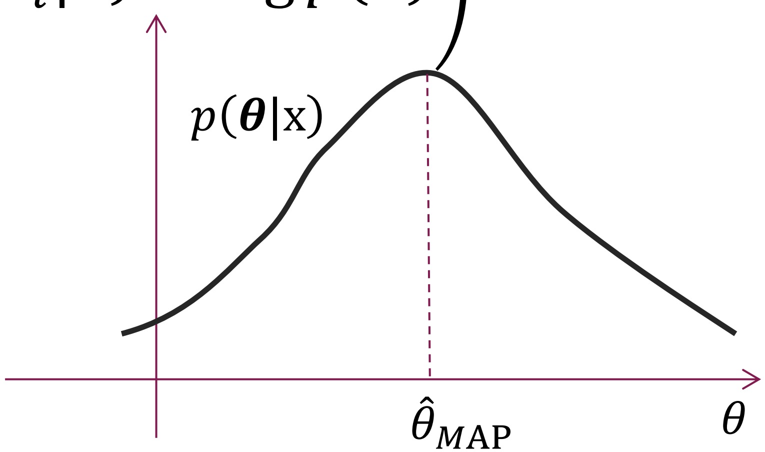
- 平均計算を省略して，事後確率の最大のパラメータを用いて，xの分布を推定する.

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\, p(\theta|X),\ \hat{p}_{MAP}(x) = f\left(x|\hat{\theta}_{MAP}\right)$$

対数化して

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} \log p(\theta|X)$$
$$= \underset{\theta}{\operatorname{argmax}} \left(\sum_{i=1}^{n} \log f(x_i|\theta) + \log p(\theta) \right)$$

- サンプルへの過適合を抑える



ベイズ推定の数値計算法

- ベイズ推定で得られる確率密度関数

$$\hat{p}_{bayes}(x) = \int f(x|\theta)p(\theta|X)d\theta$$

の積分をモンテカルロ法で近似. $p(\theta)$からの標本θ_iを用いて

$$\int f(\theta)p(\theta)d\theta \Rightarrow \frac{1}{n}\sum_{i=1}^{n} f(\theta_i)$$

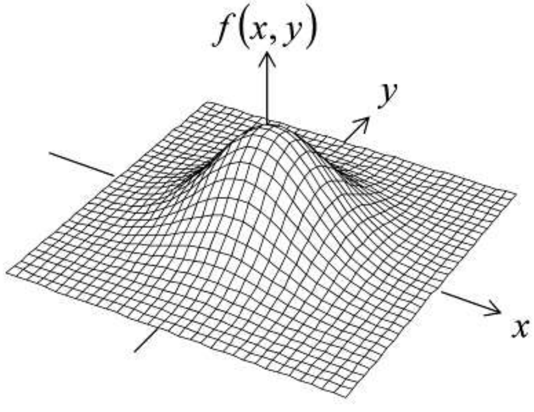
- 重点サンプリング法

$$\int f(\theta)p(\theta)d\theta = \int \left\{ f(\theta)\frac{p(\theta)}{p'(\theta)} \right\} p'(\theta)d\theta$$

計算機による疑似乱数

- MCMC法
- 前の状態に基づいて（Markov Chain）新しいパラメータをランダムにサンプリング（Monte Carlo）する
- 最尤推定と何が違う
  - 最尤推定⇒最適化＝どっかで止まる
  - MCMC⇒分布を再現＝いつまでも動く

- 乱数を使った単純操作の繰り返しで確率密度の大きいところを探しながらサンプルを生成する
  - Gibbs Sampling
  - Metropolis-Hastings Sampling



ベイズ推定におけるモデル選択

- 事前確率の設定
  - 事前確率をパラメータ(ハイパーパラメータ)でコントロールする
  - 標本を用いて（客観的に）パラメータを設定する

$$p(x|\beta) = \int \prod_{i=1}^{n} f(x_i|\theta) p(\theta|\beta)\, d\theta$$

  - この式をβの関数とした周辺尤度を最大化するβを求める.
  - あわせてモデルも周辺尤度最大化で選択

- BIC(ベイズ情報量基準)
  - $\mathrm{BIC} = -\sum_{i=1}^{n} \log f\left(x_i|\hat{\theta}_{ML}\right) + \frac{t}{2}\log(n)$
    - パラメータの事前分布もデータ数に無関係に一定とされている.
    - BICはパラメータの値に比べて最尤推定値の誤差幅が極度に小さく，有意なパラメータと，そうでないものとが，容易に識別できる状況に対応するモデルから得られている．(赤池，1996)

おわりに

マイクロ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．