BESTGUYs *(BEhavioral STudy Group on Urban activitYs)*

行動モデルの基礎
推定の方法

佐々木邦明（早稲田大学）

モデルと推定など

- 離散選択モデル
- 離散連続モデル
- 最尤推定法
- 政策分析

はじめに

- 行動は選択の帰結，行動のモデル化≒選択のモデル化
- 選択の種類
  - 連続量の選択　　　　　例）ある場所での滞在時間
  - 離散量の選択　　　　　例）交通手段，目的地，経路
- 離散量の選択を表すモデルとして，ランダム効用最大化に基づく非集計離散選択モデル

合理的選択と効用最大化

例）{A,B,C,D,E}

$(A>B), (B>C), (C>D), (D<E), (C>E)$

↓

$(A>B>C>E>D)$

効用最大化:「人は最大の効用を与える選択肢を選ぶ」

Aさん：車を選択⇔U（車）＞U（バス），U（鉄道）

提供：倉内先生

ランダム効用（1）

$$
\begin{aligned}
U(car) &= \boxed{\beta_1 + \beta_3 * time_{car} + \beta_4 * \mathrm{cost}_{car} + \beta_5 * carown} + \varepsilon_{car} \\
U(bus) &= \beta_2 + \beta_3 * time_{bus} + \beta_4 * \mathrm{cost}_{bus} + \beta_6 * age60 + \varepsilon_{bus} \\
U(rail) &= \beta_3 * time_{rail} + \beta_4 * \mathrm{cost}_{rail} + \varepsilon_{rail}
\end{aligned}
$$

確定項(V)　　誤差項

分析者にとって意思決定者がもつ真の効用は不明

→ランダム(誤差)項を用いて効用を確率的に表す

提供：倉内先生

ランダム効用（2）

誤差項に含まれるもの

●非観測属性：快適性，移動の自由度etc.

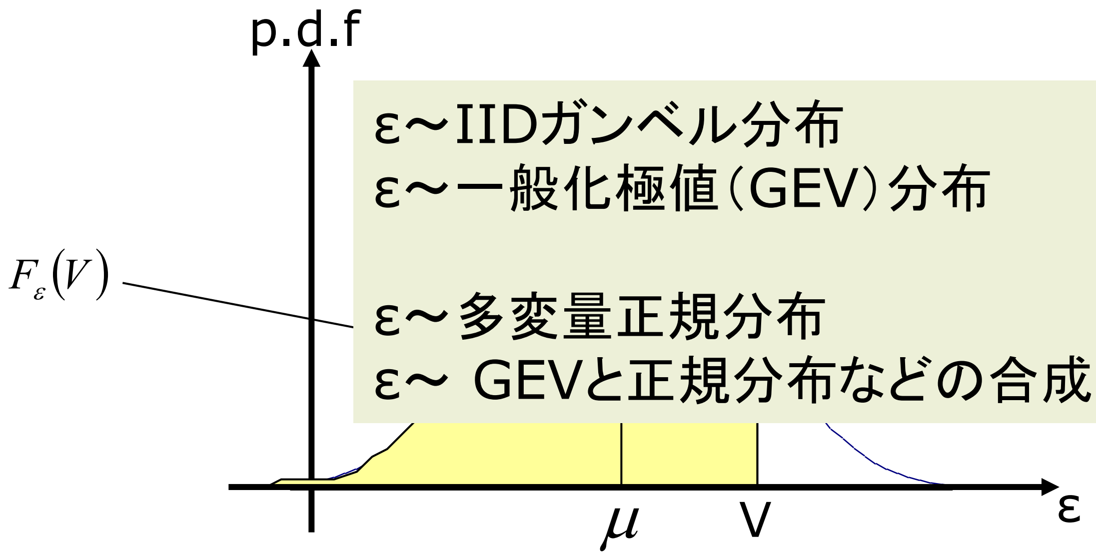
●測定誤差：駅までのアクセス時間etc.

●情報の不完全性：認知所要時間と実際の所要時間のずれetc.

●Instrumental (proxy) variables:

提供：倉内先生

誤差項の分布とモデル



$$\Pr ob(choice = car) = \Pr ob(\varepsilon < V)$$

ε～IIDガンベル分布　→　多項ロジットモデル
ε～一般化極値(GEV)分布　→　GEVモデル
(NL,PCL,CNL,GNL等)
ε～多変量正規分布　→　多項プロビットモデル
ε～ GEVと正規分布などの合成分布　→　ミックストロジットモデル

選択確率はεとVに依存

提供：倉内先生

離散選択モデル　MNL

$$U(car) = \beta X_{car} + \varepsilon_{car}$$
$$U(bus) = \beta X_{bus} + \varepsilon_{bus}$$
$$U(rail) = \beta X_{rail} + \varepsilon_{bus}$$

ε～IIDガンベル分布

$$P(i) = \frac{\exp(\mu V_i)}{\sum_{j \in C} \exp(\mu V_j)}$$

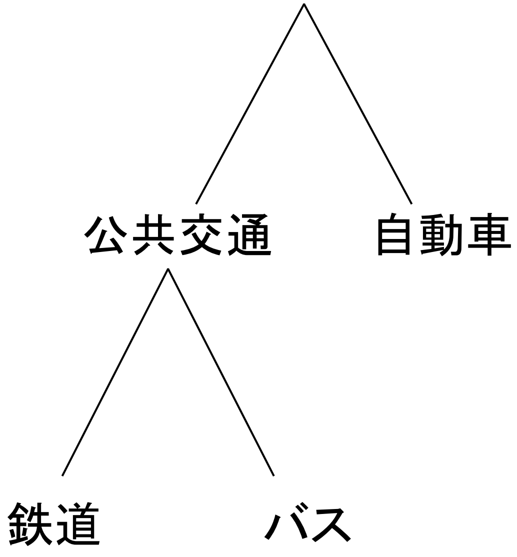
$$P(car) = \frac{\exp(\mu V(car))}{\exp(\mu V(car)) + \exp(\mu V(bus)) + \exp(\mu V(rail))}$$

Cov(U) $\begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$ $\sigma^2 = \frac{\pi^2}{6\mu^2}$

提供：倉内先生

ネスティッドロジットモデル

NLモデルの誤差構造



$Cov(U)$

$$
\begin{array}{l}
\text{鉄道} \\
\text{バス} \\
\text{自動車}
\end{array}
\begin{bmatrix}
\sigma^2 & \sigma^2_{transit} & 0 \\
\sigma^2_{transit} & \sigma^2 & 0 \\
0 & 0 & \sigma^2
\end{bmatrix}
$$

等分散

無相関

提供：倉内先生

ミックストロジット（MXL）モデル

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

ロジットモデルの操作性

IIDガンベル分布

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

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

提供：倉内先生

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

- 連続量の分析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

$$U(\boldsymbol{x})$$

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

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

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

最尤推定

行動モデルの推定と最尤推定

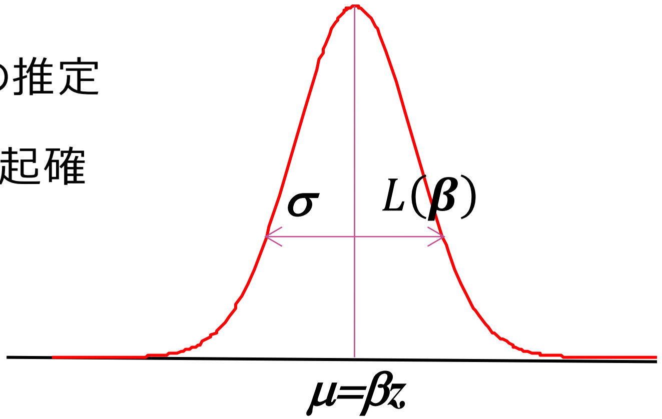
有限個のパラメータで記述される確率密度関数の推定

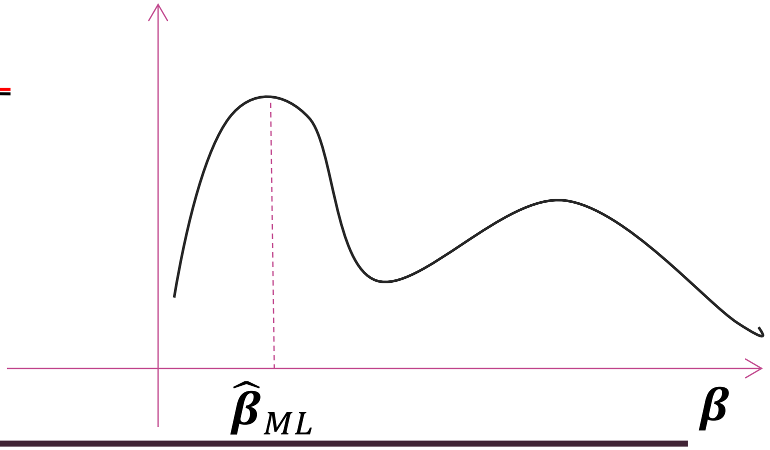
パラメータベクトル$\boldsymbol{\beta}$，モデル f による標本の生起確率を尤度とする

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

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

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





最尤推定法

点推定量を求める一般的な方法

$$\mathrm{L}(\boldsymbol{\beta}|x) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\beta})$$

上の式をβの関数とみなしたものが尤度関数

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？

視点を変えてみる

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

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

βは未知数
x は観測値

$$f(\boldsymbol{y}_i|\mu\boldsymbol{\beta}) = \prod_{j=1}^{J} \left\{ \frac{exp\left(\mu\boldsymbol{\beta}\boldsymbol{x}_{ij}\right)}{\sum_{j=1}^{J} exp\left(\mu\boldsymbol{\beta}\boldsymbol{x}_{ij}\right)} \right\}^{y_{ij}}$$

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

選択結果(y_i ： y_1＝車, y_2＝車, y_3＝鉄道, y_4＝鉄道, y_5＝鉄道, y_6＝車, ・・)が得られたとき, $\mu\boldsymbol{\beta}$がいくつだとデータへの適合がよいか？

データ(**y**)が得られやすいμ$\boldsymbol{\beta}$は？

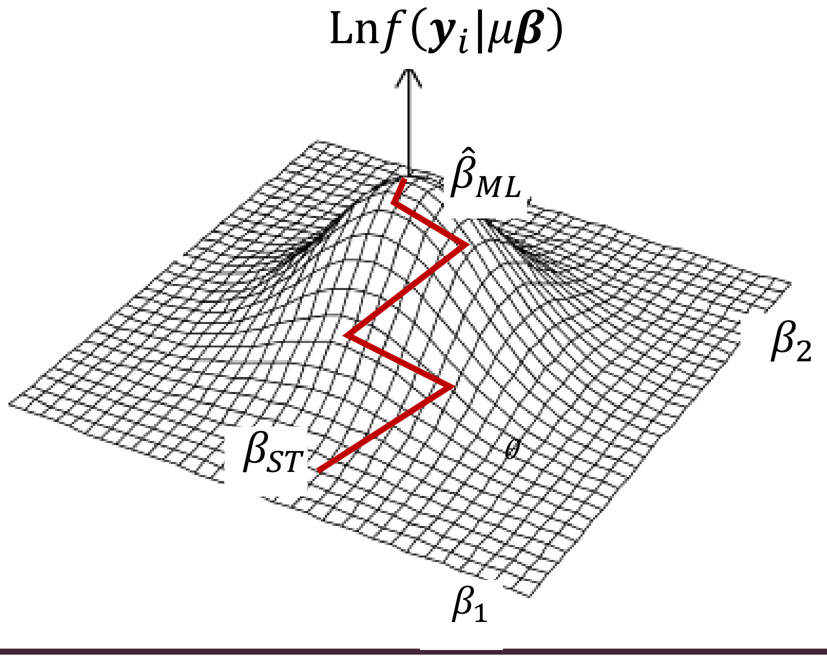
一番**L**が大きくなるμ$\boldsymbol{\beta}$を探す

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

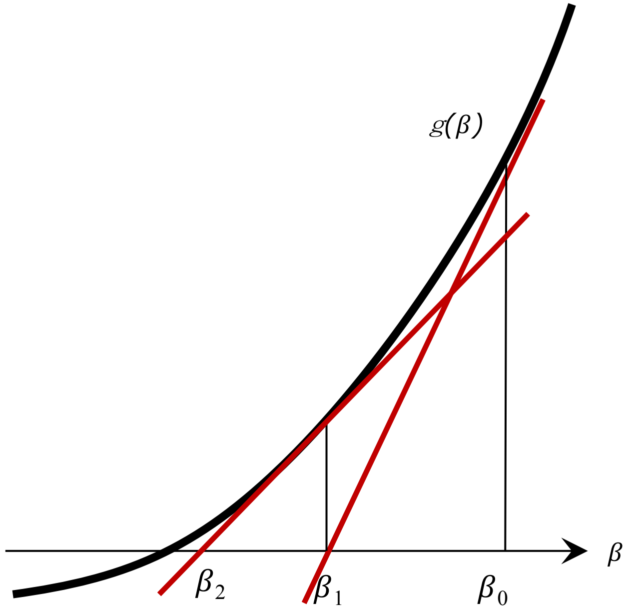
尤度関数を最大化　尤度関数の一階微分＝０を解く

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和で表す．



$$\beta_{n+1} = \beta_n - H^{-1}\beta_{\mathrm{n-1}}$$

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$\beta n+1$とβnが同じになる

- g'が0になる

収束しない

- 無限に繰り返す
- β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モデル

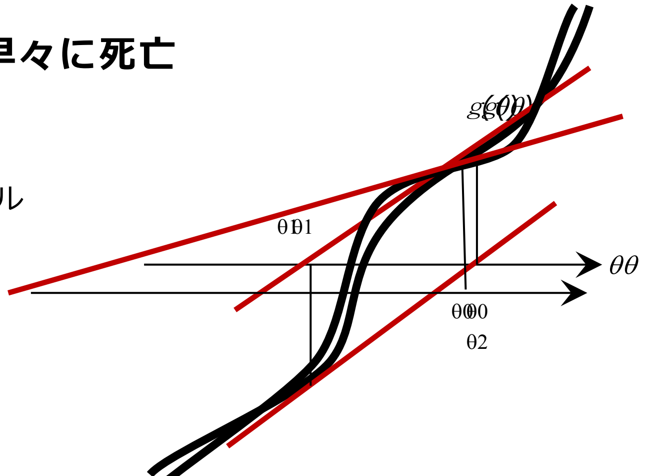
関数の近似状況

初期値の問題

モデルの問題

意思決定者間で異なるが，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



シミュレーションによる推定

シミュレーションによる尤度計算

手順

- 誤差項の密度関数から選択肢の数の次元の(準)乱数を発生させる
- この乱数を誤差の値として，各代替案の効用値を計算（積分）する
- 代替案iの効用値とその他の代替案の効用との値を比較し，それらの大小関係を1-0の変数Gで記述する.
- 1～3のステップを繰り返す．その反復回数をRとする.
- シミュレーションされた確率は $P_i = \frac{1}{R}\sum_{r=1}^{R} G^r$ となり，この値は不偏推定量である.

効用を確定値にする

確定的に選択を決定

比率を確率に置き換える

シミュレーションベースのパラメータ推定法

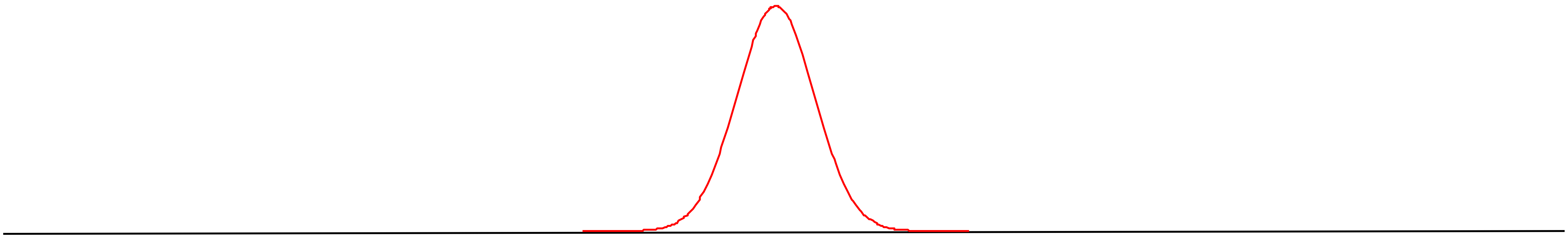
シミュレーション尤度最大化（MSL）

- シミュレーションによって計算された確率を尤度として，最大化を行う．

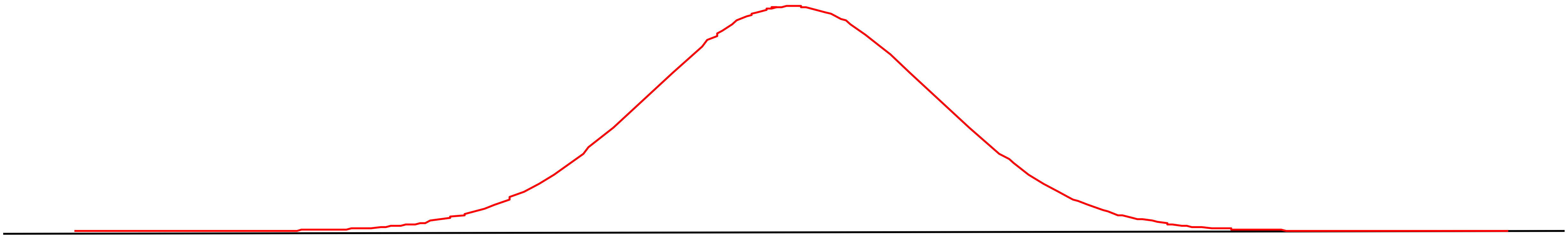
特性

- サンプル数と乱数発生回数に依存する．
- 乱数発生回数が十分大きいと一致性や漸近的有効性を持ち解析積分と一緒の特性を持つ．
- 乱数発生回数がサンプル数に対して小さく固定されると一致性もない．

非効率な乱数



非効率な乱数



乱数の効率化

確率密度が高いところ付近を重点的にサンプリングする

そもそも確率密度がどこにあるのかわかっていない問題

モデル選択

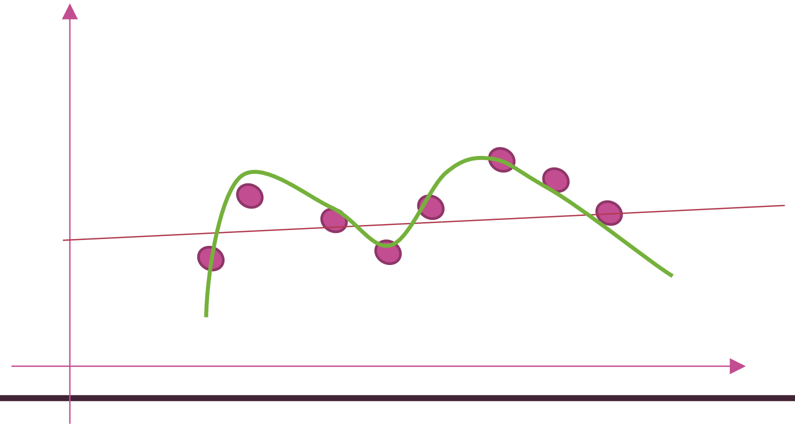
最尤推定法におけるモデル選択

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（過適合）



平均対数尤度の比較（KL情報量）

- 例：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.

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

$$\mathrm{AIC} = -\sum_{i=1}^{n} \log f\left(x_i, \hat{\theta}_{ML}\right) + t$$

パラメータ推定と過学習

機械学習における学習

- 判断の根拠となるための統計的なモデルを作る過程

機械学習の主な目的は「予測」

- ある移動手段がどの程度選ばれそうか
- ある個人が車を購入しそうか

仮説に基づく制約をモデルとして導入せず，予測精度が上がるようにモデルを自由に作る

スパース推定・モデリング

政策分析の方法

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　$S(j) = \frac{1}{N}\sum_{i=1}^{N} P_i(j)$
  最大効用の選択肢をカウント
  確率の平均値を求める

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.