推定法

理論勉強会第四回

2014/6/19

片山　元暉

そもそも推定とは？

未知の分布を持つ母集団から標本データを抽出し、そのデータの属する母集団の分布の性質を推定することである。
具体的には仮定した理論モデルのパラメータを観測した標本データに最も適合するように同定することである。

例えば母集団の分布が$y = 3x + 4$のとき

理論モデル:$y = ax + b$
パラメータ:a, b
標本データ

$a = 3$
$b = 4$

各種推定法

1. 的中率による評価
2. 最尤法
3. ベイズ推定法

1. 的中率による評価

$$\sum_{n=1}^{N} \frac{d(y_n = i_n)}{N}$$

i_n: 個人nの実際の選択
y_n : 個人nの最も選択確率の高い選択肢（効用関数の確定項が最大となる選択肢）
N : 人数
$d(y_n = i_n)$: $y_n = i_n$となったときd=1をとり、そうでないときd= 0をとる指示変数

全サンプルの選択結果と選択予測がどれだけ的中したかでモデルの適合度を評価
しかし
予測された選択結果がどれだけの確率で選択されると判断していたかを考慮しないのでモデルの適合性の評価としては不十分

２．最尤法

尤度関数：尤もらしさを表す

$$\mathrm{L}(\beta) = \prod_{n=1}^{N} \prod_{i=1}^{J} P_n(i|\beta)^{d_{in}}$$

L：尤度関数
β：パラメータ
d_{in}：個人nが選択肢iを選択したかどうかのダミー変数
$P_n(i|\beta)$：パラメータβのときに選択肢iが選択される確率

最適化手法を用いることでLを最大とするような$\hat{\beta}$を探す。

・最急降下法

・Newton-Raphson法

・BFGS法（準ニュートン法）

・BHHH法

・Nelder-Mead法

尤度比：推定モデルの評価に利用

・McFaddenの決定係数

$$\rho^2 = \frac{\ln L(0) - \ln L(\hat{\beta})}{\ln L(0)}$$

$L(\hat{\beta})$：尤度関数を最大化するパラメータ$\hat{\beta}$を採用した時の尤度関数

欠点：パラメータ数を増やすと値が増加→モデル間の比較に都合悪い

・自由度調整済み決定係数

$$\bar{\rho}^2 = \frac{\ln L(0) - (\ln L\left(\hat{\beta}\right) - K)}{\ln L(0)}$$

K：パラメータ数

赤池情報量基準を用いている。

赤池情報量基準：パラメータ数の最適化を行うモデル

３．ベイズ推定法

事前情報とデータからパラメータ値を更新していく方法

最尤法では推定結果がパラメータの値として求まる（先の場合の$\hat{\beta}$ ）のに対しベイズ推定法ではパラメータ値自体を確率変数と捉え、その分布を求める。

特徴

・極端にサンプルが少なく最尤法がうまくいかないときや、尤度関数が微分不可能な場合でも実行可能である。

・行動変化の観測やパラメータの更新にも適用可。

・尤度関数に積分型が残る場合など数値解析的に負荷の大きな最適化計算を避けられる。

３．１　ベイズの定理（1）

尤度関数

$$L(Y|\theta)=\prod_{n=1}^{N} P_n(y_n|\theta)$$

θ：パラメータ
y_n：個人nの選択結果
$P_n(y_n|\theta)$：個人nの選択結果y_nに対する選択確率
Y={$y_1, y_2, \ldots, y_n$}

条件付き確率の乗数の法則から

$$\pi(\theta|Y)\,L(Y)=L(Y|\theta)\pi(\theta)$$

$\pi(\theta)$：パラメータθに関しての事前分布
$\pi(\theta|Y)$：データYが得られて更新されたパラメータθに関しての事後分布
$L(Y)$：観測データYそのものの周辺分布

$$L(Y)=\int L(Y|\theta)\pi(\theta)d\theta$$

これより

$$\underbrace{\pi(\theta|Y)}_{\text{事後分布}}=\frac{\overbrace{L(Y|\theta)}^{\text{尤度関数}}\ \overbrace{\pi(\theta)}^{\text{事前分布}}}{\underbrace{L(Y)}_{\text{定数}}}$$

が得られこれをベイズの定理という。

３．１　ベイズの定理（２）

事後分布を解析的に求めるためには解析的な積分計算が可能とならなければならない。

そのために

事前分布と事後分布が同一の分布族となる確率分布を選ぶ（これを自然共役事前分布と呼ぶ）

尤度	事前分布	事後分布
二項分布	ベータ分布	ベータ分布
多項分布	ディリクレ分布	ディリクレ分布
正規分布	正規分布（分散は固定）	正規分布
正規分布	正規分布（分散σ^2）	ガンマ分布

３．２　事後分布の推定法

事後分布の計算では積分計算が必要となり解析的に困難

そこで

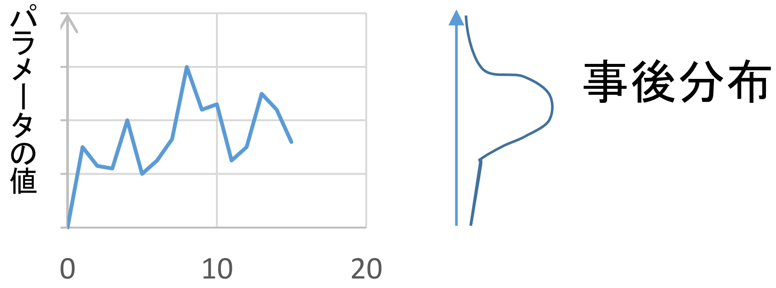
MCMC法(マルコフ連鎖モンテカルロ法)

一般的な方法

①パラメータに任意の初期値を与える

②他のパラメータを固定した時の条件付き事後確率分布からパラメータの値を順次サンプリング

③サンプルが多くなったところでパラメータの値の分布を得る



３．２．１ギブスサンプリング法

パラメータ間の条件付き確率が既知の場合に適用する。

①初期値$x^{(0)}=(x_1^{(0)},\ldots,x_k^{(0)})$を決める。

②t=0,1,…に対して次を繰り返す。（順番は関係ない）

（1）$x_1^{(t+1)}$を$\pi(x_1|x_2^{(t)},\ldots,x_k^{(t)})$からサンプリング

（2）$x_2^{(t+1)}$を$\pi(x_2|x_1^{(t+1)},x_3^{(t)}\ldots,x_k^{(t)})$からサンプリング

…

（k）$x_k^{(t+1)}$を$\pi(x_k|x_1^{(t+1)},\ldots,x_{k-1}^{(t+1)})$からサンプリング

③サンプルから事後分布を得る。

３．２．２MH法（メトロポリス・ヘイスティングス）

①初期値$\theta^{(0)}$を設定する。

②t=0,1,…に対して次を繰り返す。

（１）$x^{new(t)}$を$\pi(x^{new(t)} \mid x^{(t-1)})$からサンプリングする。

（２）$x^{new(t)}$の採択確率$\alpha(x^{(t-1)}, x^{new(t)})$を求める

$$\alpha = \min(1, \frac{\pi(x^{new(t)})\pi(x^{(t-1)} \mid x^{new(t)})}{\pi(x^{(t-1)})\pi(x^{new(t)} \mid x^{(t-1)})})$$

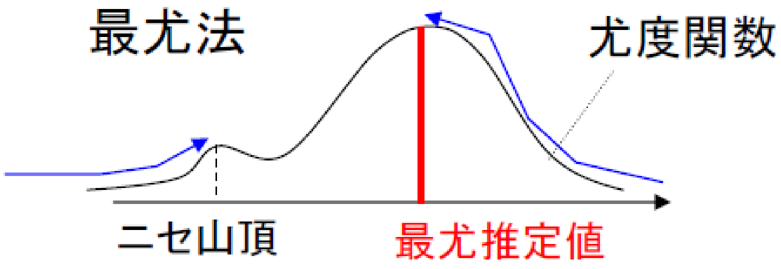
（３）$x^{(t)}$を決定する。

$$x^{(t)} = \begin{cases} x^{new(t)} \ (\text{確率}\alpha) \\ x^{(t-1)} (\text{確率}1 - \alpha) \end{cases}$$

③サンプルから事後分布を得る。

まとめ

・最尤法はパラメータ関数の山頂を求めるのに対しベイズ推定法は関数の形を求める。そのため最尤法は初期値を誤ると誤った結果が出力されることもある。ベイズ推定法はその心配はないが計算量が多い。



ひたすら山頂を目指す
⇒初期値によっては間違った結果

・ベイズ推定法は更新型の観測に適している。

・ギブスサンプリング法はパラメータ間の条件付き確率が既知の場合MH法よりも容易