ARモデルの理論と計算アルゴリズム

ARモデルとは

ARMAモデル（Autoregressive Moving Average model、自己回帰移動平均モデル）とは時系列y_nを過去の観測値y_{n-i}と白色雑音の現在および過去の値の線形和で表現したモデルをいう。

$$y_n = \sum_{i=1}^{m} a_i y_{n-i} + v_n - \sum_{i=1}^{l} b_i v_{n-i}$$

mとa_iは自己回帰の次数（order)、自己回帰係数（AR coefficient）と呼ばれ、lとb_iは移動平均の次数（order)、移動平均係数（MA coefficient）と呼ばれる。二つの次数をまとめて(m,l)次のARMAモデルと呼ぶこともある。

v_nは平均0、分散σ^2の正規分布に従う白色雑音で、時系列の過去y_{n-i}と独立と仮定、すなわち次の条件を満たす。

$$\begin{aligned}
&\mathrm{E}(v_n) = 0 \\
&\mathrm{E}(v_n^2) = \sigma^2 \\
&\mathrm{E}(v_n v_m) = 0, n \neq m \text{のとき} \\
&\mathrm{E}(v_n y_m) = 0, n > m \text{のとき}
\end{aligned}$$

ARMAモデルに従う時系列y_nはARMA過程と呼ばれる。
$l=0$と仮定して時系列をその過去の値と白色雑音v_nだけで表現したモデルをm次の自己回帰モデル（Autoregressive model、ARモデル）と呼び、実用上はほとんどこちらが用いられる。

ARモデルの推定

時系列$y_1,....,y_N$が与えられたとき、この時系列を表現するARモデル、

$$y_n = \sum_{i=1}^{m} a_i y_{n-i} + v_n$$

をあてはめるとする。パラメータは自己回帰係数と分散であり、$\theta = (a_1,...,a_m,\sigma^2)^T$と表すことにする。今次数$m$は与えられているとして、パラメータ$\theta$を最尤法で推定する。ARモデルに従う時系列$y = (y_1,....,y_N)^T$の同時分布は多変量正規分布となる。$y$の平均ベクトルは0、分散共分散行列は、ユールウォーカー方程式（Yule-Walker equation）によって定ま

る自己共分散関数C_kを用いて、

$$\Sigma = \begin{bmatrix} C_0 & C_1 & \dots & C_{N-1} \\ C_1 & C_0 & \dots & C_{N-2} \\ \dots & \dots & \dots & \dots \\ C_{N-1} & C_{N-2} & \dots & C_0 \end{bmatrix}$$

で与えられる。

従って、AR モデルの尤度は、

$$\begin{aligned} L(\theta) &= p\left(y_1,\dots,y_N \middle| \theta\right) \\ &= (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} y^T \Sigma^{-1} y\right\} \end{aligned}$$

によって計算できる。しかしこの方法では$N \times N$行列Σの逆行列や行列式の演算を伴うので、データ数Nが大きな場合には計算が困難となる。さらに最尤推定には数値的方法が必要になる。

そこで次のように条件付分布の積で尤度を表せば、尤度を効率的に計算できる。

$$\begin{aligned} L(\theta) &= p\left(y_1,\dots,y_N \middle| \theta\right) \\ &= p\left(y_1,\dots,y_{N-1} \middle| \theta\right) p\left(y_N \middle| y_1,\dots,y_{N-1},\theta\right) \\ &\dots\dots \\ &= \prod_{n=1}^{N} p\left(y_n \middle| y_1,\dots,y_{n-1},\theta\right) \end{aligned}$$

カルマンフィルタを用いる方法では右辺の各項を厳密に評価でき、最尤推定値を求めることができる。ここでは近似による方法をまとめる。また、次数mは AIC を計算し最小のもので選択する。

1. ユールウォーカー法

m次の AR モデルの自己共分散関数はユールウォーカー方程式（Yule-Walker equation）を満たす。

$$C_0 = \sum_{i=1}^{m} a_i C_i + \sigma^2$$

$$C_j = \sum_{i=1}^{m} a_i C_{j-i}$$

逆に時系列が得られると、まず標本自己共分散関数$\hat{C}_k$を計算し、これを$C_j = \sum_{i=1}^{m} a_i C_{j-i}$に

代入すると以下の一次方程式が得られる。

$$\begin{bmatrix} \hat{C}_0 & \hat{C}_1 & \dots & \hat{C}_{m-1} \\ \hat{C}_1 & \hat{C}_0 & \dots & \hat{C}_{m-2} \\ \dots & \dots & \dots & \dots \\ \hat{C}_{m-1} & \hat{C}_{m-2} & \dots & \hat{C}_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_m \end{bmatrix} = \begin{bmatrix} \hat{C}_1 \\ \hat{C}_2 \\ \dots \\ \hat{C}_m \end{bmatrix}$$

なお標本自己共分散関数$\hat{C}_k = \frac{1}{N}\sum_{n=k+1}^{N}(y_n - \hat{\mu})(y_{n-k} - \hat{\mu})$である。

この方程式の解として AR 係数の推定値$\hat{a}_i$が得られる。

さらにσ^2の推定値は、$C_0 = \sum_{i=1}^{m} a_i C_i + \sigma^2$より、

$$\hat{\sigma}^2 = \hat{C}_0 - \sum_{i=1}^{m} \hat{a}_i \hat{C}_i$$

であり、$\hat{a}_1, \dots, \hat{a}_m, \hat{\sigma}^2$はユールウォーカー推定値（Yule-Walker estimates）と呼ばれる。

a_iを係数とする AR モデルによる予測誤差分散は、

$$\mathrm{E}\left(v_n^2\right) = \mathrm{E}\left(y_n - \sum_{i=1}^{m} a_i y_{n-1}\right)^2$$

$$= C_0 - 2\sum_{i=1}^{m} a_i C_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j C_{i-j}$$

となるので、

$$\frac{\partial \mathrm{E}\left(v_n^2\right)}{\partial a_i} = -2C_i + 2\sum_{j=1}^{m} a_j C_{i-j} = 0$$

より$C_j = \sum_{i=1}^{m} a_i C_{j-i}$が得られる。従って、$C_i$を$\hat{C}_i$で置き換えて得られるユールウォーカー推定値は近似的に予測誤差分散を最小にしているものと解釈できる。

レビンソンのアルゴリズムを用いれば、1 次からM次までの AR 係数を逐次的にといて効率よく計算することができる。（詳細は省略）

また、AIC_mは、

$$\mathrm{AIC}_m = N\left(\log 2\pi\hat{\sigma}_m^2 + 1\right) + 2(m+1)$$

である。

2. PARCOR 法

（中略）

3.　最小二乗法

$\theta = \left(a_1,...,a_m,\sigma^2\right)^T$ とおくと、AR モデルの対数尤度は、$L(\theta) = \prod_{n=1}^{N} p\left(y_n \middle| y_1,...,y_{n-1},\theta\right)$ より、

$$l(\theta) = \sum_{n=1}^{N} \log p\left(y_n \middle| y_1,...,y_{n-1}\right)$$

となる。ここで m 次の AR モデルの場合、$y_{n-1},...,y_{n-m}$ の値によって y_n の分布が定まるので、$m+1$ 番目の項以降については、

$$p\left(y_n \middle| y_1,...,y_{n-1}\right) = p\left(y_n \middle| y_{n-m},...,y_{n-1}\right)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(y_n - \sum_{i=1}^{m} a_i y_{n-1}\right)^2\right)$$
$$\log p\left(y_n \middle| y_1,...,y_{n-1}\right) = -\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left(y_n - \sum_{i=1}^{m} a_i y_{n-1}\right)^2$$

となる。したがって最初の $M(M \geq m)$ 項を無視すると、

$$l(\theta) = -\frac{N-M}{2}\log 2\pi\sigma^2 - -\frac{1}{2\sigma^2}\sum_{n=M+1}^{N}\left(y_n - \sum_{i=1}^{m} a_i y_{n-i}\right)$$

によって AR モデルの対数尤度の近似値が得られる。自己回帰係数 $a_1,...,a_m$ に対して $l(\theta)$ を最大にする分散 σ^2 は、

$$\frac{\partial l(\theta)}{\partial \sigma^2} = -\frac{N-M}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\sum_{n=M+1}^{N}\left(y_n - \sum_{i=1}^{m} a_i y_{n-i}\right) = 0$$

より、

$$\hat{\sigma}^2 = -\frac{1}{N-M}\sum_{n=M+1}^{N}\left(y_n - \sum_{i=1}^{m} a_i y_{n-i}\right)$$

となりこれを $l(\theta)$ の式に代入すると対数尤度は自己回帰係数だけの関数、

$$l\left(a_1,...,a_m\right) = -\frac{N-M}{2}\log 2\pi\hat{\sigma}^2 - \frac{N-M}{2}$$

となる。ここで対数は単調増加関数なので、近似対数尤度は分散 $\hat{\sigma}^2$ の最小化で実現できることがわかる。ハウスホルダー法によって M 次までの AR モデルの最小二乗解をすべて求めることができる。

まず、

$$\mathbf{Z} = \begin{bmatrix} y_M & y_{M-1} & \cdots & y_1 \\ y_{M+1} & y_M & \cdots & y_2 \\ \cdots & \cdots & \cdots & \cdots \\ y_{N-1} & y_{N-2} & \cdots & y_{N-M} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_{M+1} \\ y_{M+2} \\ \cdots \\ y_N \end{bmatrix}$$

とし、以下の$(N-M)\times(M+1)$行列を作る。

$$\mathbf{X}=[\mathbf{Z}|y]=\begin{bmatrix} y_M & \cdots & y_1 & y_{M+1} \\ y_{M+1} & \cdots & y_2 & y_{M+2} \\ \cdots & \cdots & \cdots & \cdots \\ y_{N-1} & \cdots & y_{N-M} & y_N \end{bmatrix}$$

ハウスホルダー変換によって、

$$\mathbf{HX}=\begin{bmatrix} \mathbf{S} \\ \mathbf{O} \end{bmatrix}=\begin{bmatrix} s_{11} & \cdots & s_{1M} & s_{1,M+1} \\ & \cdots & \cdots & \cdots \\ & & s_{MM} & s_{M,M+1} \\ \mathbf{O} & & & s_{M+1,M+1} \end{bmatrix}$$

の上三角行列に変換する。
このとき、j次$(0\le j\le M)$の自己回帰モデルの残差分散および AIC は、

$$\hat{\sigma}_j^2=\frac{1}{N-M}\sum_{i=j+1}^{M+1}s_{i,M+1}^2$$
$$\mathbf{AIC}_j=(N-M)(\log 2\pi\hat{\sigma}_j^2+1)+2(j+1)$$

によって計算できる。自己回帰係数は以下をとけばよい。

$$\begin{bmatrix} s_{11} & \cdots & s_{1j} \\ & \cdots & \cdots \\ & & s_{jj} \end{bmatrix}\begin{bmatrix} a_1 \\ \cdots \\ a_j \end{bmatrix}=\begin{bmatrix} s_{1,M+1} \\ \cdots \\ s_{j,M+1} \end{bmatrix}$$

システム化のためのアルゴリズム
ここでは扱いの容易さから、ユールウォーカー法を採用する。

1. データYを読み込み、Nを設定
2. 次数mを入力
3. 標本自己共分散関数$\hat{C}_k=\dfrac{1}{N}\displaystyle\sum_{n=k+1}^{N}(y_n-\hat{\mu})(y_{n-k}-\hat{\mu})$を$k=0,1,\ldots,m$まで計算
4. $\begin{bmatrix} \hat{C}_0 & \hat{C}_1 & \cdots & \hat{C}_{m-1} \\ \hat{C}_1 & \hat{C}_0 & \cdots & \hat{C}_{m-2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{C}_{m-1} & \hat{C}_{m-2} & \cdots & \hat{C}_0 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_m \end{bmatrix}=\begin{bmatrix} \hat{C}_1 \\ \hat{C}_2 \\ \cdots \\ \hat{C}_m \end{bmatrix}$を解く。

$$\mathbf{V} = \begin{bmatrix} \hat{C}_0 & \hat{C}_1 & \dots & \hat{C}_{m-1} \\ \hat{C}_1 & \hat{C}_0 & \dots & \hat{C}_{m-2} \\ \dots & \dots & \dots & \dots \\ \hat{C}_{m-1} & \hat{C}_{m-2} & \dots & \hat{C}_0 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_m \end{bmatrix}, \mathbf{C} = \begin{bmatrix} \hat{C}_1 \\ \hat{C}_2 \\ \dots \\ \hat{C}_m \end{bmatrix}$$

とすると、

$$\mathbf{VA} = \mathbf{C}$$

$$\mathbf{A} = \mathbf{V}^{-1}\mathbf{C}$$

5. 分散 $\hat{\sigma}^2 = \hat{C}_0 - \sum_{i=1}^{m} \hat{a}_i \hat{C}_i$ を計算

6. $\mathrm{AIC}_m = N\left(\log 2\pi\hat{\sigma}_m^2 + 1\right) + 2(m+1)$ を計算