VARモデル

1 ラグオペレーター

表記を簡単にするために、ラグオペレーターという演算子を導入する。

$$L(y_t) = y_{t-1} \quad L^k(y_t) = y_{t-k}$$

2 n変数VAR(p)

$$y_t = W^T \Phi(y_t) + \epsilon_t \qquad \epsilon_t \sim \mathrm{W.N.}(\Sigma)$$

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} \quad W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{np+1,1} & w_{np+1,2} & \cdots & w_{np+1,n} \end{bmatrix} \quad \boldsymbol{\Phi}(y_t) = \begin{bmatrix} \Phi_0(y_t) \\ \Phi_1(y_t) \\ \vdots \\ \Phi_{np+1}(y_t) \end{bmatrix}$$

基底関数$\boldsymbol{\Phi}(yt)$はΦｊ(yt)が第ｊ成分のnp+1×1の列ベクトルである。各成分はラグオペレーターを用いて次のように定義される。

$$\Phi_0(y_t) = 1$$

$$\Phi_j(y_t) = \begin{cases} L(y_{jt}) & j = 1, 2, \cdots, n \\ L^2(y_{j-n,t}) & j = n+1, n+2, \cdots, 2n \\ \vdots & \\ L^p(y_{j-n(p-1),t}) & j = n(p-1)+1, n(p-1)+2, \cdots, np \end{cases}$$

3 2変量VAR(2)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \\ w_{51} & w_{52} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1 \\ \Phi_1(y_t) \\ \Phi_2(y_t) \\ \Phi_3(y_t) \\ \Phi_4(y_t) \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

$$= \begin{bmatrix} w_{11} + w_{21}\Phi_1(y_t) + w_{31}\Phi_2(y_t) + w_{41}\Phi_3(y_t) + w_{51}\Phi_4(y_t) \\ w_{12} + w_{22}\Phi_1(y_t) + w_{32}\Phi_2(y_t) + w_{42}\Phi_3(y_t) + w_{52}\Phi_4(y_t) \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

$$= \begin{bmatrix} w_{11} + w_{21}y_{1,t-1} + w_{31}y_{2,t-1} + w_{41}y_{1,t-2} + w_{51}y_{2,t-2} \\ w_{12} + w_{22}y_{1,t-1} + w_{32}y_{2,t-1} + w_{42}y_{1,t-2} + w_{52}y_{2,t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

4 n変量VAR(p)の最尤推定

VAR(p)が時刻tの状態、ytが得られる確率は

$$
\begin{aligned}
& y_t = W^T\Phi(y_t) + \epsilon_t \qquad \epsilon_t \sim \mathrm{W.N.}(\Sigma) \\
\Rightarrow\ & \epsilon_t = W^T\Phi(y_t) + y_t \\
\Rightarrow\ & P(y_t | y_{t-p}, y_{t-p+1}, \cdots, y_{t-2}, y_{t-1}; \theta) = \frac{1}{\sqrt{(2\pi)^N \det\Sigma}} \exp\left\{ -\frac{1}{2}\left(y_t - W^T\Phi(y_t)\right)^{\mathrm{T}} \Sigma^{-1} \left(y_t - W^T\Phi(y_t)\right) \right\}
\end{aligned}
$$

θはパラメータをまとめたものである。ここでこれまでのデータ、$y1, y2, \cdots, yt{-}1$が得られる確率は

$$
\begin{aligned}
& P(y_{p+1}, y_{p+2}, \cdots, y_{t-2}, y_{t-1} | y_1, y_2, \cdots, y_{p-1}, y_p; \theta) \\
& = P(y_{p+1} | y_1, y_2, \cdots, y_{p-1}, y_p; \theta) \times P(y_{p+2} | y_2, y_3, \cdots, y_p, y_{p+1}; \theta) \times \cdots \\
& \qquad\qquad \cdots \times P(y_{t-1} | y_{t-p-1}, y_{t-p}, \cdots, y_{t-3}, y_{t-2}; \theta) \\
& = \frac{1}{\sqrt{(2\pi)^N \det\Sigma}} \exp\left\{ -\frac{1}{2}\left(y_{p+1} - W^T\Phi(y_{p+1})\right)^{\mathrm{T}} \Sigma^{-1} \left(y_{p+1} - W^T\Phi(y_{p+1})\right) \right\} \times \\
& \qquad \times \frac{1}{\sqrt{(2\pi)^N \det\Sigma}} \exp\left\{ -\frac{1}{2}\left(y_{p+2} - W^T\Phi(y_{p+2})\right)^{\mathrm{T}} \Sigma^{-1} \left(y_{p+2} - W^T\Phi(y_{p+2})\right) \right\} \times \cdots \\
& \qquad \cdots \times \frac{1}{\sqrt{(2\pi)^N \det\Sigma}} \exp\left\{ -\frac{1}{2}\left(y_{t-1} - W^T\Phi(y_{t-1})\right)^{\mathrm{T}} \Sigma^{-1} \left(y_{t-1} - W^T\Phi(y_{t-1})\right) \right\} \\
& = \left(\frac{1}{\sqrt{(2\pi)^N \det\Sigma}} \right)^{t-p-1} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{t-p-1} \left(y_{p+k} - W^T\Phi(y_{p+k})\right)^{\mathrm{T}} \Sigma^{-1} \left(y_{p+k} - W^T\Phi(y_{p+k})\right) \right\}
\end{aligned}
$$

となる。

最大化すべきパラメータθの関数は

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \ln P(y_{p+1}, y_{p+2}, \cdots, y_{t-2}, y_{t-1} | y_1, y_2, \cdots, y_{p-1}, y_p; \theta) \\
&= -\frac{N(t-p-1)}{2}\ln(2\pi) - \frac{t-p-1}{2}\ln(\det\Sigma) \\
&\qquad - \frac{1}{2}\sum_{k=1}^{t-p-1} \left(y_{p+k} - W^T\Phi(y_{p+k})\right)^{\mathrm{T}} \Sigma^{-1} \left(y_{p+k} - W^T\Phi(y_{p+k})\right)
\end{aligned}
$$

4.1 wについて微分

$L(\boldsymbol{\theta})$のwに関する項は

$$\Rightarrow \quad -\frac{1}{2}\sum_{k=1}^{t-p-1}\left(y_{p+k} - W^T\Phi(y_{p+k})\right)^{\mathrm{T}}\Sigma^{-1}\left(y_{p+k} - W^T\Phi(y_{p+k})\right)$$

$$= -\frac{1}{2}\sum_{k=1}^{t-p-1}\left\{y_{p+k}^T\Sigma^{-1}y_{p+k} - y_{p+k}^T\Sigma^{-1}W^T\Phi(y_{p+k})\right.$$

$$\left. - \left(W^T\Phi(y_{p+k})\right)^T\Sigma^{-1}y_{p+k} + \left(W^T\Phi(y_{p+k})\right)^T\Sigma^{-1}W^T\Phi(y_{p+k})\right\}$$

となる。ここで第一項はwを含まないので、微分すると消える。また、共分散行列であるΣは対称行列であるので、Σ^{-1}もまた対称行列である。これを利用すると、第二項と第三項は等しいということがわかる。これにより実際に微分に関わる項は次のようになる。

$$\Rightarrow -\frac{1}{2}\sum_{k=1}^{t-p-1}\left(-2y_{p+k}^T\Sigma^{-1}W^T\Phi(y_{p+k}) + \left(W^T\Phi(y_{p+k})\right)^T\Sigma^{-1}W^T\Phi(y_{p+k})\right)$$

この式を、6.行列の微分の式（*）（⁑）を利用して解くと

$$\frac{\partial}{\partial W} = -\frac{1}{2}\sum_{k=1}^{t-p-1}\left(-2\Phi(y_{p+k})y_{p+k}^T\Sigma^{-1} + \Phi(y_{p+k})\Phi(y_{p+k})^T W\Sigma^{-1}\right)$$

$$\frac{\partial}{\partial W} = 0 \text{より}$$

$$\sum_{k=1}^{t-p-1}\Phi(y_{p+k})y_{p+k}^T\Sigma^{-1} = \sum_{k=1}^{t-p-1}\Phi(y_{p+k})\Phi(y_{p+k})^T W\Sigma^{-1}$$

両辺に右からΣをかけて

$$\sum_{k=1}^{t-p-1}\Phi(y_{p+k})y_{p+k}^T = \sum_{k=1}^{t-p-1}\Phi(y_{p+k})\Phi(y_{p+k})^T W$$

ここで、ψ、Yを次のように定める。

$$\boldsymbol{\Psi} = \begin{bmatrix} \Phi_0(y_{p+1}) & \Phi_1(y_{p+1}) & \cdots & \Phi_{np}(y_{p+1}) \\ \Phi_0(y_{p+2}) & \Phi_1(y_{p+2}) & \cdots & \Phi_{np}(y_{p+2}) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_0(y_{t-1}) & \Phi_1(y_{t-1}) & \cdots & \Phi_{np}(y_{t-1}) \end{bmatrix} \qquad Y = \begin{bmatrix} y_{p+\mathbf{1}}^T \\ y_{p+\mathbf{2}}^T \\ \vdots \\ y_{t-\mathbf{1}}^T \end{bmatrix}$$

$$\Rightarrow \quad \Psi^T Y = \Psi^T\Psi W$$

$$W_{ML} = \left(\Psi^T\Psi\right)^{-1}\Psi^T Y$$

4.2 Σについて微分

行列の公式を用いると、パラメータ$\boldsymbol{\theta}$の関数の指数部分がトレースで書き換えることができる。

$$L(\theta) = -\frac{N(t-p-1)}{2}\ln(2\pi) + \frac{t-p-1}{2}\ln\left(\det\Sigma^{-1}\right)$$

$$-\frac{1}{2}\sum_{k=1}^{t-p-1}Tr\left\{\Sigma^{-1}\left(y_{p+k} - W^T\Phi(y_{p+k})\right)\left(y_{p+k} - W^T\Phi(y_{p+k})\right)^T\right\}$$

ここで、Σの(i, j)成分をσ_ij、Σ^{-1}の(i, j)成分をλ_ijとすると

$$
\begin{aligned}
\frac{\partial L(\theta)}{\partial \sigma_{ij}} &= \frac{\partial L(\theta)}{\partial \lambda_{ij}} \times \frac{\partial \lambda_{ij}}{\partial \sigma_{ij}} \\
&= \quad -\frac{1}{2}\left\{\sum_{k=1}^{t-p-1} \frac{\partial}{\partial \lambda_{ij}} Tr\left\{\Sigma^{-1}\left(y_{p+k} - W^T\Phi(y_{p+k})\right)\left(y_{p+k} - W^T\Phi(y_{p+k})\right)^T\right\}\right. \\
&\qquad\qquad \left. -(t-p-1)\frac{\partial}{\partial \lambda_{ij}}\ln\left(\det\Sigma^{-1}\right)\right\} \times \frac{\partial \lambda_{ij}}{\partial \sigma_{ij}} \\
&= -\frac{1}{2}\left\{\sum_{k=1}^{t-p-1}\left\{\left(y_{p+k} - W^T\Phi(y_{p+k})\right)\left(y_{p+k} - W^T\Phi(y_{p+k})\right)^T\right\}_{ji} - (t-p-1)\,\sigma_{ji}\right\}
\end{aligned}
$$

$\dfrac{\partial L(\theta)}{\partial \sigma_{ij}} = 0$より

$$
\sigma_{ji} = \frac{1}{t-p-1}\sum_{k=1}^{t-p-1}\left(\left(y_{p+k} - W^T\Phi(y_{p+k})\right)\left(y_{p+k} - W^T\Phi(y_{p+k})\right)^T\right)_{ji}
$$

$$
\Sigma_{ML} = \frac{1}{t-p-1}\sum_{k=1}^{t-p-1}\left(y_{p+k} - W^T\Phi(y_{p+k})\right)\left(y_{p+k} - W^T\Phi(y_{p+k})\right)^T
$$

5 VARモデルの実装

In [1]:

```
module TSA

using Statistics
using PyCall
np = pyimport("numpy")

mutable struct var
    p::Int
    params::Array{Float64}
    sig_matrix::Array{Float64}
end

mutable struct ar
    p::Int
    params::Array{Float64}
    sig::Float64
end

# fit
function fit!( model::var, data )

    nvar = size(data)[2]

    global design_matrix = zeros( size(data)[1]-model.p, nvar*model.p+1 )
    # 定数項
    design_matrix[:,1] .+= 1
    for i in 0 : model.p-1
        design_matrix[ :, i*nvar+2 : i*nvar+nvar+1 ] += data[ model.p-i : end-1-i, : ]
    end

    data_matrix = data[ model.p+1 : end, : ]

    model.params = inv(design_matrix'design_matrix)design_matrix'data_matrix
    model.sig_matrix = (data_matrix'data_matrix
        - data_matrix'design_matrix*model.params
        - model.params'design_matrix'data_matrix
        + model.params'design_matrix'design_matrix*model.params )
        ./ ( length(data) - model.p*( nvar + 1 ) - 2 )

    return
end

function fit!( model::ar, data )

    design_matrix = zeros( size(data)[1]-model.p, model.p+1 )
    # 定数項
    design_matrix[:,1] .+= 1
    for i in 0 : model.p-1
        design_matrix[ :,i+2 ] += data[ model.p-i : end-1-i ]
    end

    data_matrix = data[ model.p+1 : end ]

    model.params = inv(design_matrix'design_matrix)design_matrix'data_matrix
    model.sig = mean( ( data_matrix .- (model.params'design_matrix')' ).^2 )

    return
end
```

```julia
# predict
function predict( model::var, data,; length::Int=1 )
    nvar = size(data)[2]
    tmp_data = data

    for i in 1:length
        rev_data = reverse( tmp_data', dims=2 )

        basis_func = [1 ; rev_data[begin:model.p*nvar]]
        pre = model.params'basis_func
                    + np.random.multivariate_normal(zeros(nvar), model.sig_matrix, size=1)'

        tmp_data = cat( tmp_data, pre', dims=1 )
    end

    return tmp_data
end

function predict( model::ar, data, ; length::Int=1 )

    tmp_data = data

    for i in 1:length
        rev_data = reverse( tmp_data )
        basis_func = [ 1; rev_data[begin:model.p] ]
        pre = model.params'basis_func + np.random.normal(0.0,np.sqrt( model.sig ) )
        push!( tmp_data, pre )
    end

    return tmp_data
end

# 区間推定
function interval_estimation( model::var, data, ;len::Int=1, iteration::Int=100 )

    resid = zeros( iteration, size(data)[2] )
    Ex = conditional_expectation( model, data, length_pre=len )
    for i in 1:iteration
        tmp = predict( model, data, length=len )[end,:] - Ex
        resid[i,:] = tmp'
        # 上記の書き方のほうがiterationが大きいとき早い
        # catを使うと10^6の時、三日たっても終わらなかった
        #resid = cat(resid, tmp', dims=1 )
        #resid = [ resid' tmp ]'
    end

    mse = mean( resid.^2, dims=1 )
    #return Ex
    # 95%区間予測
    return hcat(Ex - 1.96*np.sqrt( mse' ), Ex + 1.96*np.sqrt( mse' ))

end

function conditional_expectation( model::var, data, ;length_pre::Int=1 )
    nvar = size(data)[2]
    tmp_data = data

    for i in 1:length_pre
        rev_data = reverse( tmp_data', dims=2 )

        basis_func = [1 ; rev_data[begin:model.p*nvar] ]
```

```
        pre = model.params'basis_func

        tmp_data = cat( tmp_data, pre', dims=1 )
    end

    return tmp_data[end,:]

end
end;
```

6 行列の微分

$$x^T \Sigma W^T y \text{を} w \text{で微分}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \quad W = \begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \\ w_{13} & w_{32} \end{bmatrix} \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

$$\begin{aligned}
\left(\Sigma W^T\right)_{ij} &= \sum_k \Sigma_{ik} W^T_{kj} \\
x^T \Sigma W^T y &= \sum_i \sum_j \sum_k \Sigma_{ik} W^T_{kj} x_i y_j \\
\frac{\partial}{\partial W_{sm}} &= \sum_i \Sigma_{ik} W^T_{kj} x_i y_j \\
&= \left(\sum_i x_{i1} \Sigma_{is} \right) y_m \\
&= \left(x^T \Sigma\right)_s y_m \\
\frac{\partial x^T \Sigma W^T y}{\partial W} &= y x^T \Sigma
\end{aligned} \tag{$*$}$$

$$y^T W \Sigma W^T y \text{を} w \text{で微分}$$

$$\left(W\Sigma W^T\right)_{ij} = \sum_k \sum_l W_{ik}\Sigma_{kl}W^T_{lj}$$

$$y^TW\Sigma W^Ty = \sum_i \sum_j \sum_k \sum_l W_{ik}\Sigma_{kl}W^T_{lj}y_i yj$$

$$\frac{\partial}{\partial W_{sm}} = \sum_i \Sigma_{ik}W^T_{kj}x_i y_j$$

$$W_{sm}\text{で微分} \Rightarrow \begin{cases} ① \quad i = s, j = s, k = m, l = m \quad \Rightarrow \quad 2W_{sm}y_s^2\Sigma_{mm} \\ ② \quad i = s, j \neq s, k = m, l \neq m \quad \Rightarrow \quad \sum_{j\neq s}\sum_{l\neq m} W_{jl}\Sigma_{sl}y_s y_j \\ ③ \quad i \neq s, j = s, k \neq m, l = m \quad \Rightarrow \quad \sum_{i\neq s}\sum_{k\neq m} W_{ik}\Sigma_{km}y_i y_s \end{cases}$$

$\Sigma = \Sigma^T$ であることから、② = ③

$$\begin{aligned} ① + ② + ③ &= 2\left(W_{sm}y_s^2\Sigma_{mm} + \sum_{j\neq s}\sum_{l\neq m} W_{jl}\Sigma_{sl}y_s y_j\right) \\ &= 2\left(\sum_j \sum_l W_{jl}\Sigma_{ml}y_s y_j\right) \\ &= 2\left(\sum_j \left(\sum_l W_{jl}\Sigma_{ml}\right) y_s y_j\right) \\ &= 2\left(\sum_j (W\Sigma)_{jm} y_s y_j\right) \\ &= 2\left(\left(\sum_j y_{j1}(W\Sigma)_{jm}\right) y_s\right) \\ &= 2\left(y^TW\Sigma\right)_m y_s \\ &= 2\left(yy^TW\Sigma\right)_{sm} \end{aligned}$$

$$\frac{\partial y^TW\Sigma W^Ty}{\partial W} = yy^TW\Sigma \qquad (⁑$$