時系列解析入門 – 最小二乗法

1.　回帰モデルと最小二乗法

y_n を目的変数、$x_{n1},\ldots,x_{nm}$ を説明変数とするとき、y_n の変動を説明変数の線形和で表現したモデル、

$$y_n = \sum_{i=1}^{m} a_i x_{ni} + \varepsilon_n$$

を回帰モデル（regression model）と呼ぶ。a_i は説明変数 x_{ni} に対する回帰係数（regression coefficient）、m は説明変数の個数で次数（order）とも呼ばれる。また、y_n の変動のうち説明変数の変動によっては説明できない部分 ε_n は残差（residual）と呼ばれ、平均 0、分散 σ^2 の正規分布に従う独立な確率変数と仮定する。ここで、N 次元ベクトル y（目的変数ベクトル）と $N\times m$ 行列 Z（説明変数ベクトル、デザイン行列）を、

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_N \end{bmatrix}、Z = \begin{bmatrix} x_{11} & \ldots & x_{1m} \\ x_{21} & \ldots & x_{2m} \\ \ldots & \ldots & \ldots \\ x_{N1} & \ldots & x_{Nm} \end{bmatrix}$$

と定義すると、回帰モデルは、

$$y = Za + \varepsilon$$

と表記できる。ここで $a = (a_1,\ldots,a_m)^T$ は回帰係数ベクトル、$\varepsilon = (\varepsilon_1,\ldots,\varepsilon_N)^T$ は残差ベクトル、N はデータの個数である。

この回帰モデルは、回帰係数 $a_1,\ldots,a_m$ および分散 σ^2 をパラメータとしてもっているので、

これらをまとめて $\theta = (a_1,\ldots,a_m,\sigma^2)^T$ と表すことにする。

N 組の観測値 $\{y_n, x_{n1}\ldots,x_{nm}\}(n=1,\ldots,N)$ が与えられたとき、回帰モデルの尤度および対数尤度は θ の関数となり、

$$L(\theta) = \prod_{n=1}^{N} p(y_n|\theta, x_{n1},\ldots,x_{nm})$$

$$l(\theta) = \sum_{n=1}^{N} \log p(y_n|\theta, x_{n1},\ldots,x_{nm})$$

と表される。ここで右辺各項は、

$$p\left(y_n \middle| \theta, x_{n1}, \ldots, x_{nm}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \left(y_n - \sum_{i=1}^{m} a_i x_{ni} \right)^2 \right)$$

$$\log p\left(y_n \middle| \theta, x_{n1}, \ldots, x_{nm}\right) = -\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \left(y_n - \sum_{i=1}^{m} a_i x_{ni} \right)^2$$

となるので、対数尤度は、

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

と表される。この $l(\theta)$ を最大化する θ を求めることにより、パラメータ θ の最尤推定値 $\hat{\theta} = \left(\hat{a}_1, \ldots, \hat{a}_m, \hat{\sigma}^2\right)^T$ が求められる。任意に定められた回帰係数 $a_1, \ldots, a_m$ に対してこの対数尤度を最大とする分散 σ^2 を求めるために、

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

を解くと、

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

が得られる。次にこれを前出の対数尤度の式に代入すると、対数尤度は $a_1, \ldots, a_m$ だけの関数となり、

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

で与えられる。ここで対数が単調増加関数であることを考慮すると、この対数尤度を最大とする回帰係数 $a_1, \ldots, a_m$ を求めるためには分散 $\hat{\sigma}^2$ を最小とすればよいことがわかる。こうして回帰モデルのパラメータの最尤推定値は最小二乗法（least squares method）によって求められることがわかる。

2.　ハウスホルダー法

前出の最小二乗値（残差平方和）を行列・ベクトルで表現すると以下のような簡潔な表現が可能である。

$$\sum_{n=1}^{N}\left(y_n-\sum_{i=1}^{m}a_i x_{ni}\right)^2=\|y-Za\|_N^2=\|\varepsilon\|_N^2$$

ただし$\|y\|_N$はN次元ベクトルのyのユークリッドノルムを表す。（注：ユークリッドノルムとはノルムのうちの 2-ノルムのことを指し、たとえばベクトル$y=\{y_1,\ldots,y_N\}$が与えられると、$\|y\|_N=\left(\sqrt{|y_1|^2+\ldots+|y_N|^2}\right)$となる。ここではさらにそれを二乗しているので、結局残差平方和に一致する。）

最小二乗法の解法としてよく知られているのは、$\|y-Za\|_N^2$をaで偏微分したものを 0 とおくことにより、正規方程式$Z^T y=Z^T Za$を導き、これを解いて$\hat{a}=\left(Z^T Z\right)^{-1}Z^T y$を求めるものである。

（注：ややわかりにくいので計算過程を記す。$\left(y-Za\right)^T\left(y-Za\right)=y^T y-2y^T Za+a^T Z^T Za$

aで偏微分したものを 0 とおくと、$0-2Z^T y+2Z^T Za=0$で同様の式が得られる。）

しかし、実際の計算では精度がよく操作がしやすい直交変換がしやすい以下の方法が便利である。

Uを任意の$N\times N$直交行列とすると、ベクトル$y-Za$をUで変換してもそのノルムは変化しない。つまり、

$$\|y-Za\|_N^2=\|U(y-Za)\|_N^2=\|Uy-UZa\|_N^2$$

が成り立つので、$\|Uy-UZa\|_N^2$を最小にするベクトルaと$\|y-Za\|_N^2$を最小にするaは同一である。これは、回帰係数ベクトルaの最小二乗解を求めるためには、まずはUZを都合のよい形に変換してから$\|Uy-UZa\|_N^2$を最小とするaを求めればよいことを示している。

これは、ハウスホルダー変換（Householder transformation）を利用して効率よく実現できる。まず説明変数行列Zの右側に目的変数ベクトルyを付加して$N\times(m+1)$行列を作る。

$$X=[Z|y]$$

この行列Xに適当なハウスホルダー変換Uを適用すると、次のような上三角行列Sに変換できる。

$$UX = S = \begin{bmatrix} s_{11} & \dots & s_{1m} & s_{1,m+1} \\ & \dots & \dots & \dots \\ & & s_{mm} & s_{m,m+1} \\ \mathbf{O} & & & s_{m+1,m+1} \end{bmatrix}$$

このとき、S の$1 \sim m$ 列はUZ に、$m+1$ 列はUy に対応するので、

$$\|Uy - UZa\|_N^2 = \left\| \begin{bmatrix} s_{1,m+1} \\ \dots \\ s_{m,m+1} \\ s_{m+1,m+1} \\ 0 \\ \dots \\ 0 \end{bmatrix} - \begin{bmatrix} s_{11} & \dots & s_{1m} \\ & \dots & \dots \\ & & s_{mm} \\ & & \\ & \mathbf{O} & \\ & & \end{bmatrix} \begin{bmatrix} a_1 \\ \dots \\ a_m \end{bmatrix} \right\|_N^2$$

$$= \left\| \begin{bmatrix} s_{1,m+1} \\ \dots \\ s_{m,m+1} \end{bmatrix} - \begin{bmatrix} s_{11} & \dots & s_{1m} \\ & \dots & \dots \\ \mathbf{O} & & s_{mm} \end{bmatrix} \begin{bmatrix} a_1 \\ \dots \\ a_m \end{bmatrix} \right\|_N^2 + s_{m+1,m+1}^2$$

が成り立つ。ここで最右辺の第 2 項$s_{m+1,m+1}^2$ はa の値に依存せず一定の値をとる。したがって第 1 項を最小すなわち 0 とするベクトル$a = (a_1, \dots, a_m)^T$ が最小二乗解となる。これはa の最小二乗推定値が一次方程式、

$$\begin{bmatrix} s_{11} & \dots & s_{1m} \\ & \dots & \dots \\ \mathbf{O} & & s_{mm} \end{bmatrix} \begin{bmatrix} a_1 \\ \dots \\ a_m \end{bmatrix} = \begin{bmatrix} s_{1,m+1} \\ \dots \\ s_{m,m+1} \end{bmatrix}$$

の解として求められることを示している。この一次方程式は左辺の行列が上三角なので後退代入により簡単に解くことができる。すなわち、

$$\hat{a}_m = \frac{s_{m,m+1}}{s_{mm}}$$

$$\hat{a}_i = \frac{s_{i,m+1} - s_{i,i+1} - \hat{a}_{i+1} \dots - s_{i,m}\hat{a}_m}{s_{ii}}, i = m-1, \dots, 1$$

とする。また$s_{m+1,m+1}^2$ が残差ベクトルの長さの 2 乗を与えるので、m 次の回帰モデルの残差分散σ^2 の推定値は、

$$\sigma_m^2 = \frac{s_{m+1,m+1}^2}{N}$$

で求められる。

（注：行列キーワード＝ハウスホルダー変換、直交行列、後退代入）