周期時系列の統計解析

（12）横断面データの主成分分析

nino

2019 年 6 月 4 日

前報では，コサイン関数モデルを用いて周期時系列データの主成分分析について検証したが，横断面データ（クロスセクションデータ）に主成分分析を適用する場合も多い．ここでは，コサイン関数モデルを用いた横断面データの主成分分析について検討した．

周期時系列データの主成分分析結果の概要（参考文献１）

まず，準備として，前報のコサイン関数モデルの相関係数行列による主成分分析の検討結果の概要を以下に示す．コサイン関数モデルは次の変数u_1,u_2を用い，位相は$\alpha > \beta \geqq 0$とした．

$$\begin{aligned} u_1 &= \cos(\theta+\alpha) \\ u_2 &= \cos(\theta+\beta) \end{aligned} \tag{1}$$

相関係数行列による主成分分析では，u_1,u_2を基準化した次の変数x_1, x_2を用いる．ここで，n：サンプル数，$\sigma_{n\text{-}1}$：不偏標準偏差である．

$$\begin{aligned} x_1 &= u_1/\sigma_{n-1} = \cos(\theta+\alpha)/\sigma_{n-1} \\ x_2 &= u_2/\sigma_{n-1} = \cos(\theta+\beta)/\sigma_{n-1} \end{aligned} \tag{2}$$

変数x_1, x_2の第 1 主成分z_1と第 2 主成分z_2は，固有ベクトルを$(l_1,\ l_2)$と$(m_1,\ m_2)$とすると，次式で表される．

$$\begin{aligned} z_1 &= l_1x_1 + l_2x_2 \\ z_2 &= m_1x_1 + m_2x_2 \end{aligned} \tag{3}$$

固有ベクトルが$(l_1,\ l_2) = (1/\sqrt{2},\ 1/\sqrt{2})$，$(m_1,\ m_2) = (-1/\sqrt{2},\ 1/\sqrt{2})$の場合の$z_1$，$z_2$は次式が得られた．ここで，相関係数を$r$とすると，固有値は$\lambda_1 = 1+r$，$\lambda_2 = 1-r$ である．

$$z_1 = \frac{1}{\sqrt{2}}(x_1+x_2) = \frac{\sqrt{\lambda_1}}{\sigma_{n-1}}\cos\left(\theta+\frac{\alpha+\beta}{2}\right) \tag{4}$$

$$z_2 = \frac{1}{\sqrt{2}}(-x_1+x_2) = \frac{\sqrt{\lambda_2}}{\sigma_{n-1}}\sin\left(\theta+\frac{\alpha+\beta}{2}\right) \tag{5}$$

また，主成分z_iと変数x_jの因子負荷量$r_{z_ix_j}$は次式で表された．

$$r_{z_1x_1} = \sqrt{\lambda_1/2} = \sqrt{\lambda_1}l_1 \tag{6}$$

$$r_{z_1x_2} = \sqrt{\lambda_1/2} = \sqrt{\lambda_1}l_2 \tag{7}$$

$$r_{z_2x_1} = -\sqrt{\lambda_2/2} = \sqrt{\lambda_2}m_1 \tag{8}$$

$$r_{z_2x_2} = \sqrt{\lambda_2/2} = \sqrt{\lambda_2}m_2 \tag{9}$$

具体例として，式(2)のθを離散変数θ_t＝360°t/nとし，サンプリング間隔＝15°で２周期分(n＝48，t＝0, 1, 2, ・・・, 47)，そしてα＝60°，β＝0°の場合について考察する．この例では，１周期＝360°/15°=24hrとすると，1hr間隔で，2日間（n＝2×24＝48hr）の時系列を表し，x_1はx_2より($\alpha-\beta$)/15°＝60°/15°＝4hr進みとなる．なお，x_1とx_2の相関係数rは位相差のコサイン：r＝cos($\alpha-\beta$)＝cos(60°)＝0.5＞0である(参考文献２)．

具体例の主成分分析結果（固有値，寄与率，累積寄与率）を表１に，因子負荷量を表２に示した．図１には，x_1，x_2およびz_1，z_2の主成分得点の時系列を示した．

表１

	z_1	z_2
固有値	1.50	0.50
寄与率	0.75	0.25
累積寄与率	0.75	1.00

表２

因子負荷量	z_1	z_2
x_1	0.866	−0.500
x_2	0.866	0.500

図１

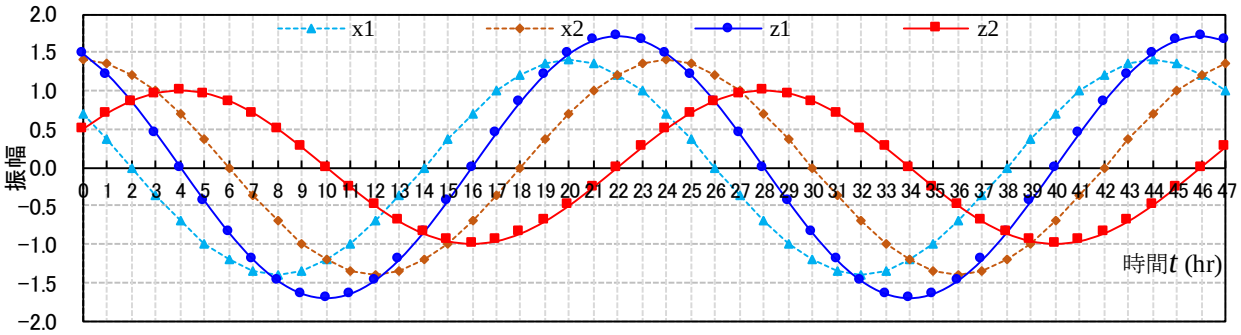


表１と表２および図１は，理論式の値と一致しており，主成分分析は変数x_1，x_2の位相関係で解釈することができる．なお，この例では，$\sigma_{n-1}=\sigma_{47}\fallingdotseq 0.7146$である．

時系列データの横断面データへの変換

図１は時間tの関数であるx_1，x_2およびz_1，z_2の時系列図であるが，それらを横断面データとした場合における座標X(x_1,x_2)と座標Z(z_1,z_2)を散布図にプロットした(図２)．

図２

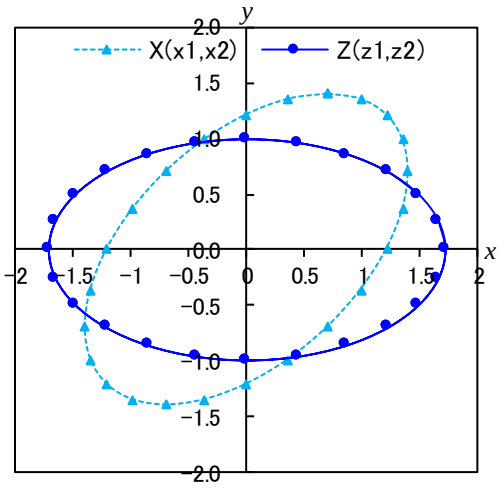


図２によると，座標X(x_1,x_2)と座標Z(z_1,z_2)は同サイズの楕円で，後者は前者を原点を中心に時計方向に45度回転させたものとみられる．

そのことを確認するため，座標の回転について検討した．座標(x_1,x_2)を反時計方向にϕ度回転させた新しい座標(z_1,z_2)は次式で表される（参考文献３）．

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{10}$$

式(10)にϕ＝－45°（時計方向に45度回転)を代入すると，次式が得られる．

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{11}$$

式(11)は，固有ベクトル(l_1, l_2)＝($1/\sqrt{2}$, $1/\sqrt{2}$)，(m_1, m_2)＝($-1/\sqrt{2}$, $1/\sqrt{2}$)と一致した．したがって，主成分分析は散布図において座標X(x_1,x_2)に時計方向に45度回転の操作を行い新しい座標Z (z_1,z_2)を求める手法といえる．また，図２から明らかなように，この操作は分散を最大化したz_1を求めることを意味している．

楕円のベクトル成分表示

変数x_1，x_2と主成分z_1，z_2は横断面データでは楕円として表された．楕円の性質はベクトル成分表示によりわかりやすくなる．その概要を以下に示した（参考文献４，５）．

楕円の媒介変数を$x=a\cos\theta$, $y=b\sin\theta$とした場合における楕円の相関係数rは，次式で表される．ここで，aは楕円の長半径，bは短半径である．

$$r = \frac{a^2 - b^2}{a^2 + b^2} \tag{12}$$

式(12)は，次のように書き直すことができる．

$$r = \frac{aa + b(-b)}{\sqrt{a^2 + b^2}\sqrt{a^2 + (-b)^2}} \tag{12}'$$

２つのベクトルを$\vec{P}(a,b)$，$\vec{Q}(a,-b)$とすると，式(12)’における右辺の分子はベクトル$\vec{P}$, $\vec{Q}$の内積を表し，分母はそれぞれのベクトルの長さの積を表している．一方，相関係数は位相差のコサインで表されるので，相関係数rはベクトル$\vec{P}$，$\vec{Q}$のなす角度のコサインcos(∠$P0Q$)に等しい．したがって，式(12)’は内積の定義を楕円の長半径と短半径を用いて示したものといえる．

このことを，先のコサイン関数モデルの主成分分析結果に当てはめてみる．２つのベクトル$\vec{P}(a,b)$，$\vec{Q}(a,-b)$において，aとbをそれぞれ式(4)の振幅$\sqrt{\lambda_1}/\sigma_{n-1}$と式(5)の振幅$\sqrt{\lambda_2}/\sigma_{n-1}$に置き換えれば，２つのベクトルは$\vec{P}(\sqrt{\lambda_1}/\sigma_{n-1}, \sqrt{\lambda_2}/\sigma_{n-1})$，$\vec{Q}(\sqrt{\lambda_1}/\sigma_{n-1}, -\sqrt{\lambda_2}/\sigma_{n-1})$となり，主成分$z_1$,$z_2$をベクトルの成分表示で示すことができる（図３）．

図中には，直線の方程式：$x=\pm\sqrt{\lambda_1}/\sigma_{n-1}$および$y=\pm\sqrt{\lambda_2}/\sigma_{n-1}$を点線で示し，$\angle Pox=\angle Qox=\mu$ とした．また，内積の定義より$\angle PoQ=(\alpha-\beta)$であるから，$\mu=(\alpha-\beta)/2$となる．

図３

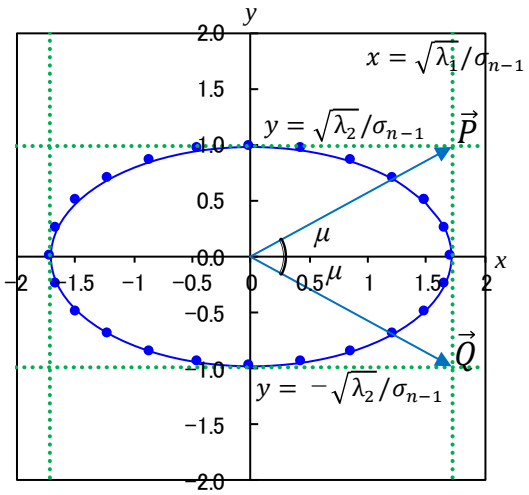


図３によると，ベクトル$\vec{P}$およびベクトル$\vec{Q}$のx軸への射影は主成分z_1の振幅に相当し，y軸への射影は主成分z_2の振幅に相当する．したがって，時系列データの主成分z_1とz_2の振幅は散布図ではそれぞれ楕円の長半径$\sqrt{\lambda_1}/\sigma_{n-1}$と短半径$\sqrt{\lambda_2}/\sigma_{n-1}$に等しい．

一方，式(12)は楕円の長半径と短半径の比b/aを用いると次式となる．

$$r=\frac{1-(b/a)^2}{1+(b/a)^2} \qquad (12)''$$

式(12)”は，相関係数はb/a比で規定され，楕円が相似ならば相関係数は等しいことを示している．また，b/a比が0に近づくほど，相関は高くなる．
さらに，式(12)”を書き換えると次式が得られる．相関係数からb/a比が求まり，そのb/a比から簡単に横断面データに対応した楕円曲線が得られる．

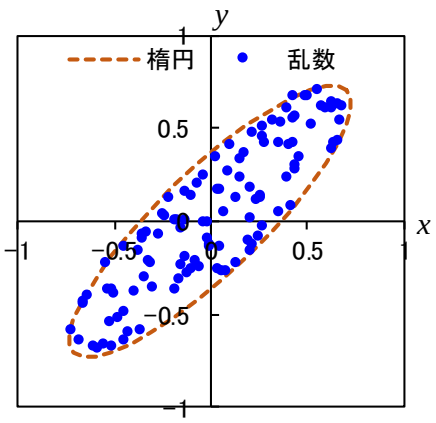
$$\frac{b}{a}=\sqrt{\frac{1-r}{1+r}} \qquad (13)$$

横断面データの具体例の主成分分析

前報では時系列データの具体例として，太陽高度と気温の観測値を用いて検討した．しかし，この２変数を横断面データの具体例として主成分分析を行っても，楕円曲線に類似したデータにしかならない．通常，散布図では楕円内にほぼ均等に分布するような横断面データを扱うことが多いので，模擬的に楕円内に生成させた乱数を具体例として用いることとした．

乱数データは，参考文献４で報告したのと同じデータ，すなわち楕円（長半径：$a=1$，短半径：$b=0.268$，相関係数：$r=\cos 30^\circ=0.866$）内に生成させた乱数（112個のデータ）を反時計方向に45度回転させたものを用いた（図４）．

図４



乱数データに主成分分析を適用した結果を表３，表４に示した．表中の（）内の数値は楕円の主成分分析結果（理論値）である．なお，乱数データの相関係数rは0.873（楕円では0.866）を示し，乱数データのb/a比は式(13)から0.260（楕円では0.268）が得られた．また，図５に，乱数データの主成分分析結果（分布）を示し，あわせてb/a比が0.260の楕円曲線を示してある．

表 3

	z_1	z_2
固有値	1.873 (1.866)	0.127 (0.134)
寄与率	0.936 (0.933)	0.064 (0.067)

表４

因子負荷量	z_1	z_2
x_1	0.968 (0.966)	−0.252 (−0.259)
x_2	0.968 (0.966)	0.259 (0.259)

図５

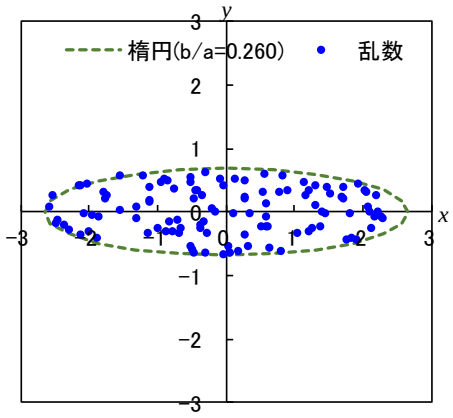


表３，表４によると，乱数データの固有値と因子負荷量は，楕円の理論値とほぼ同じ値を示した．また，図５の乱数分布は，変数を基準化しているため，図４とはx軸とy軸のスケールが異なっているが，図４の乱数分布を時計回りに45度回転させたものと一致した．このことは，横断面データの主成分分析結果は，その相関係数に対応したb/a比をもつ楕円に主成分分析を適用した結果と一致することを示唆している．

これまで述べてきたように，コサイン関数モデルにより時系列データと横断面データを関連付けることができたが，留意すべき点がある．それは，時系列データでは変数と主成分の関係が位相によって明確に解釈できるが（参考文献１），横断面データでは位相に関する情報が失われるあるいは読み取り困難になることである．時系列データは時系列データとして主成分分析することが望ましい．

参考文献

１．物理のかぎしっぽ，周期時系列の統計解析（11)コサイン関数モデルによる主成分分析の検証

http://hooktail.sub.jp/contributions/shuki11.pdf
２．物理のかぎしっぽ，周期時系列の統計解析(1)相関係数と位相差
http://hooktail.sub.jp/contributions/shuki01.pdf
３．高校数学の基本問題，回転移動の１次変換
http://www.geisya.or.jp/~mwm48961/kou2/linear_image3.html
４．物理のかぎしっぽ，楕円の相関係数(その1)
http://hooktail.sub.jp/contributions/ellipse_cor_01v4.pdf
５．物理のかぎしっぽ，楕円の相関係数(その2)
http://hooktail.sub.jp/contributions/ellipse_cor_02v2.pdf