主成分分析における軸の回転について

2002MM018 林 邦好

指導教員　田中 豊

1　研究の動機

多変量統計解析法には, 主成分分析という, 変量の次元を縮約する手法がある. 我々がある多変量データを目の前にした時, そのデータに潜む様相を簡単に把握したいのであれば, 上意の分析法を用いることが推奨される. 解析された結果を解釈することが容易な場合は何も問題は生じない. ただ, 出力された結果を解釈する上で, 非常に困難な場合が多々存在する. このような状況において, 主成分の軸の意味がより容易に付与できるように, 軸を回転させることが, これまでに行なわれてきている.Jolliffe は [1] において, 回転対象である主成分係数 A(固有ベクトル) を回転の対象とし,A に対して 3 つの基準化を考えている.

$$(1)\ \underset{p\times q}{\boldsymbol{A}} = \underset{p\times q}{\boldsymbol{U}}\,\underset{q\times q}{\boldsymbol{L}}^{\frac{1}{2}},(2)\ \underset{p\times q}{\boldsymbol{A}} = \underset{p\times q}{\boldsymbol{U}},(3)\ \underset{p\times q}{\boldsymbol{A}} = \underset{p\times q}{\boldsymbol{U}}\,\underset{q\times q}{\boldsymbol{L}}^{-\frac{1}{2}}$$

$$(\text{ただし}\ \underset{q\times p}{\boldsymbol{U}}^{T}\underset{p\times q}{\boldsymbol{U}} = \underset{q\times q}{\boldsymbol{I}},\ \underset{q\times q}{\boldsymbol{L}} = diag(\lambda_1,\cdots,\lambda_q))$$

本研究では,(a)3 つの基準化間の解釈の比較, ならびに (b) 回転の対象として主成分係数 A の他, いわゆる因子負荷量 W が考えられるが, 両者の内でどちらを回転すべきかについての検討を行なう.

2　回転行列について

回転法には, 直交回転と斜交回転があり, 類似の議論が斜交回転の場合にもできるが, 本研究では (1)Varimax 回転法と,(2)Quartimax 回転法と (3)Biquartimax 回転法に限定し, 直交行列を求める.

3　データ行列 X

本論文では, もともとの変量 X を平均 0, 分散 1 に標準化した場合と, 分散を 1 に基準化しない場合の 2 通りの解析を行なっている.

4　主成分係数及び主成分得点の定式化

4.1　Jolliffe の方法 [1]

主成分係数 $\underset{p\times q}{\boldsymbol{A}}$ は, 回転後は $\underset{p\times q}{\boldsymbol{B}} = \underset{p\times q}{\boldsymbol{A}}\,\underset{q\times q}{\boldsymbol{T}}$ となり, 回転後の主成分得点は, $\underset{n\times q}{\boldsymbol{F}} = \underset{n\times p}{\boldsymbol{X}}\,\underset{p\times q}{\boldsymbol{B}} = \underset{n\times p}{\boldsymbol{X}}\,\underset{p\times q}{\boldsymbol{A}}\,\underset{q\times q}{\boldsymbol{T}} = \underset{n\times q}{\boldsymbol{Z}}\,\underset{q\times q}{\boldsymbol{T}}$ となる.

4.2　SVD の方法

特異値分解 (SVD) によりデータ行列 X は, $\underset{n\times p}{\boldsymbol{X}} = \underset{n\times q}{\boldsymbol{Z}}\,\underset{q\times p}{\boldsymbol{W}}^{T} + V_{p-q}{}^{*}L_{p-q}{}^{\frac{1}{2}}U_{p-q}{}^{T}$ のように分解でき,W は因子負荷量 (行列) と呼ばれる.Jolliffe にならって 3 種類の基準化 $\underset{p\times q}{\boldsymbol{W}} = \underset{p\times q}{\boldsymbol{U}}\,\underset{q\times q}{\boldsymbol{L}}^{\frac{1}{2}}$,$\underset{p\times q}{\boldsymbol{W}} = \underset{p\times q}{\boldsymbol{U}}$,$\underset{p\times q}{\boldsymbol{W}} = \underset{p\times q}{\boldsymbol{U}}\,\underset{q\times q}{\boldsymbol{L}}^{-\frac{1}{2}}$ を考えることができる. ここで,SVD の式の両辺に $\underset{p\times q}{\boldsymbol{W}}(\underset{q\times p}{\boldsymbol{W}}^{T}\underset{p\times q}{\boldsymbol{W}})^{-1}$ を右から掛け合わせると, $\underset{n\times p}{\boldsymbol{X}}\,\underset{p\times q}{\boldsymbol{W}}(\underset{q\times p}{\boldsymbol{W}}^{T}\underset{p\times q}{\boldsymbol{W}})^{-1} = \underset{n\times q}{\boldsymbol{Z}}$ になり, $\underset{p\times q}{\boldsymbol{A}} = \underset{p\times q}{\boldsymbol{W}}(\underset{q\times p}{\boldsymbol{W}}^{T}\underset{p\times q}{\boldsymbol{W}})^{-1}$ が成り立つ. 回転後は, $\underset{n\times p}{\boldsymbol{X}} = \underset{n\times q}{\boldsymbol{Z}}\,\underset{q\times q}{\boldsymbol{T}}\,\underset{q\times q}{\boldsymbol{T}}^{T}\underset{q\times p}{\boldsymbol{W}}^{T} + V_{p-q}{}^{*}L_{p-q}{}^{\frac{1}{2}}U_{p-q}{}^{T}$ のように分解できるが, $\underset{p\times q}{\boldsymbol{W}}(\underset{q\times p}{\boldsymbol{W}}^{T}\underset{p\times q}{\boldsymbol{W}})^{-1}\underset{q\times q}{\boldsymbol{T}}$ を掛けることから, $\underset{n\times p}{\boldsymbol{X}}\,\underset{p\times q}{\boldsymbol{W}}(\underset{q\times p}{\boldsymbol{W}}^{T}\underset{p\times q}{\boldsymbol{W}})^{-1}\underset{q\times q}{\boldsymbol{T}} = \underset{n\times q}{\boldsymbol{Z}}\,\underset{q\times q}{\boldsymbol{T}}$ となり, $\underset{p\times q}{\boldsymbol{W}} \rightarrow \underset{p\times q}{\boldsymbol{W}}\,\underset{q\times q}{\boldsymbol{T}}$ の回転は, $\underset{p\times q}{\boldsymbol{A}}(\underset{p\times q}{\boldsymbol{W}}(\underset{q\times p}{\boldsymbol{W}}^{T}\underset{p\times q}{\boldsymbol{W}})^{-1}) \rightarrow \underset{p\times q}{\boldsymbol{A}}\,\underset{q\times q}{\boldsymbol{T}}$ の回転をもたらすことが分かる. また, $\underset{p\times q}{\boldsymbol{A}} = \underset{p\times q}{\boldsymbol{W}}(\underset{q\times p}{\boldsymbol{W}}^{T}\underset{p\times q}{\boldsymbol{W}})^{-1}$ より, $\underset{p\times q}{\boldsymbol{W}} = \underset{p\times q}{\boldsymbol{A}}(\underset{q\times p}{\boldsymbol{A}}^{T}\underset{p\times q}{\boldsymbol{A}})^{-1}$ の関係も導かれ, $\underset{p\times q}{\boldsymbol{A}} \rightarrow \underset{p\times q}{\boldsymbol{A}}\,\underset{q\times q}{\boldsymbol{T}}$ の回転は $\underset{p\times q}{\boldsymbol{W}} \rightarrow \underset{p\times q}{\boldsymbol{W}}\,\underset{q\times q}{\boldsymbol{T}}$ の回転をもたらす.

5　3 つの基準化における A と W の性質

A と W の回転前と回転後の性質について把握する.

表 1　回転前後の主成分係数 A と因子負荷行列 W の直交性

回転前		回転後	
A^TA	W^TW	T^TA^TAT	T^TW^TWT
$\underset{q\times q}{\boldsymbol{L}}$	$\underset{q\times q}{\boldsymbol{L}}^{-1}$	$\underset{q\times q}{\boldsymbol{T}}^{T}\underset{q\times q}{\boldsymbol{L}}\,\underset{q\times q}{\boldsymbol{T}}$	$\underset{p\times q}{\boldsymbol{T}}^{T}\underset{q\times q}{\boldsymbol{L}}^{-1}\underset{q\times q}{\boldsymbol{T}}$
$\underset{q\times q}{\boldsymbol{I}}$	$\underset{q\times q}{\boldsymbol{I}}$	$\underset{q\times q}{\boldsymbol{I}}$	$\underset{q\times q}{\boldsymbol{I}}$
$\underset{q\times q}{\boldsymbol{L}}^{-1}$	$\underset{q\times q}{\boldsymbol{L}}$	$\underset{q\times q}{\boldsymbol{T}}^{T}\underset{q\times q}{\boldsymbol{L}}^{-1}\underset{q\times q}{\boldsymbol{T}}$	$\underset{p\times q}{\boldsymbol{T}}^{T}\underset{q\times q}{\boldsymbol{L}}\,\underset{q\times q}{\boldsymbol{T}}$

表 2　回転前後の主成分得点の無相関性

主成分係数	因子負荷量	回転前	回転後
A	W	$Var(XA)=Var(XW^T(WW^T)^{-1})$	$Var(XAT)=Var(XW^T(WW^T)^{-1}T)$
$\underset{p\times q}{U}\,\underset{q\times q}{L}^{\frac{1}{2}}$	$\underset{p\times q}{U}\,\underset{q\times q}{L}^{-\frac{1}{2}}$	$\underset{q\times q}{\boldsymbol{L}}^{2}$	$\underset{q\times q}{\boldsymbol{T}}^{T}\underset{q\times q}{\boldsymbol{L}}^{2}\underset{q\times q}{\boldsymbol{T}}$
$\underset{p\times q}{U}$	$\underset{p\times q}{U}$	$\underset{q\times q}{\boldsymbol{L}}$	$\underset{q\times q}{\boldsymbol{T}}^{T}\underset{q\times q}{\boldsymbol{L}}\,\underset{q\times q}{\boldsymbol{T}}$
$\underset{p\times q}{U}\,\underset{q\times q}{L}^{-\frac{1}{2}}$	$\underset{p\times q}{U}\,\underset{q\times q}{L}^{\frac{1}{2}}$	$\underset{q\times q}{\boldsymbol{I}}$	$\underset{q\times q}{\boldsymbol{I}}$

表 1 及び表 2 の各列は上から基準化 1 から 3 に対応するが, これらの表から明らかなように, 基準化 1 では A と W の直交性も無相関性も成り立たない. また, 基準化 2 では,A と W の直交性のみが保持される. 基準化 3 では, 主成分得点の無相関性のみが保持される. 各々の基準化を採用した場合の回転は,8 節で数値的に詳しく検討する.

6　A の回転と W の回転

5 節では, $\underset{p\times q}{\boldsymbol{A}}$ の回転 $\underset{p\times q}{\boldsymbol{A}}\,\underset{q\times q}{\boldsymbol{T}}$ は, $\underset{p\times q}{\boldsymbol{W}}$ の回転 $\underset{p\times q}{\boldsymbol{W}}\,\underset{q\times q}{\boldsymbol{T}}$ と対応していることを示した. 問題は,(1) $\underset{p\times q}{\boldsymbol{A}}$ の単純構造化を図る,(2) $\underset{p\times q}{\boldsymbol{W}}$ の単純構造化を図る, のどちらが良いかということである.　そこで,8 節では, 数値例で両者の場合の回転を行ない,(1) の場合と (2) の場合で回転後の $\underset{p\times q}{\boldsymbol{A}}$ と $\underset{p\times q}{\boldsymbol{W}}$ の V 値 (normalized Varimax criterion) の値がどのくらい異なるか, 解釈上重要なもとの変数と主成分得点との相関行列の V 値 (normalized Varimax criterion) にどのくらいの差があるか, 等について検討した.(これ

以降は $A^* = AT$,$W^* = A^*(A^{*T}A^*)^{-1}$,$W^{**} = WT$, $A^{**} = W^{**}(W^{**T}W^{**})^{-1}$ と定義する.)

7　分析手順のフローチャート

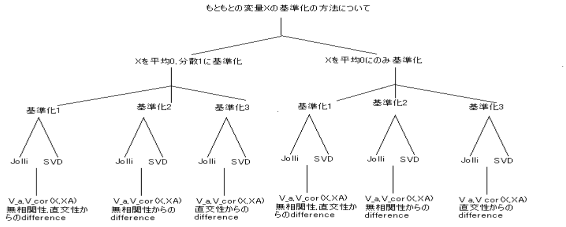

図 1　フローチャート

8　1 節の (a),(b) における解析例

8.1　食品嗜好のデータ [2](a) Varimax 回転の場合

平均 0, 分散 1 に標準化されたデータ X を基準化 1(Jolliffe) と基準化 2(Jolliffe) と基準化 3(Jolliffe) で, 回転後の主成分係数と主成分得点を比較した.(有効主成分数 3)

表 3　基準化 $1A^*$, 基準化 $2A^*$, 基準化 $3A^*$

	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
1	−0.95	−0.13	−0.09	−0.56	−0.02	0.08	−0.00	0.30	0.30
2	−0.85	−0.44	−0.14	−0.45	−0.25	0.10	−0.18	0.18	0.36
3	−0.46	−0.81	−0.20	−0.16	−0.54	0.09	−0.40	−0.06	0.35
4	−0.22	−0.90	−0.29	0.02	−0.60	0.00	−0.44	−0.17	0.22
5	0.03	−0.84	−0.46	0.21	−0.51	−0.19	−0.35	−0.23	−0.06
6	−0.91	−0.05	−0.33	−0.48	0.13	−0.15	0.14	0.32	0.00
7	−0.81	−0.19	−0.50	−0.36	0.07	−0.28	0.11	0.25	−0.13
8	−0.58	−0.31	−0.67	−0.16	0.02	−0.42	0.09	0.14	−0.32
9	−0.39	−0.45	−0.77	−0.00	−0.07	−0.50	0.03	0.04	−0.41
10	−0.18	−0.38	−0.88	0.15	0.02	−0.64	0.11	−0.01	−0.63

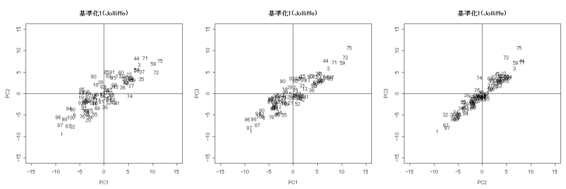

図 2　基準化 1(Jolliffe) XA^*

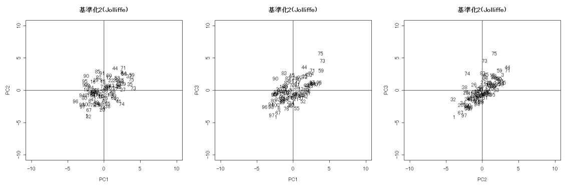

図 3　基準化 2(Jolliffe) XA^*

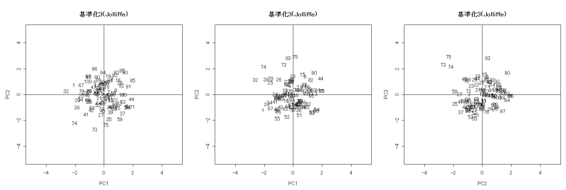

図 4　基準化 3(Jolliffe) XA^*

表 4　Harman の指標,normalized Varimax criterion

	基準化 1			基準化 2			基準化 3		
Harman	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
PC1	1.00	0.50	0.60	1.00	−0.00	−0.00	1.00	0.50	−0.52
PC2	0.50	1.00	0.70	−0.00	1.00	−0.00	0.50	1.00	0.03
PC3	0.60	0.70	1.00	−0.00	−0.00	1.00	−0.52	0.03	1.00
V_{A^*} 値		0.30			0.45			0.28	

上の表, 図から, 基準化 1 の主成分得点は無相関性が保持されず, また PC2 と PC3 は非常に構造の似た主成分に導かれている. 基準化 2 は, 軸同士の明確な分離が導かれるが, 主成分得点の相関が存在している. 基準化 3 は, 主成分得点の無相関性は保持されるが, 軸間の分離が明確ではない. ただし, 基準化 1 ほど Harman の指標で高い値を与える軸同士はない. 単純構造度 V 値も基準化 1 と基準化 3 ではあまり差は見受けられない.

8.2　食品嗜好のデータ [2](b) Varimax 回転の場合

(a) の基準化間の比較から, 基準化 2, 基準化 3 の有用性が確認される. 基準化 2 では,Jolliffe の回転からも SVD の回転からも同じ結果を得る. しかし, 基準化 3 では両者の立場で異なる. そこで, 基準化 3 内で両者の比較を行なう. 評価尺度としては,(1) 直交性からの差を測る VH 値 (VH 値とは K.Hayashi が定義した指標),(2)loadings の V 値,(3) X と rotated SC との相関値に対する V 値,(4) 主成分軸同士の類似度を測る Harman の指標を用いる.

表 5　有効主成分数 3 の場合

	Jolliffe			SVD		
Harman	PC1	PC2	PC3	PC1	PC2	PC3
PC1	1.00	0.50	−0.52	1.00	−0.14	−0.41
PC2	0.50	1.00	0.03	−0.14	1.00	−0.58
PC3	−0.52	0.03	1.00	−0.41	−0.58	1.00
V_A 値		0.28			0.27	
VH 値		0.41			0.35	
$V_{Cor(X,XA)}$ 値		0.27			0.30	

表 6　有効主成分数 4 の場合

	Jolliffe				SVD			
Harman	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
PC1	1.00	−0.09	−0.51	−0.38	1.00	−0.05	−0.27	0.47
PC2	−0.09	1.00	0.10	0.68	−0.05	1.00	−0.55	0.13
PC3	−0.51	0.10	1.00	0.33	−0.27	−0.55	1.00	0.16
PC4	−0.38	0.68	0.33	1.00	0.47	0.13	0.16	1.00
V_A 値		0.34				0.28		
VH 値		0.29				0.43		
$V_{Cor(X,XA)}$ 値		0.08				0.29		

有効主成分数 3 の場合における回転後の主成分係数の単純構造は,Jolliffe の立場からの回転と SVD の立場からの回転でほぼ同じ V 値を与える. 変量 X と回転後の主成分得点との相関値における V 値は, 後者が高くなる. 一方 Harman の指標からは, 前者の方が比較的軸同士の分離を導いていることが分かる. 有効主成分数 4 の場合は, 主成分係数の V 値に関しては,Jolliffe の立場からの回転の方が大きな値を与えるものの, 変量 X と回転後の主成分得点との相関値における V 値は,SVD の立場からの回転における $V_{Cor(X,XA)}$ 値の方が高くなり, また軸同士の分離も導いていることが示唆される.

9　研究報告

主成分の軸の解釈は,(1)loadings の単純構造と,(2) 主成分軸間の分離の 2 つを念頭に置く必要がある. このため基準化 3 における SVD の立場からの有用性が検証された.(有効主成分数 4) 固有値が類似している主成分軸の回転の場合は, 分散を 1 に基準化することで, 基準化させない場合よりも, 高い V_A,$V_{Cor(X,XA)}$ を与えた. また固有値が大きく異なる主成分軸の回転では,3 つの回転間で差が生じる. 理論的な成果としては, 基準化 3 の SVD の立場からの回転は, もともとの変量 X と主成分得点との相関係数を回転対象としているということが証明された.

参考文献

[1] IAN T.JOLLIFFE: Rotation of principal components, Journal of Applied Statistics, Vol.22, No.1, 1995

[2] 田中 豊, 脇本和昌:多変量統計解析法, 現代数学社,1983