DEA における特異値分解の活用について

02401514 東洋紡(株) *野口 弘 NOGUCHI Hiroshi
01005194 大阪大学 石井博昭 ISHII Hiroaki

1.はじめに

DEA においては、入力〔出力〕項目の値のｵｰﾀﾞｰの大きさが異なる程、DEA の目的関数の基準により入力〔出力〕項目の重み係数v〔u〕が 0となるｹｰｽ がある。ところが固有技術の立場から、入力と出力との間には何らかの関係があるのが通常で、係数が 0となることについては、解釈上困ることが多い。その一つの解決策として

(1)入力〔出力〕項目ごとの数値を、各項目ごとの標準偏差を求めてそれで割り、各項目間の数値ｵｰﾀﾞｰを揃える。

(2)v〔u〕の係数の相対的な大きさに制限を設ける領域限定法を用いる。

(3)入力〔出力〕項目間で主成分分析を用いて、入力〔出力〕項目の整理を行い、入力〔出力〕項目を選択する。等の工夫が提唱されている。

特に、多変量解析法の活用により(3)の考察報告が注目されている。それは、DEA では、nコの事業体の入力と出力の対を $(x_j, y_j)(j=1,2,\cdots,n)$ とおいた時、入力項目がmコ、出力項目をsコとすると(x,y) は(m+s) 次元空間上にあると見なすことができ、この(m+s) 次元空間を具体的な幾何学空間として捉える一つの方法として主成分分析が有用だからである。(m+s) 次元空間の幾何学的空間を考察するには特異値分解がある。特異値分解は(m+s) 次の行列を、低い階数を持つ行列による近似を行うものであり、Hottellingの主成分分析法だけでなく、図示表現法であるGabriel のﾊﾞｲﾌﾟﾛｯﾄ表現や、尺度構成法の一つであるBenzécriのｺﾚｽﾎﾟﾝﾃﾞﾝｽｱﾅﾘｼｽ がある。

本報告は、これら多変量解析手法間の特徴を論じ、その特徴に応じたDEA における(m+s) 次元の幾何学的考察の方法を提言する。その結果v(u)の係数が単に数値上の関係から 0となることを防ぎ実質的な結果をもたらすことを述べる。

2. 特異値分解を用いる多変量解析諸法の関係

n×p (p≦m+s) 行列Aが与えられたとき、Aの列ﾍﾞｸﾄﾙについて内積を考える時には、n×nの正定符号行列Mが、Aの行ﾍﾞｸﾄﾙについて内積を考える時には、p×pの正定符号行列Nが、重みとして与えられている重み付きのﾕｰｸﾘｯﾄﾞの内積を考えることにする。この時、一般化特異値分解とは、ﾗﾝｸ r (r≦p) のn×p行列Aが与えられたとき $A = U\ D_\rho V'$ と分解されることをいい $U'MU=I_r$, $V'NV=I_r$ を満たす。Uはn×r、D_ρ はr×r、V' はr×pの行列である。

(1)主成分分析でq次元までとると

$A \fallingdotseq A_{(q)} = U_{(q)}\ D_{\rho(q)} V'_{(q)}$

$U_q\ D_{\rho(q)}$ は主成分得点となり事業個体間のﾕｰｸﾘｯﾄﾞの距離を示す。また、V'_q はq次元のpコの変量項目ﾍﾞｸﾄﾙを示す。

即ち $M=(A'A)^{-1}$　$N=I_p$ である。

(2)ﾊﾞｲﾌﾟﾛｯﾄは 2次元図示表現である。主成分型ﾊﾞｲﾌﾟﾛｯﾄは 2次元までの主成分分析であり、共分散型は

$A \fallingdotseq A_{(2)} = U_{(2)}\ D_{\rho(2)} V'_{(2)}$

$F=\sqrt{n}\,U_{(2)}$ となり、Fは事業個体間のﾏﾊﾗﾉﾋﾞｽの距離を近似し、$G=1/\sqrt{n}\,D_{\rho(2)} V'_{(2)}$ はpコの変量項目間のﾕｰｸﾘｯﾄﾞ距離を近似している。これら(1)(2)のいずれも、その行列の近似誤差は

$||A_{n\times p} - A_{n\times p(q)}|| = \lambda_{q+1}{}^2 + \cdots + \lambda_p{}^2$

となる。DEA の適用は、q次元以上の固有値を捨てて進める方法や全てを取り込んで実施することもあり、これらの考察は上田[1] の報告がある。

筆者等は、Anderson検定で

$\lambda_{q+1} \fallingdotseq \lambda_{q+2} \fallingdotseq \cdots \fallingdotseq \lambda_p$ が採択できるのなら項目間と事業体間においてはきわだった特徴がないと判断できることからq+1次元以上はDEA の解析の対象から外すことも提言している。

(3)ｺﾚｽﾎﾟﾝﾃﾞﾝｽｱﾅﾘｼｽ と特異値分解について

詳細な理論的関係は磯貝・野口[2] の報告に委ねるとし、ここでは結果のみを示すと

n×pの確率行列 $A=(a_{ij})$ が与えられ時その特異値分解は

$R^{-1}AC^{-1}=UD_\rho V'$ となる。ここで、

$R=\mathrm{diag}(a_{1.}, a_{2.}, \cdots, a_{n.})$

$C=\mathrm{diag}(a_{.1}, a_{.2}, \cdots, a_{.p})$ であり、

$U'RU=I_{r+1}$, $V'CV=I_{r+1}$ の基準化条件を満たす。

$D_\rho=\mathrm{diag}(\rho_0, \rho_1, \rho_2, \cdots, \rho_q)$, $\rho_0=1$

$U=(u_{ij})=(u_0, u_1, u_2, \cdots, u_q), u_0=I_n$

$V=(v_{ij})=(v_0, v_1, v_2, \cdots, v_q), v_0=I_n$

である。

ρは $A=(a_{ij})$ の確率分布において、IとJについての2 つの実数値確率変数f(I)とg(J)の相関係数である。ρが最大になるようなｽｺｱ x, yのｽｺｱ･ﾍﾞｸﾄﾙが (u_i, v_j) である。

これらの幾何学的図示表現を考えてみると、まず行列Aについての特異値分解の結果を用いて、自明な解 $\rho_0=1, x=I_n, y=I_p$ を取り除く。そして

$R^{-1}(A-RI_n I_p{}'C)C^{-1}=UD_\rho V'$

と更めて書く。ここで自明でない解を、r=rank(A)-1 として、

$D_r=\mathrm{diag}(\rho_1, \rho_2, \cdots, \rho_r)$

$\rho_1 \geqq \rho_2 \geqq \cdots \geqq \rho_r$

$U=(u_{ij})=(u_1, u_2, \cdots, u_r)$

$V=(v_{ij})=(v_1, v_2, \cdots, v_r)$

とおくと、UとVは基準化条件を満たす。

即ち ｺﾚｽﾎﾟﾝﾃﾞﾝｽｱﾅﾘｼｽは、上からｋ番目までの解を用いて、次の行列

$U_K D_k = F$

$V_K D_k = G$

の行ﾍﾞｸﾄﾙによる同時図示表現を行うことになる。

$F' = (f_1 : f_2 : \cdots : f_n) = D_K U_k'$

$G' = (g_1 : g_2 : \cdots : g_p) = D_K V_k'$ と表すとＦは主成分分析の $U_q D_{p(q)}$ の主成分得点に対応し、Ｆの行ﾍﾞｸﾄﾙ f_i' の布置における f_i のﾕｰｸﾘｯﾄﾞの意味での距離関係 $|f_i - f_j|$ は行列

$R^{-1}(A - R I_n I_p' C) C^{-1/2}$

の、ｎコの行ﾍﾞｸﾄﾙの中の、対応する第ｉ行ﾍﾞｸﾄﾙと第ｊ行ﾍﾞｸﾄﾙのﾕｰｸﾘｯﾄﾞの意味での距離関係

$$\left\{ \sum_{s=1}^{n} \left(\frac{a_{is}}{a_{i.}\sqrt{a_{.s}}} - \frac{a_{js}}{a_{j.}\sqrt{a_{.s}}} \right)^2 \right\}^{1/2} \qquad \text{(a)}$$

を近似している。

Ｇは共分散型ﾊﾞｲﾌﾟﾛｯﾄ $D_{p(2)} V'_{(2)}$ 対応し、Ｇの行ﾍﾞｸﾄﾙ g_j' の布置における g_j のﾕｰｸﾘｯﾄﾞの意味での距離関係 $|g_i - g_j|$ は、行列

$R^{-1/2}(A - R I_n I_p' C) C^{-1}$

のｐコの列ﾍﾞｸﾄﾙの中の、対応する第ｉ列ﾍﾞｸﾄﾙと第ｊ列ﾍﾞｸﾄﾙのﾕｰｸﾘｯﾄﾞの意味での距離関係

$$\left\{ \sum_{s=1}^{n} \left(\frac{a_{si}}{a_{.i}\sqrt{a_{s.}}} - \frac{a_{sj}}{a_{.j}\sqrt{a_{s.}}} \right)^2 \right\}^{1/2} \qquad \text{(b)}$$

を近似しているといえる。(a)(b)は、共にBenzécriのｶｲ二乗距離という。Ｆ，Ｇの行ﾍﾞｸﾄﾙの同時布置を行って、Ｆ，Ｇ間の関連を調べることは、ＦとＧの行ﾍﾞｸﾄﾙ間の内積の情報をとることであり、即ち、行・列間の分布間距離を比べていることになる。この考えはPearson の χ^2 統計量の考え方と一致する。

これを、行列の近似を目的にする立場に立つなら

$R^{-1/2}(A - R I_n I_p' C) C^{-1/2}$

$= R^{1/2} U D V' C^{1/2}$ のように書換え

座標表現を行うための行列

$F = R^{1/2} U_k D_k$

$G = C^{1/2} V_k D_k$

を採用することが提言できる。

即ち、Ｆの行ﾍﾞｸﾄﾙ f_i' たちの距離関係は

$$\left\{ \sum_{s=1}^{n} \left(\frac{a_{is}}{\sqrt{a_{i.}}\sqrt{a_{.s}}} - \frac{a_{js}}{\sqrt{a_{j.}}\sqrt{a_{.s}}} \right)^2 \right\}^{1/2} \qquad \text{(c)}$$

となり、Ｇの行ﾍﾞｸﾄﾙ g_j たちの距離関係は

$$\left\{ \sum_{s=1}^{n} \left(\frac{a_{si}}{\sqrt{a_{.i}}\sqrt{a_{s.}}} - \frac{a_{sj}}{\sqrt{a_{.j}}\sqrt{a_{s.}}} \right)^2 \right\}^{1/2} \qquad \text{(d)}$$

となる。即ち、行と列とを同じ重みで調整した距離である。行と列とを同時に考察するならば、これら(c)(d)の式を用いるのがよい。

3．DEA への適用

3.1 入力項目ｍ次元と出力項目ｓ次元空間

(1)主成分分析により、事業個体間の布置を求めれば、外れ事業体の存在及び事業体の層別の有無が吟味できる。そのことにより、同質な事業体を選択してのDEA 分析が可能となる。

(2)入力項目の主成分と各々の出力項目（あるいは出力項目の主成分）との相関係数を求め、相関の高い主成分の固有ﾍﾞｸﾄﾙから入力項目の集約や整理が行える。

(3)入力項目間あるいは出力項目間に相関の強いものが含まれている場合、

負(n)の数値の扱いに決定的な方法が工夫されれば

入力主成分 Z_K，出力主成分 W_L において

各事業体ごとのｳｪｲﾄを u^T，v^T とおいて

$\left[\Sigma u^{T(p)} W_L^{(p)} + \Sigma u^{T(n)} W_L^{(n)}\right] / \left[\Sigma v^{T(p)} Z_K^{(p)} + \Sigma v^{T(n)} Z_K^{(n)}\right]$

の最大化問題に帰着することができる。

(4)第 2主成分までの累積寄与率が高い、あるいは第 3主成分以下の固有値が同じとみなせれば、ﾊﾞｲﾌﾟﾛｯﾄにより入力項目と事業固体間の関係を 2次元図示により考察することができる。

但し、 2つのﾊﾞｲﾌﾟﾛｯﾄのﾀｲﾌﾟ の違いは 2節で述べたことを認識する。

(5)もし入力項目あるいは出力項目と、事業個体間を同じ重みで調整した距離で考察したいのであれば 2節のｺﾚｽﾎﾟﾝﾃﾞﾝｽｱﾅﾘｼｽ の(c)(d)式を求めて同時布置を描けばよい。

3.2 入力項目と出力項目の(m+s) 次元空間

(1)一般に、3.1 で述べたことを、入力・出力を含めた(m+s) 次元に拡張して考察することができる

(2)第 2主成分までの累積寄与率が高い、あるいは第 3主成分以下の固有値に差がないような場合にはﾊﾞｲﾌﾟﾛｯﾄにより 2次元図示を行う。出力項目ﾍﾞｸﾄﾙ合成、入力項目ﾍﾞｸﾄﾙ合成を施すと、優秀な事業体とそうでない事業体の構造をあらかじめ推測することができ、DEA の結果の検証に役立つ。

(3)入力・出力項目、事業体の関係を同時に考察するには2 節のｺﾚｽﾎﾟﾝﾃﾞﾝｽｱﾅﾘｼｽ の(c)(d)式による同時布置が特に有効である。

3.1,3.2 の詳細な検討結果については、具体例にて別途報告する。

4．おわりに

DEA は少数の優れ者によって結果が決定するだけに、用いるﾃﾞｰﾀ の吟味が、特に重要である。その一助を担う手法として特異値分解の活用の全容を検討した。DEA の活用の範囲を拡げるには、更にｶﾃｺﾞﾘｰﾃﾞｰﾀを扱える研究が必要であると思う。

1） 上田「主成分分析を用いるDEA に関する一考察」1995年OR学会春季大会

2） 磯貝・野口「特異値分解とその応用」大阪大学教部研究集録 1992年3 月号