

特異値分解 (SVD) と多次元解析

Greenacre, M.J.

特異値分解 (SVD) は線形代数におけるもっとも便利なツールのひとつだが、統計家向け教科書の多くでは扱われていない。特異値分解の源流は、1870 年代フランス・イタリアの数学者の業績にさかのぼる (たとえば Marchall&Olkin,1979 の 19 章を参照)。その最大の適用領域である行列の低ランク近似は、Psychometrika の第一巻で Eckart&Young(1936) により報告されたのが最初である。計量心理学の分野では、彼らを称えて「Eckart-Young 分解」と呼ぶことが多い。文献においてみられる他の呼称としては、「基本構造」(Horst,1963; Green&Carroll,1976 の pp.230-240), 「正準形式」(Eckart&Young,1936;Johnson,1963), 「特異分解」(Good,1969;Kshirsagar,1972), 「テンソル縮約」(Benzécri et al., 1973) がある。おそらく「基本構造」がもっとも記述的な名前であろう。「基本」ということばが、その固有性と基礎性を含意しているし、空間の基底という幾何的な概念を示唆しているからだ (2.1 節, 2.2 節を参照)。しかし、主に歴史的な理由から、特異値分解という (いささか正当性を欠いている) ことばが、もっとも広く用いられている。この長い名前のかわりに、SVD という略語が良く使われている。

SVD は多様な多次元的手法を説明することができる。いっけん全く異なるように見えるさまざまな分析を、SVD という枠組みで統一できるのである。そのため、手法を教える際には SVD が理想的なアプローチであると私たちは考えている。本章では、SVD が主成分分析、バイプロット、正準相関分析、正準変量分析、対応分析の基盤となっていることを示す。これらの手法はすべて、ひとつのテーマのヴァリエーションなのである。そのテーマとは SVD という代数学・幾何学である。

統計学における SVD 関連文献としては、Good(1969), Chambers(1977), Gabriel(1978), Rao(1989), Mandel(1982), Greenacre&Underhill(1982) がある。

本資料は、以下を翻訳し個人的なノートをつけたものです: Greenacre, M.J.(1984) "Theory and Applications of Correspondence Analysis." Academic Press. Appendix A: Singlar Value Decomposition (SVD) and Multidimensional Analysis. なお、この本は書籍としては入手困難ですが、<http://www.carme-n.org/> で PDF が公開されています。

基本的には本文を全訳していますが、一部体裁を変えているところがあります。影つき囲みと図は私の個人的ノートです。

誤訳や誤解など、きっとあると思います。何卒ご容赦ください。(2011/01/21 小野)

数カ所にわたり計算例の誤りをご指摘いただき、修正しました。ありがとうございます。あわせて、メモをちょっぴり追記しました。(2014/09/23 小野)

誤記を修正し、R でのコード例を追記しました。なお、Clausen(著)、藤本一男(訳・解説)「対応分析入門」(オーム社)で本資料をご紹介いただきました(恐縮です...)。とてもわかりやすい本ですので、対応分析についての情報を求めて本資料に辿り着いた方は、こちらの本をご覧くださいと思います。(2016/05/17 小野)

メモをちょっぴり追記しました。(2016/05/19 小野)

R のコード例を追加しました。(2018/05/09 小野)

スペルミスを直しました。ご指摘誠にありがとうございます。(2023/09/03 小野)

1 SVD と低ランク行列近似

1.1 ユークリッド幾何学における SVD の通常の形式

■SVD とは ある行列の SVD とは、その行列を、形式が簡単で幾何的な解釈が可能な 3 つの行列の積へと分解することである。その基礎的な定理は次の通りである。任意の $I \times J$ 実数行列 \mathbf{A} は以下のように表現できる。

$$\mathbf{A} = \mathbf{U} \mathbf{D}_\alpha \mathbf{V}^\top \quad (\text{A.1.1})$$

ただし、 \mathbf{D}_α は正数 $\alpha_1 \dots \alpha_K$ からなる対角行列、 K は \mathbf{A} のランク ($K \leq \min(I, J)$) である。 \mathbf{U}, \mathbf{V} はそれぞれ $I \times K, J \times K$ の行列で、その列は (通常のユークリッド的な意味において) 正規直交である。すなわち $\mathbf{U}^\top \mathbf{U} = \mathbf{V}^\top \mathbf{V} = \mathbf{I}$ である。

例として、以下の行列を特異値分解してみる:

$$\begin{bmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{bmatrix} = \begin{bmatrix} +0.14 & -0.62 & -0.05 \\ +0.34 & +0.37 & +0.81 \\ +0.55 & +0.54 & -0.58 \\ +0.75 & -0.44 & +0.06 \end{bmatrix} \begin{bmatrix} 25.35 & 0 & 0 \\ 0 & 2.15 & 0 \\ 0 & 0 & 1.71 \end{bmatrix} \begin{bmatrix} +0.56 & +0.59 & +0.59 \\ +0.68 & +0.09 & -0.73 \\ +0.48 & -0.81 & +0.35 \end{bmatrix}$$

\mathbf{U} について調べてみると ...

$$\mathbf{U}^\top \mathbf{U} = \begin{bmatrix} +0.14 & +0.34 & +0.55 & +0.75 \\ -0.62 & +0.37 & +0.54 & -0.44 \\ -0.05 & +0.81 & -0.58 & +0.06 \end{bmatrix} \begin{bmatrix} +0.14 & -0.62 & -0.05 \\ +0.34 & +0.37 & +0.81 \\ +0.55 & +0.54 & -0.58 \\ +0.75 & -0.44 & +0.06 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

ほんとうだ、直交している。また、列内の平方和が 1 になっている。

■SVD のベクトル表現 上の式は次のように書き換えることができる。

$$\mathbf{A} = \sum_{k=1}^K \alpha_k \mathbf{u}_k \mathbf{v}_k^\top \quad (\text{A.1.2})$$

ここで $\mathbf{u}_1 \dots \mathbf{u}_K$ と $\mathbf{v}_1 \dots \mathbf{v}_K$ は \mathbf{U}, \mathbf{V} の列である。 α_k は \mathbf{A} の特異値、 \mathbf{u}_k と \mathbf{v}_k はそれぞれ \mathbf{A} の左特異ベクトル、右特異ベクトルと呼ばれている ($k = 1 \dots K$)。左特異ベクトルは、 I 次元空間における \mathbf{A} の列の正規直交基底となっており、右特異ベクトルは、 J 次元空間における \mathbf{A} の行の正規直交基底となっている。上式の形式での SVD は、「標準化された」ランク 1 の行列 $\mathbf{u}_k \mathbf{v}_k^\top$ ($k=1 \dots K$) の線形結合として解釈できる。特異値は K 「次元」のそれぞれにおけるその行列の大きさを表しているわけである。特異 (値) という名前はおそらく、行列 \mathbf{A} から抜き出した任意の項 $\alpha_k \mathbf{u}_k \mathbf{v}_k^\top$ が特異行列になるという事実由来しているのだろう。

さっきの例だと:

$$\begin{bmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{bmatrix} = 25.35 \begin{bmatrix} +0.14 \\ +0.34 \\ +0.55 \\ +0.75 \end{bmatrix} \begin{bmatrix} +0.56 & +0.59 & +0.59 \end{bmatrix} + \dots$$

■固有値分解 良く知られている固有値分解は SVD の特殊ケースである。対称な $J \times J$ 実数行列 \mathbf{B} について

$$\mathbf{B} = \mathbf{V}\mathbf{D}_\lambda\mathbf{V}^\top = \sum_{k=1}^K \lambda_k \mathbf{v}_k \mathbf{v}_k^\top$$

ただし K は \mathbf{B} のランク ($K \leq J$)、 \mathbf{V} は $J \times K$ 行列である。この場合、左特異ベクトルと右特異ベクトルは同一であり、 \mathbf{B} の固有ベクトルと呼ばれる。また特異値は固有値と呼ばれる。なお、SVD は実数行列からなり、いかなる矩形行列についても存在するのに対し、正方行列の固有値分解は、行列が非対称な場合には要素が複素数になることが多い。

■SVD の存在の証明 \mathbf{A} の SVD が存在することを証明する際には、まず次のように仮定することが多い。対称で正定な正方行列 \mathbf{B} について固有値分解 $\mathbf{B} \equiv \mathbf{A}^\top \mathbf{A} = \mathbf{V}\mathbf{D}_\lambda\mathbf{V}^\top$ が存在し、正の固有値を持つとする。この仮定の下では、 \mathbf{A} の SVD が $\mathbf{A} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^\top$ 、 \mathbf{V} が \mathbf{B} の固有ベクトル、特異値が固有値の平方根 $\mathbf{D}_\alpha = \mathbf{D}_\lambda^{1/2}$ であることを容易に示すことができる (cf. Mardia et.al., 1979, pp.473-474)。より根本的な証明においては、特異値とそれに対応した右特異ベクトル・左特異ベクトルがひとつだけ存在すると仮定し、演繹的議論によって SVD を導出する (Greenacre, 1978, Appendix A.)。

■SVD のもうひとつの定義 SVD の定義という問題に対する少し異なるアプローチが、Greenacre(1978, Appendix A.) によって示されている。このアプローチは、正方行列 \mathbf{B} の固有値 λ と固有ベクトル \mathbf{x} の通常の定義 $\mathbf{B}\mathbf{x} = \lambda\mathbf{x}$ に近い。矩形行列 \mathbf{A} について、 $\mathbf{A}\mathbf{v} = \alpha\mathbf{u}$ と $\mathbf{A}^\top\mathbf{u} = \alpha\mathbf{v}$ と同時に満たす非ゼロのスカラー α を特異値と呼び、 \mathbf{u}, \mathbf{v} を特異ベクトルと呼ぶ。ただし特異値は正であるとする (上の式は、特異値 $-\alpha$ 、特異ベクトル $\mathbf{u}, -\mathbf{v}$ の場合にも満たされるからである)。また、 \mathbf{u}, \mathbf{v} の尺度にも不定性があるが、特異ベクトルは同じユークリッド基底を持っているから ($\mathbf{u}^\top\mathbf{u} = \mathbf{v}^\top\mathbf{v}$)、どちらかのベクトルを標準化しておけば十分である (ふつうはノルムを 1 にする)。残る不定性は、 \mathbf{u}, \mathbf{v} が、それぞれの双対空間で同時に反転できてしまう点だが、これは通常なんの影響ももたらさない。異なる特異値を持った特異ベクトルがそれぞれの双対空間で互いに直交することは容易に証明できる。

1.2 SVD の一意性

ここからは、特異値は大きい順に並んでおり ($\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_K > 0$)、特異ベクトルもその順番に並んでいるものとする。特異値が厳密な不等性をもって並んでいる場合は (つまり特異値に重複がない場合は)、対応する特異ベクトルの reflection によって SVD は一意に決定できる。2 つの特異値が同一である場合、たと

例えば $\alpha_k = \alpha_{k+1}$ の場合は、対応するベクトルの対が、それぞれの 2 次元下位空間における回転によってしか決定できない。実際には、特異値が等しくなるのはめったにないが、ほぼ等しいということならば起こりうる。その場合、特異ベクトルはデータ行列のわずかな違いに対し不安定になってしまう (cf. Section 8.1)。

1.3 SVD の完全な形式

\mathbf{U} と \mathbf{V} がそれぞれの空間において持っている正規直交基底を「完全なものにし」、正方行列 $\tilde{\mathbf{U}} \equiv [\mathbf{u}_1 \dots \mathbf{u}_K \dots \mathbf{u}_I]$, $\tilde{\mathbf{V}} \equiv [\mathbf{v}_1 \dots \mathbf{v}_K \dots \mathbf{v}_J]$ を得ることが有用であることが多い。たとえば、 $\mathbf{u}_{K+1} \dots \mathbf{u}_I$ が正規直交ベクトルで、 $\mathbf{u}_1 \dots \mathbf{u}_K$ に対しても直交であるとしよう。したがって $\tilde{\mathbf{U}}^\top \tilde{\mathbf{U}} = \mathbf{I}$ である。同様に、 $\tilde{\mathbf{V}}^\top \tilde{\mathbf{V}} = \mathbf{I}$ であるとしよう。この場合 SVD は下式となる:

$$\mathbf{A} = \tilde{\mathbf{U}} \mathbf{\Delta} \tilde{\mathbf{V}}^\top \quad (\text{A.1.3})$$

ただし

$$\mathbf{\Delta} \equiv \begin{bmatrix} \mathbf{D}_\alpha & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

これを \mathbf{A} の完全 SVD と呼ぶことにする (Green & Carroll, 1976, p.234 を参照)。基底を完全なものにするやりかた、つまり \mathbf{U}, \mathbf{V} の直交な補足部分のために正規直交基底を選ぶやりかたは、通常は無限に存在する。なお、この完全な形式が SVD と呼ばれていることもある (cf. Kshirsagar, 1972, pp.247-249)。

さっきの例でいえば:

$$\begin{bmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{bmatrix} = \begin{bmatrix} +0.14 & -0.62 & -0.05 & ? \\ +0.34 & +0.37 & +0.81 & ? \\ +0.55 & +0.54 & -0.58 & ? \\ +0.75 & -0.44 & +0.06 & ? \end{bmatrix} \begin{bmatrix} 25.35 & 0 & 0 \\ 0 & 2.15 & 0 \\ 0 & 0 & 1.71 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} +0.56 & +0.59 & +0.59 \\ +0.68 & +0.09 & -0.73 \\ +0.48 & -0.81 & +0.35 \end{bmatrix}$$

$\tilde{\mathbf{U}}$ の 4 列目 \mathbf{u}_4 は、正規直交であればなんでもいいわけだ。

1.4 低ランク行列近似

■低ランク近似とは (A.1.2) の形式での SVD をみると、もし固有値 $\alpha_{K^*+1} \dots \alpha_K$ が $\alpha_1 \dots \alpha_{K^*}$ と比べて小さかったら、(A.1.2) 式の右辺の最後の $K - K^*$ 個の項を落としてしまえば、 \mathbf{A} よりも低いランクの行列で \mathbf{A} をうまく近似したことになるように思われる。この近似は実際に最小二乗近似であり、このことが SVD を非常に有用なものにしている。以下では低ランク近似の定理について述べる。なおこの定理は Eckart&Young(1936) によって最初に証明された。

■低ランク近似の原理 $\mathbf{A}_{[K^*]}$ を、 \mathbf{A} の大きいほうから K^* 個の特異値と、それに対応している特異ベクトルからつくられたランク K^* の $I \times J$ 行列 とする。すなわち $\mathbf{A}_{[K^*]} \equiv \sum_{k=1}^{K^*} \alpha_k \mathbf{u}_k \mathbf{v}_k^\top$ 。このとき $\mathbf{A}_{[K^*]}$ は、ラ

ランク K^* 以下のすべての行列 \mathbf{X} について

$$\sum_i^I \sum_j^J (a_{ij} - x_{ij})^2 = \text{trace}\{(\mathbf{A} - \mathbf{X})(\mathbf{A} - \mathbf{X})^\top\} \quad (\text{A.1.4})$$

を最小化する。この意味で、 $\mathbf{A}_{[K^*]}$ は \mathbf{A} の「ランク K^* の最小二乗近似」である。

$K^* = 2$ で試してみると、

$$\mathbf{A}_{[2]} = \begin{bmatrix} +0.14 & -0.62 \\ +0.34 & +0.37 \\ +0.55 & +0.54 \\ +0.75 & -0.44 \end{bmatrix} \begin{bmatrix} 25.35 & 0 \\ 0 & 2.15 \end{bmatrix} \begin{bmatrix} +0.56 & +0.59 & +0.59 \\ +0.68 & +0.09 & -0.73 \end{bmatrix} = \begin{bmatrix} 1.2 & 2.3 & 2.7 \\ 5.3 & 5.9 & 3.5 \\ 8.4 & 9.5 & 5.7 \\ 10.0 & 12.9 & 9.9 \end{bmatrix}$$

というわけで、なかなかいい線いっている。

■証明 上記を証明するのは難しくない。その方法のひとつは、その特殊ケースである、 \mathbf{A} が正方行列である場合の証明と同じやりかたで証明することである (たとえば以下を見よ: Kshirsagar, 1972, pp.429-430; Stewart, 1973)。この場合、 \mathbf{A} の完全 SVD(A.1.3) を用いる必要がある。目的関数 (A.1.4) は以下となる:

$$\text{trace}\{\tilde{\mathbf{U}}\tilde{\mathbf{U}}^\top(\mathbf{A} - \mathbf{X})\tilde{\mathbf{V}}\tilde{\mathbf{V}}^\top(\mathbf{A} - \mathbf{X})^\top\} = \text{trace}\{(\mathbf{\Delta} - \mathbf{G})(\mathbf{\Delta} - \mathbf{G})^\top\} = \sum_k^K (\alpha_k - g_{kk})^2 + \sum_k^K \sum_{l \neq k}^K g_{kl}^2$$

ここで \mathbf{G} は $I \times J$ 行列であり $\mathbf{G} \equiv \tilde{\mathbf{U}}^\top \mathbf{X} \tilde{\mathbf{V}}$ である。 \mathbf{G} は \mathbf{X} と同じランクをもっているから、あきらかに、最適の \mathbf{G} は対角要素が $\alpha_1 \dots \alpha_{K^*}$ 、他の要素が 0 である行列であり、このときに $\mathbf{X} = \mathbf{A}_{[K^*]}$ が最適となる。

■近似の質 $\mathbf{U}, \mathbf{D}_\alpha, \mathbf{V}$ の部分行列の記号を導入しよう。

$$\begin{aligned} \mathbf{U} &\equiv [\mathbf{U}_{(K^*)} \quad \mathbf{U}_{(K-K^*)}] \\ \mathbf{D}_\alpha &\equiv \begin{bmatrix} \mathbf{D}_{\alpha(K^*)} & \mathbf{O} \\ \mathbf{O} & \mathbf{D}_{\alpha(K-K^*)} \end{bmatrix} \\ \mathbf{V} &\equiv [\mathbf{V}_{(K^*)} \quad \mathbf{V}_{(K-K^*)}] \end{aligned}$$

この記号を用いると、 $I \times K^*$ 行列 $\mathbf{U}_{(K^*)}$ 、 $K^* \times K^*$ 行列 $\mathbf{D}_{\alpha(K^*)}$ 、 $J \times K^*$ 行列 $\mathbf{V}_{(K^*)}$ の 3 つの行列が、 $\mathbf{A}_{[K^*]}$ の SVD を構成することになる:

$$\mathbf{A}_{[K^*]} = \mathbf{U}_{(K^*)} \mathbf{D}_{\alpha(K^*)} \mathbf{V}_{(K^*)}^\top \quad (\text{A.1.5})$$

あきらかに、残差行列は以下となる。

$$\mathbf{A} - \mathbf{A}_{[K^*]} = \mathbf{U}_{(K-K^*)} \mathbf{D}_{\alpha(K-K^*)} \mathbf{V}_{(K-K^*)}^\top \quad (\text{A.1.6})$$

行列 \mathbf{Y} の要素の二乗和は $\text{trace}(\mathbf{Y}\mathbf{Y}^\top)$ に等しいから、(A.1.1), (A.1.5), (A.1.6) より

- \mathbf{A} の要素の二乗和は $\sum_k^K \alpha_k^2$
- $\mathbf{A}_{[K^*]}$ の要素の二乗和は $\sum_k^{K^*} \alpha_k^2$
- $\mathbf{A} - \mathbf{A}_{[K^*]}$ の要素の二乗和は $\sum_{k=K^*+1}^K \alpha_k^2$

となる。近似の質を示す伝統的指標としては、二乗和のパーセンテージがよくつかわれている:

$$\tau_{K^*} \equiv 100 \frac{\sum_k^{K^*} \alpha_k^2}{\sum_k^K \alpha_k^2} \quad (\text{A.1.7})$$

さっきの例では、 $\alpha_1^2 = 642.4, \alpha_2^2 = 4.6, \alpha_3^2 = 2.9$, 合計 650(これが \mathbf{A} の全要素の二乗和)。 $\tau_2 = 99.5$ である。

1.5 一般化 SVD

■一般化 SVD とは さて、これからの説明を楽にするために、SVD の定義にちょっとした一般化を加えよう。 $\mathbf{\Omega} (I \times I)$, $\mathbf{\Phi} (J \times J)$ を正定対称行列とすると、任意の実行列 $\mathbf{A} (I \times J)$, ランク K は以下のように表現できる:

$$\mathbf{A} = \mathbf{N} \mathbf{D}_\alpha \mathbf{M}^\top = \sum_k^K \alpha_k \mathbf{n}_k \mathbf{m}_k^\top \quad (\text{A.1.8})$$

ここで、 \mathbf{N} , \mathbf{M} の列は $\mathbf{\Omega}$, $\mathbf{\Phi}$ について正規直交化されている:

$$\mathbf{N}^\top \mathbf{\Omega} \mathbf{N} = \mathbf{M}^\top \mathbf{\Phi} \mathbf{M} = \mathbf{I} \quad (\text{A.1.9})$$

\mathbf{N} と \mathbf{M} の列は、それぞれ一般化左特異ベクトル、一般化右特異ベクトルと呼ぶことができるだろう。これらは依然として \mathbf{A} の列と行の正規直交基底だが、 I 次元空間・ J 次元空間における距離はもはや単純なユークリッド距離ではなく、 $\mathbf{\Omega}$, $\mathbf{\Phi}$ によって定義された一般化(重みつき)ユークリッド距離となる (Section 2.3 を見よ)。同様に、対角行列 \mathbf{D}_α の対角成分(大きい順に並んでいる)は一般化特異値と呼ぶことができるだろう。

■一般化 SVD の証明 一般化 SVD は、行列 $\mathbf{\Omega}^{1/2} \mathbf{A} \mathbf{\Phi}^{1/2}$ の SVD について考えれば容易に証明できる ($\mathbf{\Omega}$ が固有値分解 $\mathbf{\Omega} = \mathbf{W} \mathbf{D}_\mu \mathbf{W}^\top$ を持つならば $\mathbf{\Omega}^{1/2} = \mathbf{W} \mathbf{D}_\mu^{1/2} \mathbf{W}^\top$)。

$$\mathbf{\Omega}^{1/2} \mathbf{A} \mathbf{\Phi}^{1/2} = \mathbf{U} \mathbf{D}_\alpha \mathbf{V}^\top, \text{ただし } \mathbf{U}^\top \mathbf{U} = \mathbf{V}^\top \mathbf{V} = \mathbf{I} \quad (\text{A.1.10})$$

ここで

$$\mathbf{N} \equiv \mathbf{\Omega}^{-1/2} \mathbf{U} \text{ かつ } \mathbf{M} \equiv \mathbf{\Phi}^{-1/2} \mathbf{V} \quad (\text{A.1.11})$$

とすれば、(A.1.8) と (A.1.9) を得る。

ひらたくいえばこういうことであろう。任意の行列 \mathbf{A} の一般化 SVD を、次の手順で求めることができる。

1. $\mathbf{B} = \mathbf{\Omega}^{1/2} \mathbf{A} \mathbf{\Phi}^{1/2}$ を求める。
2. \mathbf{B} について、通常の SVD $\mathbf{B} = \mathbf{U} \mathbf{D}_\alpha \mathbf{V}^\top$ を求める。
3. $\mathbf{N} = \mathbf{\Omega}^{-1/2} \mathbf{U}$, $\mathbf{M} = \mathbf{\Phi}^{-1/2} \mathbf{V}$ とする。
4. $\mathbf{A} = \mathbf{N} \mathbf{D}_\alpha \mathbf{M}^\top$ が \mathbf{A} の一般化 SVD である。

さっきの例で試してみよう。話を簡単にするために、 $\mathbf{\Omega}$ を適当な対角行列にして

$$\mathbf{\Omega} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 16 \end{bmatrix}, \mathbf{\Phi} = \mathbf{I}$$

とすると、最初のステップは

$$\mathbf{B} = \mathbf{\Omega}^{1/2} \mathbf{A} \mathbf{\Phi}^{1/2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 12 & 8 & 10 \\ 24 & 27 & 21 \\ 40 & 44 & 48 \end{bmatrix}$$

つまり、 \mathbf{A} の行を $\mathbf{\Omega}$ で重みづけているわけである。同様に、 $\mathbf{\Phi}$ は列への重みとなる。3 番目のステップでは、特異ベクトルをこの重みで割り戻している。

■一般化 SVD と低ランク近似 低ランク近似の定理においては以下の一般化が導出される。(A.1.8) の最後の $K - K^*$ 項を落とした $\mathbf{A}_{[K^*]} \equiv \sum_k^{K^*} \alpha_k \mathbf{n}_k \mathbf{m}_k^\top = \mathbf{N}_{(K^*)} \mathbf{D}_{\alpha(K^*)} \mathbf{M}_{\alpha(K^*)}^\top$ は、 \mathbf{A} の「一般化されたランク K^* の最小二乗近似」となる。それはランク K^* (ないしそれ以下) のすべての行列 \mathbf{X} のなかで、

$$\text{trace}\{\mathbf{\Omega}(\mathbf{A} - \mathbf{X})\mathbf{\Phi}(\mathbf{A} - \mathbf{X})^\top\} \quad (\text{A.1.12})$$

を最小化する行列となる。

■一般化 SVD の幾何的な解釈 たとえば、 $\mathbf{\Omega}$ が正数 $\omega_1 \dots \omega_I$ からなる対角行列 \mathbf{D}_ω であるとしよう。(A.1.12) は以下のようになる:

$$\text{trace}\{\mathbf{D}_\omega(\mathbf{A} - \mathbf{X})\mathbf{\Phi}(\mathbf{A} - \mathbf{X})^\top\} = \sum_i^I \omega_i (\mathbf{a}_i - \mathbf{x}_i)^\top \mathbf{\Phi} (\mathbf{a}_i - \mathbf{x}_i) \quad (\text{A.1.13})$$

ただし、 $\mathbf{a}_i, \mathbf{x}_i$ は \mathbf{A}, \mathbf{X} の行を列ベクトルに書き換えたものである。この関数は本質的には、Pearson(1901), Young(1937) らが考えた一般化主成分分析の定義である (Section 2.5 で述べた)。 \mathbf{A} の I 本の行は、 J 次元の一般化 (重みつき) ユークリッド空間における I 個の点であると考えられる。この空間における距離は $\mathbf{\Phi}$ によって定義される (Section 2.5 では $\mathbf{\Phi}$ は対角行列でもあった。これを「対角距離」と呼ぶことが多い)。 $\omega_1 \dots \omega_I$ は、それぞれの行の点自体に割り当てられた重みである。 \mathbf{X} の行は K^* 次元下位空間における未知の点であり、 $\mathbf{X} \equiv \mathbf{A}_{[K^*]}$ によって得られる (A.1.13) を最小化するということは、 \mathbf{A} の I 個の点のあつまりに対して、平方距離の重みつき和という観点からみて最も近い下位空間を決めるとい

うことである。この例の場合、ベクトル $\mathbf{m}_1 \dots \mathbf{m}_{K^*}$ はその下位空間の正規直交な「主軸」を定義し、行列 $\mathbf{N}_{(K^*)} \mathbf{D}_{\alpha(K^*)} = [\alpha_1 \mathbf{n}_1 \dots \alpha_{K^*} \mathbf{n}_{K^*}]$ の行は、点のあつまりを下位空間に射影した際の (主軸についての) 座標をあらわしている。軸と射影の正規直交性は距離 Φ の観点から定義されていることに注意してほしい。近似した行列 $\mathbf{A}_{[K^*]}$ は、 $\mathbf{M}_{(K^*)}$ によって定義された最適空間と等しく、 α_{K^*} が厳密に α_{K^*+1} より大である限りにおいて一意に定まる (前述した SVD の一意性についての説明を参照のこと)。実用的には、近似の次元数 K^* は α_{K^*} と α_{K^*+1} の差が明確であるところに決めるべきである。

■一般化 SVD を用いたデータ視覚化 $\mathbf{F} \equiv \mathbf{N}_{(K^*)} \mathbf{D}_{\alpha(K^*)}$ の行を、 K^* 次元のユークリッド空間 (K^* はたいてい 2 ないし 3) においてプロットすることで、データ行列 \mathbf{A} の行の多次元的布置を調べることも多い。しかし、Gabriel(1971, 1981) が述べているように、同じ空間に $\mathbf{G} \equiv \mathbf{M}_{(K^*)}$ の行をプロットすることもできる。このような、ある行列の行と列の両方を点として同時にあらわす方法はバイプロットと呼ばれている。それは本質的には Tucker(1960) の述べた選好のベクトルモデルと同じなのだが、より広い範囲のデータに適用できる。バイプロット表示では、 i 番目の行を点として表すベクトル (すなわち \mathbf{F} の i 番目の行 \mathbf{f}_i^\top) と、 j 番目の列を点として表すベクトル (すなわち \mathbf{G} の j 番目の行 \mathbf{g}_j^\top) のスカラー積がデータ a_{ij} を近似する:

$$a_{ij} \approx \mathbf{f}_i^\top \mathbf{g}_j = (\mathbf{f}_i \text{の長さ}) \times (\mathbf{g}_j \text{の長さ}) \times (\mathbf{f}_i \text{と } \mathbf{g}_j \text{がなす角の } \cosine) \quad (\text{A.1.14})$$

通常、データ値 a_{ij} はなんらかのかたちで中心化される (たとえば行平均によって)。それにより、正の偏差 $a_{ij} > 0$ は鋭角をなすベクトル $\mathbf{f}_i, \mathbf{g}_j$ によって表され、負の偏差 $a_{ij} < 0$ は鈍角をなすベクトルによって表される。

2 一般的分析とその特殊ケース

■一般的分析 上記の説明から、矩形行列 \mathbf{Y} の一般的分析が以下になることがわかる。

フェーズ 1 \mathbf{Y} をなんらかの方法で前処理する。なんらかの中心化ないしデータの再コード化がなされることが多い。その結果を行列 \mathbf{A} とする。

フェーズ 2 所与の $\mathbf{\Omega}, \mathbf{\Phi}$ について、 \mathbf{A} の一般化 SVD を求める ((A.1.10), (A.1.11) を参照):

$$\mathbf{A} = \mathbf{N} \mathbf{D}_{\alpha} \mathbf{M}^{\top}, \text{ただし } \mathbf{N}^{\top} \mathbf{\Omega} \mathbf{N} = \mathbf{M}^{\top} \mathbf{\Phi} \mathbf{M} = \mathbf{I}$$

次にランク K^* の近似を選択する:

$$\mathbf{A}_{[K^*]} = \mathbf{N}_{(K^*)} \mathbf{D}_{\alpha(K^*)} \mathbf{M}_{(K^*)}^{\top}$$

フェーズ 3 データの行ないし列 (またはその両方) を視覚表示するために、所与の a と b を用い、行列 $\mathbf{F} \equiv \mathbf{N}_{(K^*)} \mathbf{D}_{\alpha(K^*)}^a$ の行, ないし行列 $\mathbf{G} \equiv \mathbf{M}_{(K^*)} \mathbf{D}_{\alpha(K^*)}^b$ の行, またはその両方を, K^* 本の直交軸についてプロットする。

■パラメータ この分析はコンピュータ・サブルーチン/プロシジャ/マクロとしてプログラミングできるし、そのプログラムに以下の「パラメータ」を与えてやればさまざまな分析が可能になる:

1. フェーズ 1 における, データ行列の中心化/再コード化のタイプ。
2. フェーズ 2 における, 正定対称行列 $\mathbf{\Omega}, \mathbf{\Phi}$ 。
3. スカラー a, b 。プロットする前に左特異ベクトルと右特異ベクトルをリスケールする際, そのために特異値をどのように配分するかを示す。

■さまざまな特殊ケース 上の分析の特殊ケースである, いくつかのよく知られた分析手法を表 A.1 に示した。それぞれの分析から得られるプロットの幾何的解釈についても簡単に説明した。より詳しい説明を求める読者は, 「参考文献」欄の教科書・論文を参照されたい。また, 具体的なデータセットを使って, この 3 つのフェーズの「パラメータ」について実験してみてほしい。

■表 A.1 について 表 A.1 を見る際の手助けとして, いくつか説明を加えておく。

1. 行の点と列の点のうち, どちらか一方にしか関心がないという場合も珍しくない。通常, リスケールされていない点 (例, $b = 0$) は関心がない点である。
2. 行の点と列の点の両方がプロットされ, かつ $a + b = 1$ である場合には, バイプロット解釈が可能である。すなわち, 表示におけるスカラー積 $\mathbf{f}_i^{\top} \mathbf{g}_j$ が行列 \mathbf{A} の要素 a_{ij} の近似となる。

3. 行の点と列の点の両方がプロットされ、かつ、たとえば $b = 0$ である場合には、列の点は行の点 ($a = 1$ によってリスケールされている) に比べて極端に「小さく」なることがあるだろう。こうした場合は、列の点を適当な量だけ一律にリスケールすると、列の点と行の点の尺度が異なることになるが、列の点の相対的位置はみやすくなる。 $a = b$ の場合、点はふつう同じ尺度でプロットされる。
4. 表が示しているのは、それぞれの手法における基本的な算出の枠組みである。手法によってはさらなる計算を必要とするものがある。たとえば主成分分析においては、成分得点 (行列 \mathbf{F} の列) と変数 (\mathbf{Y} ないし \mathbf{A} の列) との間の相関の計算が必要になる。SAS や GENSTAT のようなプログラミング言語を用いれば、これらを付け加えるのはとても簡単である (Appendix B をみよ)。
5. 手法 (4) と (5) では、データを最初に $(I - 1)^{1/2}$ で割っている。これは、分散ないし相関の真の尺度を図に復元するためである。たとえば、共分散パイプロットでは

$$\frac{\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^T\mathbf{Y}}{(I - 1)^{1/2}} = \mathbf{N}\mathbf{D}_\alpha\mathbf{M}^T$$

こうすると、 $\mathbf{M}\mathbf{D}_\alpha^2\mathbf{M}^T$ は共分散行列

$$\frac{\{\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^T\mathbf{Y}\}^T\{\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^T\mathbf{Y}\}}{(I - 1)}$$

に等しくなる。したがって、 $\mathbf{G} \equiv \mathbf{M}_{(K^*)}\mathbf{D}_{\alpha(K^*)}$ の列のスカラー積が、それに対応する共分散を近似することになる。つまり、図を標準偏差と相関の観点から解釈できるようになる (Gabriel, 1971)。

6. 手法 (2) と (3) におけるオプション (i) と (ii) のちがいについて注意してほしい。最終的な図示における行の点は同じだが、列の点が異なる。ちがう行列をパイプロット表示しているからである。
7. 距離を定義している行列の逆行列が、双対問題における「重み行列」になることに注意。正準相関分析と対応分析では、行列「パラメータ」である $\mathbf{\Omega}, \mathbf{\Phi}$ は重み行列として定義されており、その逆行列は双対空間における通常の距離となる。たとえば手法 (9) によって得られる、対応分析での行の点 \mathbf{F} の配置は、 $\mathbf{A} = \mathbf{D}_r^{-1}\mathbf{P} - \mathbf{1}\mathbf{1}^T\mathbf{D}_c, \mathbf{\Omega} = \mathbf{D}_r, \mathbf{\Phi} = \mathbf{D}_c^{-1}, a = 1$ という分析での行の点の配置と同じである。これは、 \mathbf{P} の中心化された行プロファイルをもとに \mathbf{A} の行とし、 \mathbf{D}_r の対角成分で重みづけした一般化主成分分析 (手法 (2)) である (Section 2.5 を参照)。同様に、列の点 \mathbf{G} の配置は、一般化主成分分析 $\mathbf{A} = \mathbf{D}_c^{-1}\mathbf{P}^T - \mathbf{1}\mathbf{1}^T\mathbf{D}_r, \mathbf{\Omega} = \mathbf{D}_c, \mathbf{\Phi} = \mathbf{D}_r^{-1}, b = 1$ で得られる配置と同じである。4章で示すように、集群分析法や双対尺度法といった対応分析のバリエーションは、パラメータ a, b の定義のみが異なる (両方が 0 とされることが多い)。対応分析において $a = 0, b = 0$ とすると、それぞれの行の点は、列の点の重心 (重みつき平均) に位置する。重みは分割表のその行における個々の値と比例する。

2.1 主座標分析

表 A.1 に挙げた手法のほとんどを扱うことができるもうひとつの枠組みとして、主座標分析 (Gower, 1966) がある。これはもう少し広い枠組みであり、元の矩形データ行列ではなく、点の間の距離の行列を入力とする。ここで、この分析の通常的な定義を一般化し、点に与えられる重みを扱うことができるようにしよう。まず、スカラー積の行列 \mathbf{S} を計算し (例 2.6.1(b) を見よ)、その一般化固有値分解を求めよう: $\mathbf{S} = \mathbf{N}\mathbf{D}_\mu\mathbf{N}^T$ 。ここでたとえば $\mathbf{N}\mathbf{D}_\omega\mathbf{N}^T = \mathbf{I}$ であれば、 \mathbf{D}_ω は重みの対角行列である。スカラー積の (最小二乗的に) 最適な K^T 次元表示は、 $\mathbf{F} \equiv \mathbf{N}_{(K^*)}\mathbf{D}_{\mu(K^*)}^{1/2}$ の行として与えられる。一般化固有値分解は、一般化 SVD の場合と同じく、通

常の固有値分解によって求めるのがふつうである (Section 2.5 と Appendix B を見よ)。つまり、 $\mathbf{D}_\omega^{1/2} \mathbf{S} \mathbf{D}_\omega^{1/2}$ の固有値と固有ベクトルをみつければ、それが $\mathbf{N} = \mathbf{D}_\omega^{1/2} \mathbf{U}$ である。

例 3.5.2 に、対応分析が双対主座標分析として定義できることを示した。

2.2 個人データのウェイトニング

上記の枠組みとは全然離れている問題として、 a_{ij} を指定された w_{ij} で重みづけし、行列 \mathbf{A} を重みづけ最小二乗近似する、という問題がある。これは Gabriel & Zamir(1979) によって議論されており、以下の点で有用である。

1. 欠損データを扱う際。 $w_{ij} = 0$ とする。
2. それぞれの値について信頼性の指標が手に入る場合。 w_{ij} を信頼性に比例させる。
3. 外れ値を扱う際。個別に重みを下げる。

我々の枠組みは、 $w_{ik} = s_i t_j$ 形式の重みづけスキーマを許容する。したがって、 \mathbf{A} の個別の行、個別の列、その両方を、最小二乗近似において重みづけすることができる。これは、たとえば外れ値のベクトルの役割を減らす際に有用である。しかし、一般的な重みづけスキーマにあわせようとする、もはや SVD に基づく理論に頼ることはできなくなり、その整然としたアルゴリズムや優れた最適的な特性は失われる。列ベクトル・行ベクトルが通常持っている幾何的性質も失われる。

表 A.1 さまざまな多次元手法の一般化 SVD の観点からの定義

表を簡条書きに改めた。

いくつかの手法について、さっきから使っている行列

$$\begin{bmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{bmatrix}$$

のバイプロット表示をつくってみる。行を R1 ... R4, 列を C1 ... C3 と呼ぶことにする。

なお, 手法 (1)(2)(3)(4)(5) のフェーズ 1 に登場する $\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^T\mathbf{Y}$ は,

$$\begin{aligned} & \mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^T\mathbf{Y} \\ = & \begin{bmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{bmatrix} - (1/4) \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{bmatrix} \\ = & \begin{bmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{bmatrix} - \begin{bmatrix} 6.25 & 6.5 & 6.75 \\ 6.25 & 6.5 & 6.75 \\ 6.25 & 6.5 & 6.75 \\ 6.25 & 6.5 & 6.75 \end{bmatrix} \\ = & \begin{bmatrix} -5.25 & -4.5 & -3.75 \\ -0.25 & -2.5 & -1.75 \\ +1.75 & +2.5 & +0.25 \\ +3.75 & +4.5 & +5.25 \end{bmatrix} \end{aligned}$$

つまり, データ行列 \mathbf{Y} を列ごとに中心化した行列のこと。参考のために、これを単純に特異値分解すると

$$= \begin{bmatrix} -5.25 & -4.5 & -3.75 \\ -0.25 & -2.5 & -1.75 \\ +1.75 & +2.5 & +0.25 \\ +3.75 & +4.5 & +5.25 \end{bmatrix} \begin{bmatrix} 11.63 & 0 & 0 \\ 0 & 2.12 & 0 \\ 0 & 0 & 1.64 \end{bmatrix} \begin{bmatrix} +0.56 & +0.62 & +0.56 \\ -0.71 & -0.00 & +0.71 \\ +0.44 & -0.79 & +0.44 \end{bmatrix}$$

手法 (1) 「主成分分析」(変数がプロットされている場合には「主成分バイプロット」)

特有用定義 データ行列 \mathbf{Y} は、ふつうは行がケース、列が変数。

フェーズ 1 $\mathbf{A} = \mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^\top \mathbf{Y}$

フェーズ 2 $\mathbf{\Omega} = (1/I)\mathbf{I}, \mathbf{\Phi} = \mathbf{I}$

フェーズ 3 $a = 1, b = 0$

幾何的解釈の概要 図の原点は行の重心。列の点は、 J 次元空間における列の点をもっとも「近い」(もっともフィットした) K^* 次元下位空間に直交射影したもの。もし必要なら、変数をベクトルとしてプロットする。バイプロット解釈が可能。 \mathbf{G} の列は共分散行列の固有ベクトルになる。

参考文献 Pearson(1901) と Hotelling(1933) (Bryant&Atchley(1975)はこの2つの研究を主成分分析のほかの論文とともに再現している); Morrison(1976); Gabriel(1971); Kendall(1975); Chambers(1977)

プロット表示をつくってみよう。

$$\mathbf{A} = \begin{bmatrix} -5.25 & -4.5 & -3.75 \\ -0.25 & -2.5 & -1.75 \\ +1.75 & +2.5 & +0.25 \\ +3.75 & +4.5 & +5.25 \end{bmatrix}$$

これを $\mathbf{\Omega} = (1/I)\mathbf{I}, \mathbf{\Phi} = \mathbf{I}$ を重みにして一般化 SVD すると

$$\mathbf{A} = \begin{bmatrix} -1.34 & -1 & -0.46 \\ -0.46 & +1 & +1.34 \\ +0.46 & +1 & -1.34 \\ +1.34 & -1 & +0.46 \end{bmatrix} \begin{bmatrix} 5.82 & 0 & 0 \\ 0 & 1.06 & 0 \\ 0 & 0 & 0.82 \end{bmatrix} \begin{bmatrix} +0.56 & +0.62 & +0.56 \\ -0.71 & -0.00 & +0.71 \\ +0.44 & -0.79 & +0.44 \end{bmatrix}$$

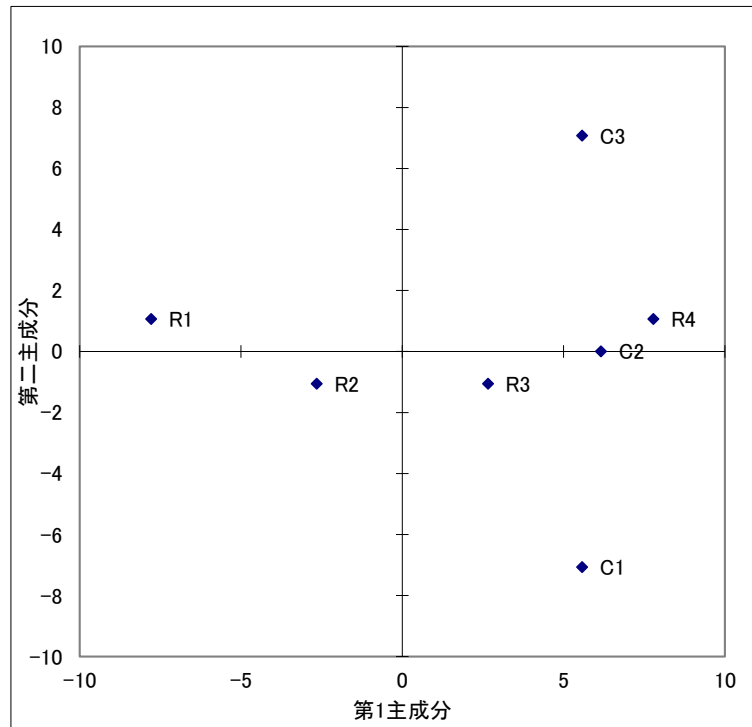
前のページの普通の SVD と比べると、 \mathbf{D}_α が $1/\sqrt{I}$ 倍、 \mathbf{N} が \sqrt{I} 倍になっている。

$K^* = 3$ として \mathbf{F}, \mathbf{G} を求めよう。

$$\mathbf{F} = \mathbf{N}_{(3)}\mathbf{D}_{\alpha(3)} = \begin{bmatrix} -7.79 & +1.06 & -0.37 \\ -2.65 & -1.06 & +1.10 \\ +2.65 & -1.06 & -1.10 \\ +7.79 & +1.06 & +0.37 \end{bmatrix}$$

$$\mathbf{G} = \mathbf{M}_{(3)} = \begin{bmatrix} +0.56 & -0.71 & +0.43 \\ +0.62 & -0.00 & -0.79 \\ +0.56 & +0.71 & +0.44 \end{bmatrix}$$

結局、得られる \mathbf{F} は $\mathbf{\Omega} = \mathbf{I}, \mathbf{\Phi} = \mathbf{I}$ とした場合と変わらない。もちろん \mathbf{G} も変わらない。ということは、わざわざ $\mathbf{\Omega} = (1/I)\mathbf{I}$ についての一般化 SVD を求めなくても、 $\mathbf{\Omega} = \mathbf{I}, \mathbf{\Phi} = \mathbf{I}$ とした一般化 SVD (つまり普通の SVD) を求め、 $\mathbf{F} = \mathbf{N}\mathbf{D}_\alpha, \mathbf{G} = \mathbf{M}$ としたほうがかんたんである。ここでは、 \mathbf{D}_α を世間でいうところの主成分分析の結果と揃えるために、わざわざこう書いているのだろう。



主成分プロット。列スコアは10倍した。

列は点ではなく原点からの矢印で表現することが多いのだが、めんどくさいので点で描いている。

Rでのコード例を示す。上述の通り、一般化SVDではなく普通のSVDで済ませている。

```
> Y <- matrix(c(1,2,3,6,4,5,8,9,7,10,11,12), ncol=3, byrow=TRUE)
> A <- scale(Y, center = TRUE, scale = FALSE)
> oSVD <- svd(A)
> oSVD$u %*% diag(oSVD$d)
      [,1]      [,2]      [,3]
[1,] -7.785228  1.06066 -0.3744716
[2,] -2.653918 -1.06066  1.0985066
[3,]  2.653918 -1.06066 -1.0985066
[4,]  7.785228  1.06066  0.3744716
> oSVD$v
      [,1]      [,2]      [,3]
[1,] 0.5570694 -7.071068e-01  0.4355155
[2,] 0.6159119 -5.551115e-17 -0.7878150
[3,] 0.5570694  7.071068e-01  0.4355155
```

Rでは、たとえば stats パッケージの prcomp 関数で主成分分析を行うことができる。返し値のうち、

- 要素 sdev は \mathbf{D}_α の対角要素の $\sqrt{I/(I-1)}$ 倍である。この値は \mathbf{F} の各列の不偏 SD に一致する。SAS や SPSS でいう「固有値」ではない。
- 要素 rotation は \mathbf{G} である。
- 要素 x は \mathbf{F} である。

コード例を示そう。

```
> Y <- matrix(c(1,2,3,6,4,5,8,9,7,10,11,12), ncol=3, byrow=TRUE)
> result <- prcomp(Y)
> result$sdev
[1] 6.7158049 1.2247449 0.9476096
> result$rotation
      PC1      PC2      PC3
[1,] 0.5570694 -7.071068e-01  0.4355155
[2,] 0.6159119 -5.551115e-17 -0.7878150
[3,] 0.5570694  7.071068e-01  0.4355155
> result$x
      PC1      PC2      PC3
[1,] -7.785228  1.06066 -0.3744716
[2,] -2.653918 -1.06066  1.0985066
[3,]  2.653918 -1.06066 -1.0985066
[4,]  7.785228  1.06066  0.3744716
```

biplot(result, scale=0) で図が出力される。scale オプションをつけないと手法 (5) になるので注意。

SAS では、proc princomp に cov オプションをつけて実行すると再現できる。

- 「固有値」は、フェーズ 2 で $\mathbf{\Omega} = (1/(I-1))\mathbf{I}$ としたときに得られる \mathbf{D}_α の対角要素の二乗、すなわち上式の \mathbf{D}_α の対角要素の二乗の $I/(I-1)$ 倍である。データ行列ではなくその不偏共分散行列を分析しているためである。
- 「固有ベクトル」は \mathbf{G} である。
- 「主成分得点」は \mathbf{F} である。

SPSS では、factor コマンドに method=covariance, extraction=pc オプションをつけて実行すると再現できる。

- 「固有値」は、SAS の proc princomp の場合と同じく、上式の \mathbf{D}_α の対角要素の二乗の $I/(I-1)$ 倍である。
- 「成分行列」は \mathbf{G} の列に上記の固有値の平方根を掛けたものである。
- 「因子得点」は \mathbf{F} の列を上記の固有値で割ったものである。

手法 (2) 「一般化主成分分析」(変数がプロットされている場合には「一般化主成分バイプロット」)

特有な定義 \mathbf{D}_ω は正の重み (合計 1) からなる対角行列。

オプション (i) :

フェーズ 1 $\mathbf{A} = \mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^\top \mathbf{D}_\omega \mathbf{Y}$

フェーズ 2 $\mathbf{\Omega} = (1/I)\mathbf{D}_\omega$, Φ は所与

フェーズ 3 $a = 1, b = 0$

オプション (ii) :

フェーズ 1 $\mathbf{A} = \{\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^\top \mathbf{D}_\omega \mathbf{Y}\} \Phi^{1/2}$

フェーズ 2 $\mathbf{\Omega} = (1/I)\mathbf{D}_\omega$, $\Phi = \mathbf{I}$

フェーズ 3 $a = 1, b = 0$

幾何的解釈の概要 (1) とのちがいは, 行の空間が距離 Φ に従う一般化ユークリッド空間であり, 行の点が \mathbf{D}_ω の対角要素 $\omega_1 \dots \omega_I$ によって重み付けられていること。 \mathbf{A} の近似の質は同じ。オプション (i) と (ii) は, \mathbf{F} の行の点については同じ布置を与えるが, \mathbf{G} の列の点については異なる布置を与える。手法 (3) を見よ。

参考文献 手法 (3), (7), (8), (9) などよく知られた手法の多くはこの手法の特殊ケースである。 \mathbf{D}_ω, Φ を変えたときに何が起こるかはこれからの課題である。

手法 (3) 「標準化データについての主成分分析」(変数がプロットされている場合には「標準化データについての主成分バイプロット」)

特有な定義 \mathbf{D}_s は変数の標準偏差からなる対角行列。

オプション (i) :

フェーズ 1 $\mathbf{A} = \mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^\top \mathbf{Y}$

フェーズ 2 $\mathbf{\Omega} = \mathbf{I}, \mathbf{\Phi} = \mathbf{D}_s^{-2}$

フェーズ 3 $a = 1, b = 0$

オプション (ii) :

フェーズ 1 $\mathbf{A} = \{\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^\top \mathbf{Y}\}\mathbf{D}_s^{-1}$

フェーズ 2 $\mathbf{\Omega} = \mathbf{I}, \mathbf{\Phi} = \mathbf{I}$

フェーズ 3 $a = 1, b = 0$

幾何的解釈の概要 手法 (2) の良く知られている特殊ケース。オリジナルの軸における行の点の分散が軸の間で等しくなっている。オプション (ii) では、 \mathbf{G} の列は相関行列の通常固有ベクトルである。オプション (i) では $\tilde{\mathbf{G}} = \mathbf{D}_s \mathbf{G}$ をつくることになり、標準偏差が列の点のベクトルによって近似されるが、共分散構造は近似されない。手法 (4) を見よ。

参考文献 多変量解析の教科書の多くは、この問題を「相関行列の主成分分析」として扱っている (すなわちオプション (ii))。

手法 (4) 「共分散バイプロット」

フェーズ 1 $\mathbf{A} = \frac{\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^\top \mathbf{Y}}{(I-1)^{1/2}}$

フェーズ 2 $\mathbf{\Omega} = \mathbf{I}, \mathbf{\Phi} = \mathbf{I}$

フェーズ 3 $a = 0, b = 1$

幾何的解釈の概要 図の原点は行の重心。列の点は、 J 次元空間における列の点をもっとも「近い」(もっともフィットした) K^* 次元下位空間に直交射影したもの。変数をベクトルとしてプロットする。その長さは変数の標準偏差を近似する。ベクトル間のコサインは変数間の相関を近似する。

参考文献 Gabriel(1971,1972,1981)

バイプロット表示を作ってみよう。

$$\mathbf{A} = \begin{bmatrix} -3.03 & -2.60 & -2.17 \\ -0.14 & -1.44 & -1.01 \\ +1.01 & +1.44 & +0.14 \\ +2.17 & +2.60 & +3.03 \end{bmatrix}$$

これを SVD して、

$$\mathbf{A} = \begin{bmatrix} -0.67 & +0.5 & -0.23 \\ -0.23 & -0.5 & +0.67 \\ +0.23 & -0.5 & -0.67 \\ +0.67 & +0.5 & +0.23 \end{bmatrix} \begin{bmatrix} 6.72 & 0 & 0 \\ 0 & 1.22 & 0 \\ 0 & 0 & 0.95 \end{bmatrix} \begin{bmatrix} +0.56 & +0.62 & +0.56 \\ -0.71 & +0.00 & +0.71 \\ +0.44 & -0.79 & +0.44 \end{bmatrix}$$

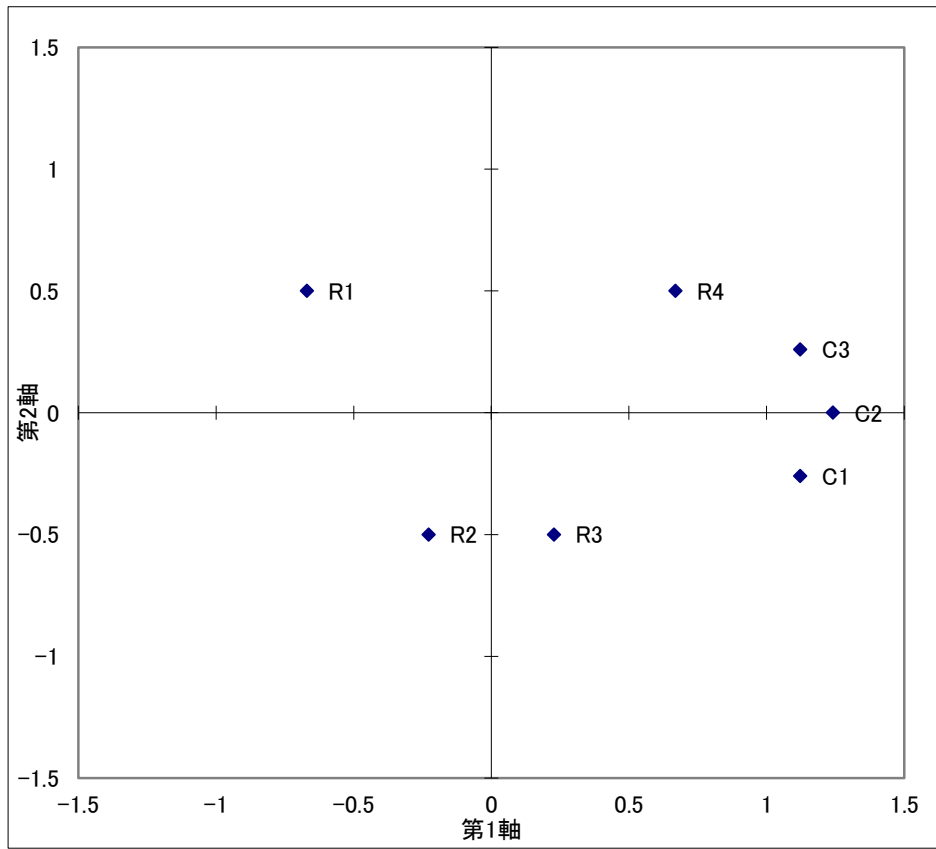
$\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^\top \mathbf{Y}$ を SVD した場合と比べて、 \mathbf{D}_α が $1/\sqrt{I-1}$ 倍になっている。

$K^* = 3$ として

$$\mathbf{F} = \mathbf{N}_{(3)} = \begin{bmatrix} -0.67 & +0.5 & -0.23 \\ -0.23 & -0.5 & +0.67 \\ +0.23 & -0.5 & -0.67 \\ +0.67 & +0.5 & +0.23 \end{bmatrix}$$

$$\mathbf{G} = \mathbf{M}_{(3)}\mathbf{D}_{\alpha(3)} = \begin{bmatrix} +0.56 & -0.71 & +0.44 \\ +0.62 & +0.00 & -0.79 \\ +0.56 & +0.71 & +0.44 \end{bmatrix} \begin{bmatrix} 6.72 & 0 & 0 \\ 0 & 1.22 & 0 \\ 0 & 0 & 0.95 \end{bmatrix} = \begin{bmatrix} +3.74 & -0.87 & +0.41 \\ +4.13 & +0.00 & -0.75 \\ +3.74 & +0.87 & +0.41 \end{bmatrix}$$

\mathbf{G} 行列の 1 行目の長さは 3.86 で、 $\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^\top \mathbf{Y}$ の 1 列目の不偏標準偏差に一致する。また \mathbf{G} 行列の 1 行目と 2 行目の内積は 15.17 で、 $\mathbf{Y} - (1/I)\mathbf{1}\mathbf{1}^\top \mathbf{Y}$ の 1 列目と 2 列目の不偏共分散に一致する。なるほど。



共分散プロット。行スコアは 0.3 倍した。

R の場合、その 1。主成分分析の関数を使わず、自力で特異値分解して **F**, **G** を求めてみる。

```
> Y <- matrix(c(1,2,3,6,4,5,8,9,7,10,11,12), ncol=3, byrow=TRUE)
> X <- scale(Y, scale=FALSE)
> A <- X / sqrt(3)
> oSVD <- svd(A)
> oSVD$u
      [,1] [,2]      [,3]
[1,] -0.6692874  0.5 -0.2281544
[2,] -0.2281544 -0.5  0.6692874
[3,]  0.2281544 -0.5 -0.6692874
[4,]  0.6692874  0.5  0.2281544
> oSVD$v %*% diag(oSVD$d)
      [,1]      [,2]      [,3]
[1,] 3.741169 -8.660254e-01  0.4126986
[2,] 4.136344  5.438960e-16 -0.7465411
[3,] 3.741169  8.660254e-01  0.4126986
```

R の場合、その 2。主成分分析の関数を使ってみる。

prcomp 関数がなにをしていたかを思い出そう。手法 (1) で調べたように、prcomp 関数は、入力行列を中心化し、 $SVD(\Omega = \Phi = I$ による一般化 SVD) によって M, D_α, N を求め、

- 要素 x として $\sqrt{I} N D_\alpha$ を
- 要素 rotation として M を
- 要素 sdev として $\sqrt{I/(I-1)}(1/\sqrt{I}) D_\alpha = \frac{1}{\sqrt{I-1}} D_\alpha$ を、

返してくるのであった。いっぽういま欲しいのは、同じように M, D_α, N を求めたときの

- N
- $\frac{1}{\sqrt{I-1}} M D_\alpha$

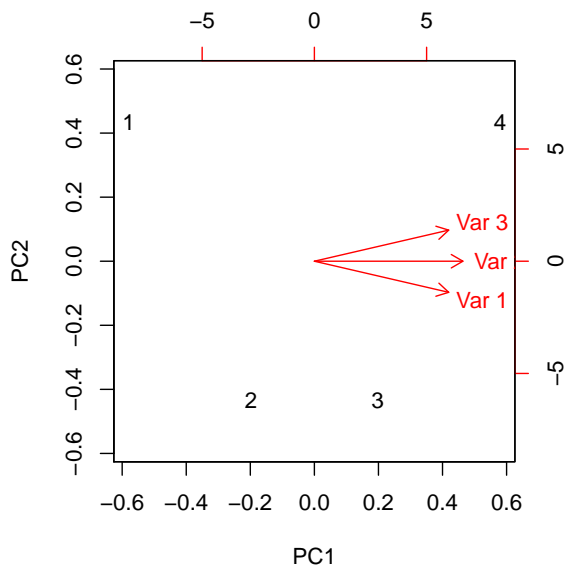
である。

従って、 F, G を求めるコードは以下となる。

```
> Y <- matrix(c(1,2,3,6,4,5,8,9,7,10,11,12), ncol=3, byrow=TRUE)
> result <- prcomp(Y)
> t(t(result$x)/ (result$sdev*sqrt(3)))
      PC1  PC2      PC3
[1,] -0.6692874  0.5 -0.2281544
[2,] -0.2281544 -0.5  0.6692874
[3,]  0.2281544 -0.5 -0.6692874
[4,]  0.6692874  0.5  0.2281544
> t(t(result$rotation) * result$sdev)
      PC1      PC2      PC3
[1,] 3.741169 -8.660254e-01  0.4126986
[2,] 4.136344 -6.798700e-17 -0.7465411
[3,] 3.741169  8.660254e-01  0.4126986
```

R の場合、その 3。もし biplot 関数で図を描いたら何が起きるか。

```
> Y <- matrix(c(1,2,3,6,4,5,8,9,7,10,11,12), ncol=3, byrow=TRUE)
> result <- prcomp(Y)
> biplot(result)
```



左と下の目盛が行の座標、上の右の目盛が列の座標を表している。列を表す矢印は、ラベルと重ならないようにするため、少し短めに描かれている。

行の座標は、`result$x` の (1,2) 列を、`result$sdev` の対応する値で割りデータの行数の平方根で割った値、すなわち

$$\begin{bmatrix} -0.58 & +0.43 \\ -0.20 & -0.43 \\ +0.20 & -0.43 \\ +0.58 & +0.43 \end{bmatrix}$$

となっている。本資料の説明とくらべると $\frac{\sqrt{I-1}}{\sqrt{I}}\mathbf{F}$ である。

列の座標は、`result$rotation` の (1,2) 列に、`result$sdev` の対応する値を掛け、さらにデータの行数の平方根を掛けた値、すなわち

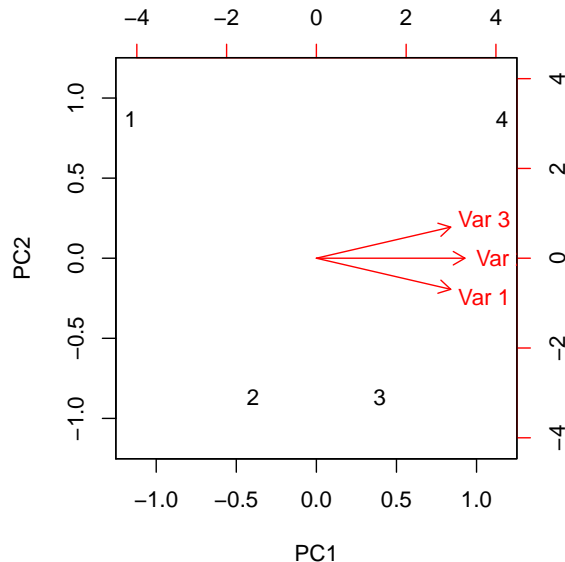
$$\begin{bmatrix} +7.48 & -1.73 \\ +8.27 & -0.00 \\ +7.48 & +1.73 \end{bmatrix}$$

となっている。本資料の説明とくらべると $\sqrt{I}\mathbf{G}$ である。

このように、`biplot(result)` の出力は本資料の説明と一致しない。しかし、行の座標と列の座標の相対的サイズがちがうだけで、同じ図である。

R の場合、その 4。biplot 関数に引数 pc.biplot=TRUE をつけるとどうなるか。

```
> Y <- matrix(c(1,2,3,6,4,5,8,9,7,10,11,12), ncol=3, byrow=TRUE)
> result <- prcomp(Y)
> biplot(result, pc.biplot=TRUE)
```



引数 pc.biplot=TRUE は、データの行数の平方根でかけたり割ったりするのをやめてくれ、という指示である。

従って行の座標は、result\$x の (1,2) 列を result\$sdev の対応する値で割った値、すなわち

$$\begin{bmatrix} -1.16 & +0.86 \\ -0.40 & -0.86 \\ +0.40 & -0.86 \\ +0.16 & +0.86 \end{bmatrix}$$

となる。本資料の説明とくらべると $\sqrt{I-1}\mathbf{F}$ である。

列の座標は result\$rotation の (1,2) 列に、result\$sdev の対応する値を掛けた値、すなわち

$$\begin{bmatrix} +3.74 & -0.86 \\ +4.14 & -0.00 \\ +3.74 & +0.86 \end{bmatrix}$$

となる。本資料の説明における \mathbf{G} と一致する。

このように、pc.biplot=TRUE をつけても、出力は本資料の説明と一致しない。しかし、行の座標と列の座標の相対的サイズがちがうだけで、同じ図である。

WRC Research Systems の Brandmap の場合,

- \mathbf{Y} 行列を転置して与えると上記の説明と一致する。これはこのソフトウェアが、列がブランド、行が属性である集計表を入力として想定しているから。
- Data Centering を”Row”とする。これは \mathbf{Y} の「行」ではなく、入力行列の「行」を指している。
- Factorization を” $G*(V*H')$ ”とする。これは入力行列の SVD ではなく、 \mathbf{Y} の SVD を指している。やれやれ。
- Brandmap はフェーズ 1 でデータを $\sqrt{I-1}$ で割らない。そのため、 \mathbf{Y} の列 (入力行列の行) をあらかずベクトルの長さは標本標準偏差の \sqrt{I} 倍、ベクトルの内積は標本共分散の I 倍になる。
- \mathbf{Y} の行 (入力行列の列) の点は、上記の座標を謎の値で定数倍した位置にプロットされるようである。おそらく、図の外観上の都合で倍率を決めているのだろう。

手法 (5) 「相関バイプロット」

特有な定義 D_s は変数の標準偏差からなる対角行列。

フェーズ 1 $A = \frac{\{Y - (1/I)\mathbf{1}\mathbf{1}^T Y\} D_s^{1/2}}{(I-1)^{1/2}}$

フェーズ 2 $\Omega = I, \Phi = I$

フェーズ 3 $a = 0, b = 1$

幾何的解釈の概要 共分散バイプロットとのちがいは、列の点が J 次元空間の原点から長さ 1 の位置にあること。したがって、それぞれの変数の図示の質は、点が K^* 次元の単位球面 (たとえば 2 次元における単位円) とどのくらい近いかを観察すればわかることになる。

参考文献 Hill(1969)

A は、標準化されたデータ行列を $\sqrt{I-1}$ で割ったものである。

$$A = \begin{bmatrix} -0.9 & -0.7 & -0.6 \\ +0.0 & -0.4 & -0.3 \\ +0.3 & +0.4 & +0.0 \\ +0.6 & +0.7 & +0.9 \end{bmatrix}$$

これを素直に SVD して

$$A = \begin{bmatrix} -0.7 & -0.5 & +0.2 \\ -0.2 & +0.5 & -0.7 \\ +0.2 & +0.5 & +0.7 \\ +0.7 & -0.5 & -0.2 \end{bmatrix} \begin{bmatrix} 1.9 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} \begin{bmatrix} +0.6 & +0.6 & +0.6 \\ +0.7 & 0 & -0.7 \\ +0.4 & -0.8 & +0.4 \end{bmatrix}$$

$K^* = 3$ として、

$$F = N_{(3)} = \begin{bmatrix} -0.9 & -0.7 & -0.6 \\ +0.0 & -0.4 & -0.3 \\ +0.3 & +0.4 & +0.0 \\ +0.6 & +0.7 & +0.9 \end{bmatrix}$$

$$G = M_{(3)} D_{\alpha(3)} = \begin{bmatrix} +0.6 & +0.7 & +0.4 \\ +0.6 & +0.0 & -0.8 \\ +0.6 & -0.7 & +0.4 \end{bmatrix} \begin{bmatrix} 1.9 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} = \begin{bmatrix} +1.1 & +0.3 & +0.1 \\ +1.1 & +0.0 & -0.2 \\ +1.1 & -0.3 & +0.1 \end{bmatrix}$$

G 行列の各行の長さは $\sqrt{4/3} = 1.2$, 行間の内積は相関の $4/3$ 倍に一致する。こうなるのは、フェーズ 1 でデータを $\sqrt{I-1}$ で割ってしまったからである。素直に \sqrt{I} で割っておけばよかったのに。

R でやってみよう。R の `scale()` で使われる SD は不偏 SD なので、わざわざ標本 SD に修正している。

```
> Y <- matrix(c(1,2,3,6,4,5,8,9,7,10,11,12), ncol=3, byrow=TRUE)
> agSD <- sqrt(apply(Y, 2, var) * (3/4))
> A <- scale(Y, center = TRUE, scale = agSD) / sqrt(3)
> oSVD <- svd(A)
> oSVD$u
      [,1] [,2]      [,3]
[1,] -0.6718767  0.5 -0.2204125
[2,] -0.2204125 -0.5  0.6718767
[3,]  0.2204125 -0.5 -0.6718767
[4,]  0.6718767  0.5  0.2204125
> oSVD$v %*% diag(oSVD$d)
      [,1]      [,2]      [,3]
[1,] 1.119862 -2.589191e-01  0.1104651
[2,] 1.133898 -2.032635e-17 -0.2181956
[3,] 1.119862  2.589191e-01  0.1104651
```

WRC Research Systems の Brandmap の場合、

- \mathbf{Y} 行列を転置して与えると上記の説明と一致する。
- Data Centering を”Row”, Data Standardization を”Row”とする。
- Factorization を” $G*(V*H')$ ”とする。
- Brandmap は、フェーズ 1 で不偏標準偏差を使った標準化を行い (したがって標準偏差は $\sqrt{(I-1)/I}$ となる), かつそれを割らない。その結果, \mathbf{Y} の列 (入力行列の行) をあらわすベクトルの長さは $\sqrt{I-1}$, ベクトルの内積は標本相関の $I-1$ 倍になる。
- \mathbf{Y} の行 (入力行列の列) の点は, 上記の座標を謎の値で定数倍した位置にプロットされる。

手法 (6) 「対称バイプロット」

特有な定義 \hat{y} はデータ行列の全要素の平均。

フェーズ 1 $\mathbf{A} = \mathbf{Y} - \hat{y}\mathbf{1}\mathbf{1}^\top$

フェーズ 2 $\mathbf{\Omega} = \mathbf{I}, \mathbf{\Phi} = \mathbf{I}$

フェーズ 3 $a = 1/2, b = 1/2$

幾何的解釈の概要 このバイプロットは、行と列の幾何ではなく、行列の構造そのものに焦点を当てている。行-列間のバイプロット解釈のみが妥当。

参考文献 Bradu & Gabriel(1978); Gabriel(1981)

手法 (7) 「正準相関分析」

特有な定義 変数が I 個の変数群と J 個の変数群に自然に分かれているとする。 $\mathbf{Y} \equiv \mathbf{Y}_1\mathbf{Y}_2$ とする。 \mathbf{Y} は列平均 (変数ごとの平均) によって中心化されているものとする。群内の共分散行列を $\mathbf{S}_{11}, \mathbf{S}_{22}$, 群間の共分散行列を \mathbf{S}_{12} とする。

フェーズ 1 $\mathbf{A} = \mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}$

フェーズ 2 $\mathbf{\Omega} = \mathbf{S}_{11}, \mathbf{\Phi} = \mathbf{S}_{22}$

フェーズ 3 $a = 1, b = 1$

幾何的解釈の概要 \mathbf{F} と \mathbf{G} の列は正準負荷 (すなわち, 正準変数を作る際のオリジナル変数に対する係数)。正準得点 (すなわち $\mathbf{Y}_1\mathbf{F}$ と $\mathbf{Y}_2\mathbf{G}$ の行) をプロットした場合は, ケースをあらわす 2 つの点の集まりは, それぞれのマハラノビス空間 (距離はそれぞれ $\mathbf{S}_{11}^{-1}, \mathbf{S}_{22}^{-1}$) においてあらわされている点を, 2 つの点の集まりの間の相関をもっとも高くする K^* 次元下位空間に射影したものである。

参考文献 Anderson(1958); Tatsuoka(1971); Morrison(1976); Mardia et al.(1979); Falkenhagen & Nash(1978); Chambers(1977); Gittins(1979)

手法 (8) 「正準変量分析」

特有益義 \mathbf{Y} の行は H 個の群に分かれる。 J 個の変数の群内平均からなる $H \times J$ 行列を $\bar{\mathbf{Y}}$ とする。各群の行数の割合からなる対角行列を \mathbf{D}_w とする ($H \times H$)。通常のプールされた群内共分散行列を \mathbf{S} とする。

フェーズ 1 $\mathbf{A} = \bar{\mathbf{Y}} - \mathbf{1}\mathbf{1}^\top \mathbf{D}_w \bar{\mathbf{Y}}$

フェーズ 2 $\mathbf{\Omega} = \mathbf{D}_w, \mathbf{\Phi} = \mathbf{S}^{-1}$

フェーズ 3 $a = 1, b = 0$

幾何的解釈の概要 図の原点は \mathbf{Y} の行の重心 (すなわち $\bar{\mathbf{Y}}$ の重心)。手法 (2) で定義した一般化主成分分析の特殊ケースである。群の重心の主軸を定義し、群のサイズで重み付け、プールした群内共分散行列 \mathbf{S} で定義されたマハラノビス空間であらわしたもの。

参考文献 正準相関分析の参考文献と同じ。さらに、判別分析についての多くの教科書。

手法 (9) 「対応分析」(「双対尺度法」「集群分析法」「分割表の正準分析」)

特有な定義 たとえば, \mathbf{Y} は分割表。 \mathbf{Y} をその合計で割ったものを \mathbf{P} とする。 \mathbf{P} の行和からなる対角行列を \mathbf{D}_r , 列和からなる対角行列を \mathbf{D}_c とする。

フェーズ 1 $\mathbf{A} = \mathbf{D}_r^{-1} \mathbf{P} \mathbf{D}_c^{-1} - \mathbf{1} \mathbf{1}^\top$

フェーズ 2 $\mathbf{\Omega} = \mathbf{D}_r, \mathbf{\Phi} = \mathbf{D}_c$

フェーズ 3 $a = 1, b = 1$

幾何的解釈の概要 手法 (7) の特殊ケースで, データが離散的な場合。分割表 \mathbf{Y} の行と列が点としてあらわされ, 点の位置は行と列の間の連関をあらわす。行と列の図示は双対一般化主成分分析である(ないし主座標分析。Example 3.5.2 をみよ)

参考文献 Benzécli et al.(1973); Hill(1974); Greenacre(1978); Nishisato(1980); Greenacre(1981); Gifi(1981); Gauch(1982); 本書

バイプロット表示を作ってみよう。

$$\mathbf{A} = \begin{bmatrix} 1/.08 & 0 & 0 & 0 \\ 0 & 1/.19 & 0 & 0 \\ 0 & 0 & 1/.31 & 0 \\ 0 & 0 & 0 & 1/.42 \end{bmatrix} \begin{bmatrix} .01 & .03 & .04 \\ .08 & .05 & .06 \\ .10 & .12 & .09 \\ .13 & .14 & .15 \end{bmatrix} \begin{bmatrix} 1/.32 & 0 & 0 \\ 0 & 1/.33 & 0 \\ 0 & 0 & 1/.35 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -.5 & +.0 & +.4 \\ +.2 & -.2 & +.0 \\ +.0 & +.1 & -.2 \\ +.1 & +.0 & +.1 \end{bmatrix}$$

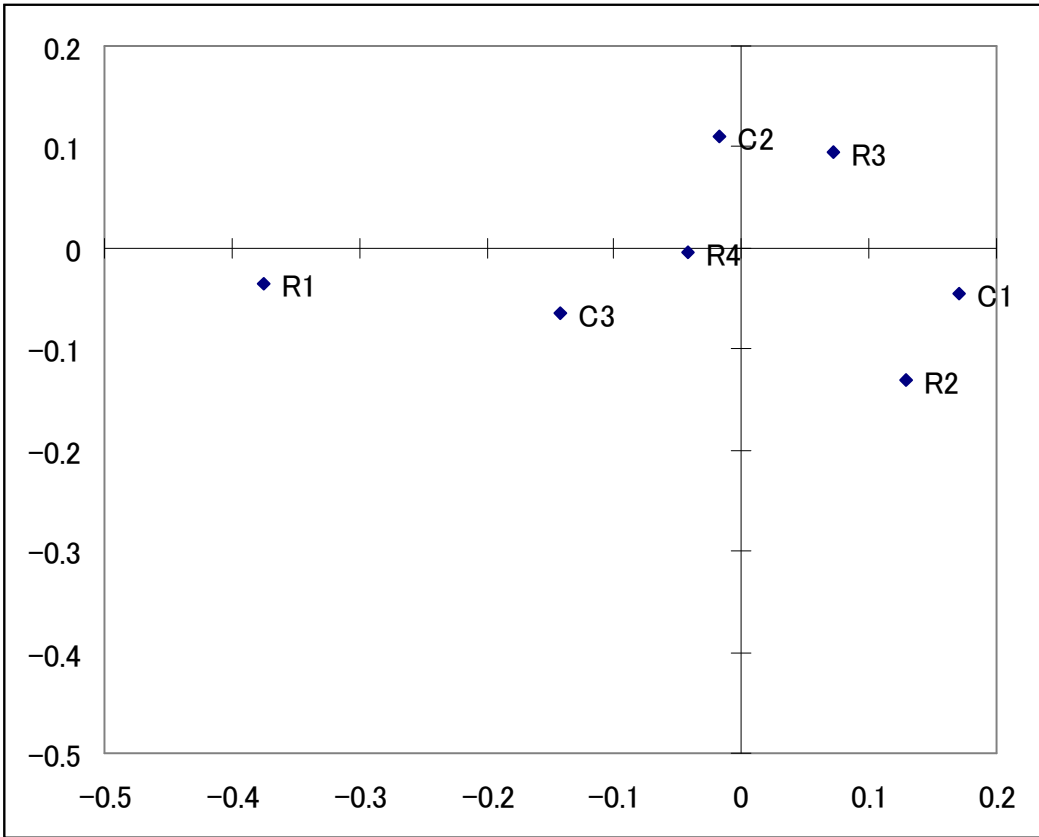
$\mathbf{D}_r, \mathbf{D}_c$ を重みにした一般化 SVD を行い

$$\mathbf{A} = \begin{bmatrix} -2.9 & +0.5 \\ +1.0 & +1.7 \\ +0.6 & -1.2 \\ -0.3 & +0.1 \end{bmatrix} \begin{bmatrix} 0.13 & 0 \\ 0 & 0.08 \end{bmatrix} \begin{bmatrix} +1.3 & -0.1 & -1.1 \\ +0.6 & -1.4 & +0.8 \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} -2.9 & +0.5 \\ +1.0 & +1.7 \\ +0.6 & -1.2 \\ -0.3 & +0.1 \end{bmatrix} \begin{bmatrix} 0.13 & 0 \\ 0 & 0.08 \end{bmatrix} = \begin{bmatrix} -0.38 & +0.04 \\ +0.13 & +0.13 \\ +0.07 & -0.10 \\ -0.04 & +0.00 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} +1.3 & +0.6 \\ -0.1 & -1.4 \\ -1.1 & +0.8 \end{bmatrix} \begin{bmatrix} 0.13 & 0 \\ 0 & 0.08 \end{bmatrix} = \begin{bmatrix} +0.17 & +0.05 \\ -0.02 & -0.11 \\ -0.14 & +0.06 \end{bmatrix}$$

データ行列 \mathbf{Y} の 1 列目は, 上から 1, 6, 8, 10, 計 25。いま, この 25 人がそれぞれ第 1 軸の行スコア (\mathbf{F} 行列の 1 列目) を持っていると考えて, その平均を求めると 0.02。これを第 1 特異値 0.13 で割ると, 第 1 軸の列スコア 0.17 (\mathbf{G} 行列の 1 列目) が得られる。このように, 列スコアは行スコアの加重平均となっている。同様に, 行スコアは列スコアの加重平均となっている。この性質を双対性という。



対応分析

R でやってみよう。

```
> Y <- matrix(c(1,2,3,6,4,5,8,9,7,10,11,12), ncol=3, byrow=TRUE)
> P <- Y / sum(Y)
> DR <- diag(rowSums(P))
> DC <- diag(colSums(P))
> A <- solve(DR) %*% P %*% solve(DC) - 1
> A
      [,1] [,2] [,3]
[1,] -0.4800000 0.000 0.444444444
[2,]  0.2480000 -0.200 -0.03703704
[3,]  0.0400000 0.125 -0.15740741
[4,] -0.05454545 0.000 0.05050505
> # 一般化 SVD
> oSVD <- svd(sqrt(DR) %*% A %*% sqrt(DC))
> N <- solve(sqrt(DR)) %*% oSVD$u
> M <- solve(sqrt(DC)) %*% oSVD$v
> # A の再現
> N %*% diag(oSVD$d) %*% t(M)
      [,1] [,2] [,3]
[1,] -0.4800000 1.200739e-16 0.444444444
[2,]  0.2480000 -2.000000e-01 -0.03703704
[3,]  0.0400000 1.250000e-01 -0.15740741
[4,] -0.05454545 1.353203e-17 0.05050505
> # F 行列
> N %*% diag(oSVD$d)
      [,1] [,2] [,3]
[1,] -0.37540739 -0.035937598 -3.680718e-17
[2,]  0.12922721 -0.129697013 -4.787124e-17
[3,]  0.07174225  0.095660282 -4.787124e-17
[4,] -0.04265993 -0.004083818 -6.557373e-17
> # G 行列
> M %*% diag(oSVD$d)
      [,1] [,2] [,3]
[1,]  0.17091524 -0.04571431 -5.540068e-17
[2,] -0.01726158  0.11046283 -5.540068e-17
[3,] -0.14163259 -0.06404355 -5.540068e-17
```

R ではたとえば `ca` パッケージの `ca` 関数で再現できる。`ca()` の返し値を `print()` したときに表示される座標、すなわち `ca()` の返し値の要素 `rowcoord`, `colcoord` は **F**, **G** ではなく、**A** の一般化 SVD の **N**, **M** である。

```
> library(ca)
> Y <- matrix(c(1,2,3,6,4,5,8,9,7,10,11,12), ncol=3, byrow=TRUE)
> oCA <- ca(Y)
> print(oCA)
Principal inertias (eigenvalues):
      1      2
Value  0.016406 0.006157
Percentage 72.71%  27.29%

Rows:
      [,1]      [,2]      [,3]      [,4]
Mass    0.076923  0.192308  0.307692  0.423077
ChiDist 0.377124  0.183087  0.119574  0.042855
Inertia 0.010940  0.006446  0.004399  0.000777
Dim. 1  -2.930910  1.008913  0.560112 -0.333058
Dim. 2  -0.458002 -1.652905  1.219129 -0.052046

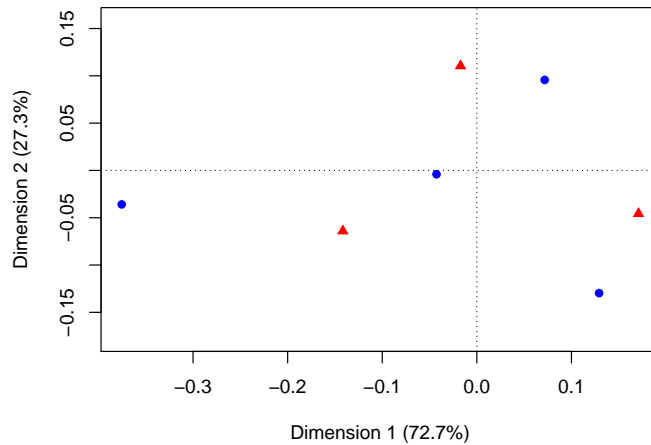
Columns:
      [,1]      [,2]      [,3]
Mass    0.320513  0.333333  0.346154
ChiDist 0.176923  0.111803  0.155439
Inertia 0.010033  0.004167  0.008364
Dim. 1   1.334383 -0.134766 -1.105765
Dim. 2  -0.582599  1.407778 -0.816194

> oCA$rowcoord %*% diag(oCA$sv)
      [,1]      [,2]
[1,] -0.37540739 -0.035937598
[2,]  0.12922721 -0.129697013
[3,]  0.07174225  0.095660282
[4,] -0.04265993 -0.004083818

> oCA$colcoord %*% diag(oCA$sv)
      [,1]      [,2]
[1,]  0.17091524 -0.04571431
[2,] -0.01726158  0.11046283
[3,] -0.14163259 -0.06404355
```

前頁の続き。前述のとおり、ca() の返し値に入っているのは **N, M** なのだが、返し値を plot した際の座標は **F, G** となる。

```
> plot(oCA)
```



WRC Research Systems の Brandmap では Factorization を”Symmetric Prn”とすると再現できる。

SAS では proc corresp で再現できる。

SPSS では correspondence コマンドで、standardize=rcmean, measure=chisq, normalization=principal オプションをつけると再現できる。

ただし、二元分割表そのものではなく、行カテゴリ、列カテゴリ、頻度を変数としたデータ行列を与える点に注意。かなり面倒である。

おわり