# PCAからICAへ?

○狩野 裕　清水昌平
大阪大学人間科学部

**0．はじめに**　統計学や，データ解析，信号処理，そしてニューラルネットワークなどの分野の中心的な問題は，多次元データの適切な線形変換を見つけることである．よく知られている変換の手法には，PCA(主成分分析)，FA(因子分析)，PP(射影追跡) などがある．

PCA(Principal Component Analysis) が Hotelling によって（再）発見されたのが 1933 年である．固有値解析というシンプルな道具を用いることで簡単に次元縮約できることから，長らく重宝な多変量解析の道具として用いられてきた．理論的な発展をかいつまむと，固有値の分布を論じ初めて主成分分析にモデルを導入し分布論を展開した Anderson (1963)，分散最大化や情報損失などの PCA の特徴づけを行った Okamoto (1969)，Flury(1988) の common PCA，高根 (1995) の制約付き PCA などがある．これらの発展は，主に（心理）統計学者によってなされてきた．一方，主成分分析と似て非なる分析方法に因子分析がある．これらは主に，目に見えない成分・因子を扱う多変量解析法である．

1990 年代になって，情報工学の分野から，突如として「独立成分分析 (Independent Component Analysis，ICA)」なる新しい主成分分析が提案され，注目を浴びている．その起源は，複数個の独立な信号が線型混合された観測値から，「独立」というキーワードを最大限使って，元の信号を取り出すという BSS(Blind Source Separation) の問題である[1]．つまり，独立な $n$ 個の信号 $\mathbf{s} = (s_1, s_2, \cdots, s_n)^T$ は観測できないが，混合行列 $A$ によって線型混合された $\mathbf{v} = (v_1, v_2, \cdots, v_m)^T$，すなわち

$$\mathbf{v} = A\mathbf{s} \tag{1}$$

のみが観測されるという状況から，混合行列 $A$ を推定し，信号源 $\mathbf{s}$ を再生するという問題である．一般には $m \geq n$ が仮定される．

**1．モデル**　数理的なモデルを紹介する．独立成分分析 (ICA)(e.g.,Comon, 1994; Jutten and Herault, 1991) は，確率変数の組を統計的独立な変数の一次結合として表現することを目的とする手法である．(1) のモデルが ICA のもっとも単純な形 (Comon, 1994) であり，観測 $v_i$ は，相互に統計的独立で平均ゼロの $n$ 個の未知独立成分 $s_i$ の一次結合であると仮定される．

ICA の基本問題は，混合物 $v_i$ から，独立成分 $s_i$ を推定する，つまり，混合行列 $A$ を推定することである．

測定に誤差が伴うという状況では

$$\mathbf{v} = A\mathbf{s} + \epsilon \tag{2}$$

となり，このモデルは因子分析モデルに他ならず，独立因子モデル，または noisy ICA とよばれている (Attias 2000)．$\mathbf{s}$ が独立に分布するという仮定を利用するとデータから因子回転を定めることができる (cf, Mooijaart 1985)．すなわち，因子スコアが，何らかの意味で最も独立に近くなるような回転解が推定される．

$s_i$ の確率密度関数 (p.d.f.) を $\kappa_i(s_i)$ とすると，独立性から，$\mathbf{s}$ の同時分布は $\prod_{i=1}^n \kappa_i(s_i)$ となる．誤差に正規性 $\epsilon \sim N(0, \Psi)$ を仮定すると，$\mathbf{v}$ の p.d.f. は

$$p(\mathbf{v}|A, \Psi, {\kappa_i}' s) = \int N(\mathbf{v}|A\mathbf{s}, \Psi) \prod_{i=1}^{n} \kappa_i(s_i) d\mathbf{s} \tag{3}$$

となる．これは，$\kappa_i$ の関数形を指定しないセミパラメトリックモデルである．

さて，$\mathbf{s} \sim N_n(\mathbf{0}, I_n)$ を仮定し，混合行列 $A$ として直交行列を考えてみる．このとき，$\mathbf{v}$ は $\mathbf{s}$ を再生するのにまったく無力であることが分か

[1] 心理学でいう「カクテルパーティ効果」と似ている．BSS の問題は独立分布という情報が使えるという点で少しやさしい．

るだろう．したがって，ICA の鍵の一つは「非正規」である．$\mathbf{s}$ の分布が同時正規であればICA は適用できない．これをモデルが識別可能でないという．ICA は非正規分布に基づく分析方法であり，非正規性を利用して因子回転を定める方法であり，一次と二次モーメント以外の情報も用いる分析方法である．

モデルの識別可能性を保証するために，以下の条件が課される．

1. 正規分布に従う独立成分は多くても 1 つで，残りの独立成分 $s_i$ は全て非正規である．
2. 観測 $\mathbf{v}$ の次元は，独立成分 $\mathbf{s}$ の次元以上である．つまり $m \geq n$.
3. 混合行列 $A$ の列ベクトルは一次独立である．

ただし，独立成分 $s_i$ のスケールも符号も推定できないし，独立成分の間の順序は定義されない．そこでスケールの不定性を消すために通常は独立成分 $s_i$ の分散を 1 に固定する．

ICA の興味深い応用として，先に紹介した BSS と特徴抽出があげられる．BSS(Cardoso, 1990; Jutten and Herault, 1991) では，$\mathbf{v}$ の観測値は，$m$ 次元の離散時間信号 $\mathbf{v}(t), t = 1, 2, \cdots$ の実現値に対応する．このとき $s_i(t)$ は，源信号と呼ばれる．特徴抽出とは，いわゆる主成分（因子）の解釈である．$A$ の列は，特徴を表し，$s_i$ は，観測データ $\mathbf{v}$ の $i$ 番目の特徴の存在と大きさを表している (Bell and Sejnowski, 1996a,1996b; Hurri et al.,1996).

**2．解析手順**　ICA の解析手順はおおまかに次のようになる．

1. PCA で前処理 (球状化).
2. 非正規性が最大になるように回転.

**PCA で前処理 (球状化)** 式 (1) で $A$ を推定するという問題は，データ $\mathbf{v}$ を球状化することによって，いくらか単純化できる (Cardoso 1990; Comon 1994; Oja and Karhunen 1995)．成分 $x_i$ が相互に無相関で，分散が 1 になるように, 観測ベクトル $\mathbf{v}$ を，$\mathbf{x} = M\mathbf{v}$ と線形変換する．すなわち，$E(\mathbf{x}\mathbf{x}^T) = I$．この変換は，常に可能で，PCA で成される．同時に，データの次元は，変換されたデータベクトル $\mathbf{x}$ の次元が $n$，つまり独立成分の数に等しくなるように減ぜられる．変換後，われわれは次式を得る．

$$\mathbf{x} = M\mathbf{v} = MA\mathbf{s} = B\mathbf{s} \tag{4}$$

ここで $B = MA$ は，独立成分 $s_i$ に関する仮定より，直交行列であり，$E\{\mathbf{x}\mathbf{x}^T\} = BE\{\mathbf{s}\mathbf{s}^T\}B^T = BB^T = I$．こうすることで，任意のフルランク行列 $A$ を求める問題を，$\mathbf{s} = B^T\mathbf{x}$ をあたえる直交行列 $B$ を求めるという，より単純な問題にすることができる．$B$ の第 $i$ 列を $\mathbf{b}_i$ と表すと，第 $i$ 独立成分 $s_i$ は，観測 $\mathbf{x}$ から $s_i = (\mathbf{b}_i)^T\mathbf{x}$ として計算できる．

**非正規性が最大になるように回転**．非正規性を測るメジャーとして，尖度や負のエントロピー，相互情報量などが考えられる．これらを最大，あるいは最小化するように回転する．

**3 尖度の最大化**　もっとも簡単な ICA 問題の解法は，4 次のキュムラント，つまり，信号 $v$ の尖度を用いる：

$$\mathrm{kurt}(v) = E\{v^4\} - 3(E\{v^2\})^2 \tag{5}$$

独立な確率変数 $v_1, v_2$ とスカラー $\alpha_1, \alpha_2$ に対して

$$\mathrm{kurt}(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1^4 \mathrm{kurt}(v_1) + \alpha_2^4 \mathrm{kurt}(v_2)$$

が成り立つ．

球状化された観測 $x_i$ の線形結合 $\mathbf{w}^T\mathbf{x}$ の中で，尖度が最大，あるいは最小となるものを求める．ここで $\|\mathbf{w}\| = 1$ と仮定する．(4) 式の直交混合行列 $B$ を用い $\mathbf{z} = B^T\mathbf{w}$ とおく．このとき $\|\mathbf{z}\| = 1$ である．式 (4) と尖度の性質から次式を得る．

$$\begin{aligned} \mathrm{kurt}(\mathbf{w}^T\mathbf{x}) &= \mathrm{kurt}(\mathbf{w}^T B\mathbf{s}) \\ &= \mathrm{kurt}(\mathbf{z}^T\mathbf{s}) \\ &= \sum_{i=1}^{n} z_i^4 \mathrm{kurt}(s_i) \end{aligned} \tag{6}$$

条件 $\|\mathbf{w}\| = \|\mathbf{z}\| = 1$ の下で一般に，関数 (6) は極大点・極小点をもつ．独立成分 $s_i$ の中でゼロでない尖度をもつ成分が少なくとも一つあると仮定する．このとき，(6) の極値は，通常，ある基底ベクトル $\mathbf{z} = \mathbf{e}_j$ で達成されることが示される．対応する重みベクトルは $\mathbf{w} = B\mathbf{z} = B\mathbf{e}_j = \mathbf{b}_j$，つまり，直交混合行列 $B$ の列である．以上より，式 (6) における尖度を最小化あるいは最大化す

ることによって，混合行列の一つの列が得られる．また，${\mathbf{b}_j}^T\mathbf{x} = {\mathbf{b}_j}^T B\mathbf{s} = s_j$ であるから，一次結合自体が独立成分の一つと等しくなる．

この結果は，独立な非正規成分の線型結合の中で非正規性を最大にするものは，成分自身であることを示している．中心極限定理は，独立成分を加算していくと正規分布に近づくことを示すので，一つの成分が一番非正規性が大きいということが期待される．これが上記結果の直感的な解釈である (Hyvarinen, in press)．

式 (6) はまた，正規分布に従う成分は，その尖度が 0 であるために，この方法では推定できないことを示している．ただし，正規分布に従う成分が 1 つだけの場合は，独立成分を抽出したあとの残りとして求めることができる．

実際に $\mathrm{kurt}(\mathbf{w}^T\mathbf{x})$ を最適化するために，グラディエント降下法をつかうことができる (Delfosse and Loubaton, 1995; Hyvarinen and Oja, 1996)．

目的関数は，入力が球状化されているため単純化でき，次式が成り立つ．

$$\begin{aligned}\mathrm{kurt}(\mathbf{w}^T\mathbf{x}) &= E\{(\mathbf{w}^T\mathbf{x})^4\} - 3[E\{\mathbf{w}^T\mathbf{x}\}^2]^2 \\ &= E\{(\mathbf{w}^T\mathbf{x})^4\} - 3\|\mathbf{w}\|^4 \qquad (7)\end{aligned}$$

制約 $\|\mathbf{w}\| = 1$ の扱いは種々考えられるが，たとえば，ペナルティー項を付加する方法が提案されている (Hyvarinen and Oja, 1996)．そのとき，最終的な目的関数は，

$$J(\mathbf{W}) = E\{(\mathbf{w}^T\mathbf{x})^4\} - 3\|\mathbf{w}\|^4 + F(\|\mathbf{w}\|^2) \quad (8)$$

となる．ここで，$F$ は制約 $||\mathbf{w}|| = 1$ を達成するためのペナルティー項である．具体的なペナルティー項がいくつか，Hyvarinen and Oja (1996) によって提案されているが，実は，$F$ の正確な形は重要ではない．

$\mathbf{x}(t)$，$\mu(t)$ をそれぞれ観測の列，学習率の列とし，$f$ を $F/2$ の微分係数とすると，オンライン学習アルゴリズムは，次のような形になる．

$$\begin{aligned}\mathbf{w}(t+1) = \mathbf{w}(t) \pm \mu(t)[\mathbf{x}(t)(\mathbf{w}(t)^T\mathbf{x}(t))^3 \\ -3\|\mathbf{w}(t)\|^2\mathbf{w}(t) + f(\|\mathbf{w}(t)\|^2)\mathbf{w}(t)] \quad (9)\end{aligned}$$

括弧の前の正の符号は，極大点を求めることを意味し，負の符号は，極小点を求めることに対応する．

Hyvarinen (1997) は，(9) の不動点を求めるというアイデアから Fast ICA のアルゴリズムを開発した．すなわち，Fast ICA は

$$E[\mathbf{x}(\mathbf{w}^T\mathbf{x})^3] - 3||\mathbf{w}||^2\mathbf{w} + f(||\mathbf{w}||^2)\mathbf{w} = 0$$

の解として $\mathbf{w}$ を定める．Hyvarinen (1999) は，この考えをさらに推し進め，Fast and Robust Fixed-Point Algorithm を提唱している．

Amari-Cardoso (1997) や川鍋-村田 (2000) は，セミパラメトリック統計モデルにおける推定関数の理論を用いて (noisy) ICA の推定アルゴリズムを提唱している．

**4．負のエントロピーの最大化**　尖度ははずれ値に対して大変敏感であり，非正規性の頑健なメジャーとは言えないことがある．そこで，負のエントロピーを非正規性のメジャーとする方法がある．$f(\mathbf{v})$ を $\mathbf{v}$ の密度関数とすると，エントロピー $H(\mathbf{v})$ は

$$H(\mathbf{v}) \quad = \quad -\int f(\mathbf{v})\log f(\mathbf{v})d\mathbf{v}$$

となる．負のエントロピー $J(\mathbf{v})$ (negentropy) は

$$J(\mathbf{v}) \quad = \quad H(\mathbf{v}_{gauzz}) - H(\mathbf{v})$$

と定義される．

二次モーメントが存在し，サポートが $R^1$ 全体である連続分布の族の中で，正規分布はエントロピーを最大化する．したがって，負のエントロピーは非正規分布に対して正の値をとり，大きな値は非正規性が大きいと解釈できよう．しかし, $f(\mathbf{v})$ は未知であるから, 負のエントロピーを直接計算するのは難しいので，いくつかの近似方法が考案されている．その一つにキュムラントをつかう方法:

$$J(v) \approx \frac{1}{12}\kappa_3(v)^2 + \frac{1}{48}\kappa_4(v)^2$$

があるが，先に指摘したのと同様，はずれ値に対して頑健でない．一方，最大エントロピー原理に基づく方法は，$J(v)$ を以下の量で近似する：

$$\begin{aligned}J(v) &\approx c[E\{G(v)\} - E\{G(\nu)\}]^2 \\ &= J_G(\mathbf{w}),\ \text{と書く}\end{aligned}$$

ここで $G, c, \nu$ はそれぞれ, 非二次の関数，定数，標準化された確率変数である．Hyvarinen (1999) は，$G$ の選び方や最大化のアルゴリズムを議論し，$E[\mathbf{x}g(\mathbf{w}^T\mathbf{x})] - \beta\mathbf{w} = 0$ の解として $\mathbf{w}$ を定めている．ここで $\beta = E\{\mathbf{w}_0^T\mathbf{x}g(\mathbf{w}_0^T\mathbf{x})\}$, $g(\cdot)$ は $G(\cdot)$ の微分係数である．

**5．PP, FA, PCA の関連**　(2) は，ノイズのあるデータを処理できるという意味で Noisy ICA

と呼ばれ，これに対して，式 (1) を Noise-free ICA という．

非正規性を最大にするという手順は，射影追跡 (PP, Projection Pursuit) と同等である (e.g., 岩崎 1990, 1991; 小山他 1998).

射影追跡は，多次元データの興味深い射影を見つけるための統計手法であり,「最も興味深い分布」をもつ方向の射影が探索される.「最も興味深い分布」は，正規分布から最もはなれている分布と定義される．つまり，正規分布が最もつまらない分布で，最も正規分布から遠いものが，最も興味深い分布であると考える．非正規性を最大にする射影（線型結合）を探すという意味でICA と同等と考えられる．ICA で，非正規性のメジャーとそのメジャーに対応したアルゴリズムは，PP では，射影指標と対応するアルゴリズムと考えられる．しかし，PP は独立成分という概念が不必要であり，理論的には，ICA より広い概念である．

ICA モデルが成り立つなら，ICA の非正規性のメジャーを最適化することは，独立成分を生み出し，モデルが成り立ってないなら，我々が得るのは射影追跡方向である．したがって，非正規データで，独立成分についてなんの仮定もなく，ノイズがないと仮定されているなら，ICA は，PP のような探索的データ解析と考えられる．ICA も PP も，データの興味深さ，つまり，非正規性が最大になる方向を探している．

Noisy ICA モデルは非正規データに対する FA である．FA では，因子空間を見つけたあと，回転を行うのが普通であるが，ICA はその回転である．ただし，その基準は $A$ の構造ではなく，因子の高次統計量に依存し非正規性の最大化である．これは，もし潜在因子が独立であれば，その独立成分を同定することと同等である．これは，前処理としてPCA で次元縮小して，さらに次元縮小することなくICA を実行することにおおよそ等しい．

PCA とICA の関連はあまりない．どちらともデータに線型変換を施すわけであるが，PCA は2次統計量のみをつかい，ICA は高次統計量をつかうというように基準がまったく異なる．PCA は分散を最大にし，ICA は非正規性を最大にする．また PCA は次元縮小を目的とするのに対し，ICA は次元を縮小することもあれば，拡大する (over-complete basis) こともあるし，維持することもあるという違いもある．

**6. さいごに**　近年，さまざまな分野で，PCA で分析されていたデータに対して ICA による分析が試みられている．しかし，ICA が PCA を凌駕するということはなく，目的によって使い分けるということになるであろう．分散最大化と非正規性最大化は，その意味も方法も結果もまったく異なるからである．

主成分分析・因子分析という伝統的かつ代表的多変量解析の手法を生み出し洗練させてきた計量心理学者や統計学者が，ICA の開発に遅れをとったことは残念である．ICA の方法論的な発展は山場を越えたと思われるが，その社会科学データへの応用は未知数である.「独立」というキーワードが社会科学の分野でどのように意味付けられるのか，さまざまなデータへの応用を通して検討されなければならない．

## Selected references


Amari, S. and Cardoso, J. (1997). Blind source separation: Semiparametric statistical approach. *IEEE Transactions on Signal Processing*, **45**, 2692-2700.

Cardoso, J-F. (1990). Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem. In *Proc. ICASSP'90*, 2655-2658.

Comon, P. (1994). Independent component analysis, A new concept?. *Signal Processing*, **36**, 287-314.

Delfosse,N. and Loubaton,P. (1995). Adaptive blind separation of independent sources: A deflation approach. *Signal Processing*, **45**, 59-83.

Hyvarinen, A. (1999). Survey on independent component analysis. *Neural Computing Surveys*, **2**, 94-128.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks.*

Hyvarinen, A. and Oja, E. (in press). Independent component analysis: Algorithms and applications. *Neural Networks.*

Hyvarinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, **9(7)**, 1483-1492

Hyvarinen, A. and Oja, E. (1996). A neuron that learns to separate one independent component from linear mixtures. In *Proc IEEE Int. Conf. on Neural Networks*, 62-67.

Mooijaart, A. (1985). Factor analysis for non-normal variables. *Psychometrika*, **50**, 323-342.

Oja, E. and Karhune, J. (1995). Signal separation by nonlinear hebbian learning. In Palaniswami, M. et.al., Editors, *Computational Intelligence - a Dynamic System Perspective*, 83-97.

岩崎 (1990). 射影追跡と多変量データ解析. 柳井晴夫他編 人間行動の計量分析. pp.95-111.

岩崎 (1991). 射影追跡. 計算機統計学. **4**, 41-56

川鍋-村田. (1999.12). 独立成分解析のセミパラメトリック推定と情報幾何. 情報幾何ワークショップ資料.

川鍋-村田. (2000.7). 正規測定ノイズがある場合の独立成分分析. 第 68 回日本統計学会.

小山-森田-水田-佐藤. (1998). 3 次元空間の射影追跡. 行動計量学会, **48**, 1-9.