

統計科学のフロンティア 5

多変量解析の展開

統計科学のフロンティア 5

甘利俊一 竹内啓 竹村彰通 伊庭幸人 編

多変量解析の展開

隠れた構造と因果を推理する

甘利俊一 狩野裕 佐藤俊哉
松山裕 竹内啓 石黒真木夫

岩波書店

編集にあたって

多変量解析の新しい方向

この巻は「多変量解析の展開」と名づけられているが、内容はふつうの多変量解析とはかなり異なっている。

これまでの統計的多変量解析の方法は、大きく分けて2つの分野からなっている。1つは1変量の場合の統計的推測、すなわち推定や検定の理論を多変量(ベクトル変量)の場合に拡張したものであり、ほとんどの場合、多変量正規分布を前提とする線形モデルに関する推定理論が扱われてきた。その中では検定統計量などの分布の正確な表現や、その漸近展開を求めることに多くのエネルギーが費やされる。

もう1つの分野は、多くの変量からなるデータを何らかの形で縮約して、そこから情報を得ようとするものであり、因子分析法や、主成分分析法がその代表である。その場合にも線形空間の中に、データを小さい次元に何らかの方法で射影することが中心であった。この場合、確率モデルが想定されることはあっても、関心はモデルの中の母数の推測よりも、データそのものの構造を見るにあった。

しかし、このどちらの場合にも、ベクトル変数の各成分は、いずれも形式的に対等なものとして扱われ、それらの変量の間には論理的な前後関係、あるいは因果関係の存在が仮定されていなかった。因果関係が想定される場合には、原因を表す変量は外生変数、または独立変数として、モデルの外で決定されるものとして扱われ、確率的に変動する変量はそれらによって決定される内生変数、または従属変数とされたのである。そして内生変数の外生変数に対する回帰関係が、因果関係を数量的に表現するものとしてモデル化されたのである。

■因果関係の解析

ところが場合によっては、というよりいろいろな分野の多くの場合に、「因果関係」が観測される変量の間に存在し、しかもそれが最初から明確に

与えられているのではなく、データから検出、または検証しなければならないことがある。それが因果性をめぐる問題であり、この巻の3つの部でくわしく論じられる。

まず最初に明確にしておかなければならぬのは、因果性の非存在、すなはちAの変化がBに影響を及ぼすことではなく、したがってAがBの原因とは考えられないということをデータから検証することはできるが、AとBが仮に同じ方向に動くことが明らかになったとしても、AがBの原因であるか、BがAの原因であるかをデータの上で決めるることはできないということである。つまり因果性の存在や、その方向を積極的に確立することは不可能である。このことは統計的方法にのみつかわることではなく、そもそも本質的に「因果性」は経験的事実だけからは言うことができないものなのである。

もちろんAの変化がBの変化に先行することが経験的に明らかになれば、BがAの原因でないことは確かに言えるが、それでもAがBの原因であることが立証されたといはることはできない。ここでAを気象の予報値、Bを実際の気象の観測値とすれば、予報がよく当たる場合、単に時間的な前後関係だけで因果性を判定すれば、予報が実際の気象の原因ということになってしまう。しかしこのことが馬鹿げていると思うとすれば、われわれは気象予報というものの構造を知っているからあって、もしデータとして予報値と現実の観測値の2つの系列しかなく、またその意味について何の説明もなければ、予報値が原因、現実の観測値が結果を表すと判断されることがあっても、とくに不自然ではない。

■因果関係のパターン

A,B, 2つの量の間に何らかの関係があると考えられる場合、その関係には因果性の方向を矢印で表すと次のようなパターンがある。

- (a) $A \rightarrow B$, (b) $B \rightarrow A$, (c) $A \rightleftharpoons B$, (d) $\alpha \xrightarrow[B]{A}$

ここで(c)は因果関係が双方向であり、その結果として均衡が成立する場合を、(d)は直接観測されない変量 α があって、それがA,B両者の原因となっているために、A,Bの間に相関関係が成立する場合を表している。そ

ここで A, B の間の関係がどのパターンに属するかは理論的に決められることであって、それをデータに照らして検証、あるいは検定することが可能であっても、純粹に経験的に決める事はできない。

前記の気象予報の場合については、次のような構造を考えることができよう。気象に関し、時間的に変化する基礎構造 $X(t)$ (たとえば気圧配置など)があつて、それが t 時点における気象 B を生み出す。一方、その 1 時点前において、その何らかの特性 A を観測することができる。そしてそれを手掛りとして B が予測される。すなわち基礎構造における因果系列 $\cdots \rightarrow X(t-1) \rightarrow X(t) \rightarrow \cdots$ を媒介として、A と B の間に相関が生じ、それによって予測が可能になっているわけである。

$$\begin{array}{c} \cdots \rightarrow X(t-1) \rightarrow X(t) \rightarrow \cdots \\ \downarrow \qquad \downarrow \\ A \dashrightarrow B \end{array}$$

統計的手法としてはパターン(a), (b)に対応するものは回帰分析である。そしてパターン(a)の場合、A が独立変数、B が従属変数、(b)の場合、A が従属変数、B が独立変数となることはいうまでもない。パターン(c)の典型的な場合は、マクロ計量経済学における同時方程式モデルである。これについては経済システムの中の均衡関係によって決まる内生変数 A, B のほかに、経済システム外部で決まる量 Z があつて、それがこのシステムに影響を与えていると見える。すなわち、次のようになる。

$$\begin{array}{c} A \rightleftharpoons B \\ \uparrow \\ Z \end{array}$$

このような変数 Z を考えなければならないのは、もしこのようなものが存在しなければ、A と B が均衡点に達してしまうとそれ以上変動しなくなるから、A と B の関係を知ることができなくなるからである。

具体的には次のようにモデル化される。内生変数を y_1, \dots, y_p とし、外生変数を z_1, \dots, z_q とする。各 y_i は他の y_1, \dots, y_n (y_i を除く) と z_1, \dots, z_q 、および偶然的な変動を表す攪乱項 u_1, \dots, u_n によって決定される。すなわち、

$$y_i = f_i(y_1, \dots, y_p, z_1, \dots, z_q, \theta_1, \dots, \theta_t, u_i), \quad i = 1, \dots, p$$

ただし $\theta_1, \dots, \theta_t$ は未知の母数, f_i は既知の関数である. z_1, \dots, z_q と u_1, \dots, u_p の値が与えられれば, y_1, \dots, y_p は上記の連立方程式の解として与えられることになる. このような方程式を構造方程式という. そうして連立方程式の解として与えられる y_1, \dots, y_p を

$$y_i = g_i(z_1, \dots, z_q, \theta_1, \dots, \theta_i, u_i), \quad i = 1, \dots, p$$

と表すことができる. これを誘導方程式という.

このようなモデル, とくに f_i が線形である場合の母数に関する推測問題は 1950 年代から 60 年代にかけてくわしく研究され, 体系的結果が得られている.

また, このようなモデルはマクロ経済学におけるケインズ理論と結びついて, 現実の経済予測や経済計画にもしばしば用いられた. しかしその後アメリカ, イギリスなどにおける経済学の主流がケインズ経済学からいわゆる新古典派経済学に移り, 「経済計画」の概念も用いられないようになって, マクロ同時計量経済モデルもあまり論じられなくなった. しかし, その統計的理論の有用性はとくにケインズ的経済学にのみ結びついているわけではない.

パターン(d)のモデルは, 古典的な因子分析法の考え方にも通じるが, 潜在変数モデルという形に表現されることもある. すなわち観測される変数 x_1, \dots, x_p の背後に, それらを決定する未知の(あるいは観測されない)変数 w_1, \dots, w_t (その個数 t も未知の場合もある)があって,

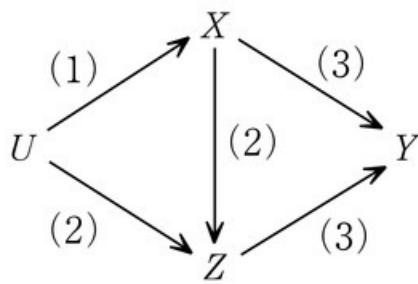
$$x_i = h_i(w_1, \dots, w_t, \beta_1, \dots, \beta_s, u_i)$$

と表されるとするものである. ここで h_i は既知の関数, β_1, \dots, β_s は未知母数, u_i は確率的誤差項である.

いろいろな対象分野において複雑な因果関係が存在している場合, 2つの量 X, Y の因果関係をどのように把握するかの問題がおこる. たとえばこれ以外に Z, U が存在し, 図のような因果構造が考えられるとしよう.

そこで矢印の数字に対応して次のような関係が成り立つとする.

- (1) $X = \alpha_1 + \alpha_2 U + e_1$
- (2) $Z = \beta_1 + \beta_2 U + \beta_3 X + e_2$
- (3) $Y = \gamma_1 + \gamma_2 Z + \gamma_3 X + e_3$



ここに e_1, e_2, e_3 はすべての変数と独立な(ただし互いに相関を持つかもしれない)攪乱項である。

このとき X の Y に対する因果関係の強さをどのように定義したらよいであろうか。

- (i) 最も簡単に考えれば、それは Y の X への回帰係数 γ_3 で与えられる。
- (ii) しかし X の Y への影響は直接的なもののほかに Z を通じる部分もある。そこで Z の式を代入して

$$Y = (\gamma_1 + \beta_1\gamma_2) + \beta_2\gamma_2 U + (\gamma_3 + \beta_3\gamma_2)X + \gamma_2e_2 + e_3$$

とすれば、 $\gamma_3 + \beta_3\gamma_2$ が X の Y に対する影響の大きさを与えると考えることもできる。

- (iii) さらに X も Y も共通の変数 U によって影響されている。そこで X と Y の関係だけに注目して

$$Y = \delta_1 + \delta_2X + e_4$$

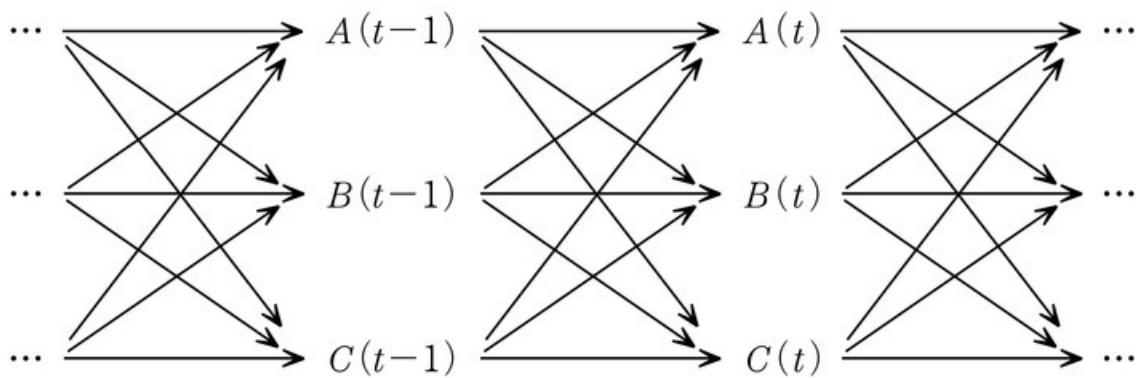
というモデルを想定すれば、係数 δ_2 は U を通じる影響を含むことになる。

このような定義の中でどれが適当であるかは問題による。もし $X \rightarrow Y$ を純粹に予測として考えるならば、つまり X の一定量の変化が観測されたとき、それに応じてどれだけが変動すると推定されるかが問題であれば、他の変数すべてと通じる間接的な関係も含めて考えなければならない。

これに対して、制御、つまり X を操作して変化させたとき、 Y がどれだけ変動するであろうかが問題であれば、それはそこで他の変数をどのように固定するか、あるいは X の変動に応じて変動することを許すかによって答えが変わってくる。

因果関係は原因変数の操作可能性と結びついて理解されなければならない。時間的前後関係が含まれる時系列データについては様相はもっと複雑に

なるが、一般に次のような形になる。



いくつかの時系列変数を、 $x_i(t)$, $i = 1, \dots, p$, $t = 1, 2, \dots$ とすると、一般にモデルは

$$x_i(t) = f_i(x_1(t), \dots, x_p(t), x_1(t-1), \dots, x_p(t-1), u_i(t))$$

という形になる。 $u_i(t)$ は独立な確率項である。これを $x_1(t), \dots, x_p(t)$ に関して解けば、

$$x_i(t) = g_i(x_1(t-1), \dots, x_p(t-1), u_1, \dots, u_p)$$

という形の誘導形が得られる。これは多変量自己回帰モデルと呼ばれるものである。ただし、この場合も f_i および g_i には未知母数が含まれているが、記号が繁雑になるのでここでは省略してある。このような関係から、 $y_i(t)$ の間の時間的前後関係を分析し、因果関係をチェックすることができる。

■正規性と線形性

上記のような問題において、これまでに展開した理論と手法はほとんどすべて、関係の線形性と確率分布の多変量正規性を前提としてきた。それは解析的に明確な答えが得られるのはほとんどそのような場合に限定されるからであり、またそれによって応用上有効な答えが得られる場合も多いからであるが、しかし現実に正規性や線形性が明らかに成立しない場合もある。

そのような場合には、何らかの非線形モデルや非正規モデルを想定しなければならないが、しかしそこには困難がある。とくに構造方程式から誘導形を解析的に導くことは、構造方程式が非線形になると一般に不可能になる。またその場合、解の一意性や連続性も保証されなくなる。構造方程

式内の外生変数の連続的な変化に対して内生変数が不連続に変化するような、いわゆるカタストロフィーがおこることもある。そのような場合を含むモデルの統計的解析はきわめて困難である。

しかし逆に、非正規、非線形モデルを想定すると、正規線形モデルを前提にすると求められないような答えが得られる場合がある。たとえば、前述のパターン(a),(b)に対応する 2 変数間の 2 つの線形回帰モデル

$$y = \alpha + \beta x + u \quad x \text{ と } u \text{ は独立}$$

$$x = \alpha' + \beta'y + v \quad y \text{ と } v \text{ は独立}$$

は正規分布を仮定すれば区別できない。すなわち、どちらの場合も x, y の同時分布が 2 変量正規分布になるからである。しかし関係が非線形であれば、2 つの関係式、たとえば、

$$y = \alpha + \beta x + \gamma x^2 + u \quad x \text{ と } u \text{ は独立}$$

$$x = \alpha' + \beta'y + \gamma'y^2 + v \quad y \text{ と } v \text{ は独立}$$

はまったく別の関係になって、相互に変換されることはない。したがって、このような場合には因果の方向性をデータから知ることができる。

同様なことがいろいろな場合におこることが本書の各部でくわしくのべられている。

(竹内啓)

目 次

編集にあたって

第Ⅰ部 独立成分分析とその周辺	甘利俊一	1
第Ⅱ部 構造方程式モデリング、因果推論、 そして非正規性	狩野裕	65
第Ⅲ部 疫学・臨床研究における因果推論	佐藤俊哉・松山裕	131
補論 A 分布の非正規性の利用	竹内啓	177
補論 B 多次元 AR モデルと因果関係	石黒真木夫	195
索 引	221	

装丁 蛾名優子

I

独立成分分析とその周辺

甘利俊一

目 次

1	信号の混合と分離——独立成分分析の枠組み	4
2	問題の定式化	6
3	独立成分分析, 主成分分析, 因子分析	8
4	確率変数の従属性コスト関数	13
5	最急降下学習法	18
6	自然勾配学習法	22
7	独立成分分析における最急降下学習	25
7.1	学習アルゴリズム	25
7.2	非ホロノームアルゴリズム	27
7.3	白色化——独立成分の数が少ない場合	28
8	推定関数と学習アルゴリズム	30
8.1	推定関数	30
8.2	推定誤差	32
8.3	推定関数を用いた学習アルゴリズム——学習の安定論	34
8.4	標準推定関数とニュートン法	38
8.5	諸パラメータの適応的決定	39
8.6	雑音のある場合の推定関数	40
8.7	観測信号の数が少ないとときのスパース解	41
9	独立成分の逐次の抽出	43
9.1	キュムラントに着目した抽出	43
9.2	確率分布とコスト関数	45
9.3	白色化した2段階アルゴリズム	47
9.4	独立成分の同時抽出	48
10	信号の時間相関を利用する方法	50
10.1	相関行列の同時対角化	50
10.2	時間相関がある場合の推定関数	51
11	時間的な混合とデコンボリューション	55
12	画像の分解と独立成分解析	57
	参考文献	62



独立成分分析(independent component analysis, ICA)という言葉が10年ほど前から登場し、評判になっている。これは、相互に確率的に絡まっている多数の変数を、その実現値を観測するだけで、独立なものに分解する新しい手法である。わかりやすい例は、カクテルパーティ効果である。何人かが同じ部屋で同時に話をしているとしよう。これを部屋に設置してあるいくつかのマイクで録音しておく。マイクに拾われた音は各人の話が混ざり合っているが、これから個々の人の話を分離することである。

これまで多変量解析では、標準的な手法として主成分分析(principal component analysis, PCA)が採用されてきた。また、因子分析(factor analysis, FA)も確率変量を独立な因子に分解する話である。しかし、これらは変量間の相関を主に考えていたために、こうした場合に適用できなかった。無相関と独立とは違うからである。独立なら無相関であるが、無相関であるからといって独立とは限らない。多変量が正規分布(ガウス分布)に従うときにのみ、無相関が独立になる。

独立成分解析は、確率に従うと想定される多変量の新しい分解の方法を与える。これには、理論と手法の新しい枠組みが必要になる。たとえば、神経回路網の学習の手法、相間に代わる高次のキュムラントの利用、情報幾何などである。

また、応用の分野も開けてきている。音声認識はもとより、画像情報の分解と処理、さらには携帯電話などの混線やエコーの消去などに活躍する。医用情報の分野では、脳の活動を外部から測定するMEG(脳磁計), fMRI(磁気共鳴画像)などのデータから、脳内の独立な情報を分離して取りだす試みへの応用が有力である。これはさらに心理学的な測定データや、社会科学的な測定データの解析にも応用されよう。

本稿では、独立成分分析の理論の枠組みとその応用にまつわる話をわかりやすく述べてみたい。

1 | 信号の混合と分離——独立成分分析の枠組み

いくつかの信号源があって、離散時間 $t=1, 2, 3, \dots$ にそれぞれ信号 $s_1(t), \dots, s_n(t)$ を発生するとしよう。信号の個数は n 個である。それぞれの信号は確率的に独立であるとしよう。各信号については、時間的には独立であってもよいし、相関をもっていてもよいが、定常的であるとする。 n 人の人が勝手に発話する場合は、各人の音声は時間的に独立ではない。だがしばらくは簡単のため、時間的にも独立であるとしておこう。

1つの部屋で n 人が話をしているとしよう。部屋のどこかにマイクがあると、 n 人の話がすべて混合して入り、録音される。話者がマイクから遠ければ、音は減衰するから、混合した音は

$$x(t) = \sum_{i=1}^n A_i s_i(t)$$

のように書ける。係数 A_i は話者 i とマイクの距離による。マイクを異なる場所に m 個仕掛けておけば、第 j 番目のマイクに入る音は、

$$x_j(t) = \sum_{i=1}^n A_{ji} s_i(t), \quad j = 1, \dots, m$$

である。マイクの場所が違うから、 A_{ji} はそれぞれの距離関係に応じて決まる。これは、空間的に線形な瞬時的な混合である。本稿ではこれを主に扱うが、時間的な混合もある。

1つの信号 $s(t)$, $t=1, 2, 3, \dots$ があったとしよう。このとき、信号は時間的に独立であるとする。さて、この信号を携帯電話などの無線で送ろうとすると、電波は山やビルで反射して時間がずれて重なり合う。つまり、受信する信号は

$$x(t) = \sum_k A_k s(t-k)$$

のようになって、過去の信号が現在の信号に混ざり合う。これが時間的混

合である。多数の信号が時間的にも空間的にも混ざり合うのが時空間的な混合である。

いま、空間的に混合した信号 $x_j(t)$, $j=1, \dots, m$; $t=1, 2, \dots$ が測定されたとしよう。この情報だけを使って、元の信号 $s_i(t)$ を復元するのが独立成分分析である。混合の行列 $\mathbf{A} = (A_{ji})$ は未知である。元の信号 $s_i(t)$ が互いに独立であるということだけが頼りである。

少し問題を数学的に整理してみよう。 n 個の信号をまとめてベクトル

$$\mathbf{s}(t) = [s_1(t), \dots, s_n(t)]^T$$

で表すことにする。信号を縦ベクトルで表すことにし、 T で転置を表す。測定値もベクトル

$$\mathbf{x}(t) = [x_1(t), \dots, x_m(t)]^T$$

となる。混合の係数をまとめて、 $m \times n$ 行列

$$\mathbf{A} = (A_{ji})$$

で表そう。このとき、信号の混合は

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

と書ける。時間混合、さらに時空間混合の場合は

$$\mathbf{x}(t) = \sum_k \mathbf{A}_k \mathbf{s}(t - k)$$

となる。ここで \mathbf{A}_k は k 時間遅れた信号との混合を表す行列である。時間混合はフィルターをかけることであるから、時間遅れの混合をコンボリューション * を用いて表し、時空間混合を、

$$\mathbf{x}(t) = \tilde{\mathbf{A}} * \mathbf{s}(t)$$

と書いててもよい。ここで $\tilde{\mathbf{A}} = (\mathbf{A}_1, \dots, \mathbf{A}_k, \dots)$ がフィルターの行列である。

信号の混合は線形であるから、分離も線形ができる。そこで、行列 \mathbf{W} を用いて測定信号 $\mathbf{x}(t)$ を線形変換し、元の信号の候補

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$$

が得られるとしよう。（時空間混合の場合は、 \mathbf{W} はフィルターからなる行列でこれもまたコンボリューションとなる。） \mathbf{A} は未知であるから、その逆演算 \mathbf{W} も未知である。問題は、測定値 $\mathbf{x}(t)$ だけを頼りに、うまい \mathbf{W} を見つけ出すことである。

こんなことが可能かといえば、可能である。たとえば、 \mathbf{W} をいいかげんに選んで $\mathbf{y}(t)$ を求めてみる。このとき、 $\mathbf{y}(t)$ が空間的に、つまりその異なる成分がすべて独立に分布していれば成功であるから、成功するように \mathbf{W} を探していけばよい。もちろん、手探りはいけない。なにかうまい探索手法がありそうではないか。

測定データ $\mathbf{x}(t)$ を全部集めておいて、一括して \mathbf{W} を探そうというのが、バッチ処理である。これに対して、現在の候補 $\mathbf{W}(t)$ があるときに、次の時間の測定データ $\mathbf{x}(t+1)$ を見て、これを少し変更して次の候補 $\mathbf{W}(t+1)$ に改良していくのが、学習とよばれる手法である。これは神経回路網における学習の考え方からきているが、統計学的に言えば、逐次推定法である。これは混合行列 \mathbf{A} が時間と共にゆっくりと(または突然に)変わるときにも、この変化に追従できるので都合がよい。カクテルパーティでは人は動き回るし、移動携帯電話でも混合フィルターは移動につれて変わるからである。

2 | 問題の定式化

信号分離の問題を、いちばん簡単な線形瞬時空間混合、しかも $m=n$ の場合に数学的に定式化しておこう。もっと一般の場合も似たような仕方でできる。測定数 m のほうが信号源の数 n より大きいときは簡単であるが、逆の場合は難しい。これについては後にふれる。

信号 s_i の従う確率密度関数を $r_i(s_i)$ としよう。すると、どの時間 t においても、信号のベクトル $\mathbf{s}(t)$ の従う確率分布 $r(\mathbf{s})$ は積の形の密度関数

$$r(\mathbf{s}) = r_1(s_1) \cdots r_n(s_n)$$

で表される。これが独立性の仮定である。では、測定信号 $\mathbf{x}(t)$ の分布はどうなっているのだろうか。 \mathbf{x} と \mathbf{s} とは

$$\mathbf{x} = \mathbf{As} \tag{1}$$

で結ばれている。逆に、

$$\mathbf{W} = \mathbf{A}^{-1} \tag{2}$$

とおけば、

$$\mathbf{s} = \mathbf{W}\mathbf{x} \quad (3)$$

である。測定信号 \mathbf{x} の確率密度関数を $p(\mathbf{x}, \mathbf{W})$ とおこう。これは $\mathbf{W}=\mathbf{A}^{-1}$ に依存する。 \mathbf{x} と $\mathbf{x} + d\mathbf{x}$ の間に信号ができる確率は $p(\mathbf{x}, \mathbf{W})d\mathbf{x}$ であるが、これは対応する信号が \mathbf{s} と $\mathbf{s} + d\mathbf{s}$ の間に出て来る確率に等しい。

$$p(\mathbf{x}, \mathbf{W})d\mathbf{x} = r(\mathbf{s})d\mathbf{s}$$

ここで、 $\mathbf{s}=\mathbf{W}\mathbf{x}$, $d\mathbf{s}=|\mathbf{W}|d\mathbf{x}$ を代入する。 $|\mathbf{W}|$ は行列 \mathbf{W} の行列式である。これより \mathbf{x} の確率分布

$$p(\mathbf{x}, \mathbf{W}) = |\mathbf{W}|r(\mathbf{W}\mathbf{x}) \quad (4)$$

が得られる。

これが、測定データ \mathbf{x} の確率密度関数である。これは未知のパラメータ $\mathbf{W}=\mathbf{A}^{-1}$ を含んでいる。したがって、測定されたデータ $\mathbf{x}(1), \mathbf{x}(2), \dots$ をもとに未知の行列 \mathbf{W} を推定しようというのは統計モデル式(4)を用いた通常の統計的推定問題に帰着されることがわかる。しかし、この統計モデルはパラメータ \mathbf{W} 以外に、未知の関数 $r(\mathbf{s})$ を含んでいる。 $r(\mathbf{s})$ は、各変数の積に分解されることは独立性からわかっているが、成分の n 個の関数 $r_i(s_i)$ は未知である。してみると、確率分布を定める未知のパラメータは、 n^2 個の成分をもつ行列 \mathbf{W} のほかに n 個の関数を含み、しかも関数は無限大の自由度を含む。このような統計モデルはセミパラメトリック統計モデルとよばれ、たいへん厄介な代物である。

最後に、推定の不定性について触れておこう。われわれは、 $\mathbf{y}=\mathbf{W}\mathbf{x}$ により分解された \mathbf{y} の成分 y_i の独立性を頼りに \mathbf{x} を分解し、そうなるように \mathbf{W} を求めるのであった。しかし、独立な成分が得られたとして、それが 1 番目であり、それが 2 番目か、これは皆目わからない。独立成分の順番のつけ方は任意である。だから順番のつけ方は不定のままであるが、われわれは成分が得られればよいのだからこれはかまわない。ところで、分解された信号 y_i の平均の大きさはどうだろうか。元の信号 s_i の大きさが 2 倍になっても、未知の行列 \mathbf{A} の第 i 列が 2 分の 1 になっていれば測定データ \mathbf{x} は同じである。だから、 \mathbf{A} の各列の絶対的な大きさ、または復元した成分 y_i の絶対的な大きさは、不定のままで残る。しかし、時系列 $s_i(t)$ の

時間波形が復元できれば、全体の大きさは一応気にしなくてよいであろう。このことを数学的にいえば、行列 \mathbf{W} は、順番のパーミュテイション(これを行列 \mathbf{P} で表す)と、各信号のスケールを決める対角行列 Λ の不定性を除いて決まる。つまり得られる答えは、 \mathbf{A} の真の逆行列を \mathbf{W} としたとき、

$$\mathbf{W}' = \mathbf{P} \Lambda \mathbf{W}$$

であることになる。

以後このことは気にせず、どの \mathbf{W}' でも真の解であるとする。もちろん、信号の混合過程は物理的に決まる。たとえば、カクテルパーティ効果ならば、 \mathbf{A} の各成分は距離によって決まり、すべて正である。脳の情報でも、脳内の物理的な配置から、 \mathbf{A} には自然の制約が入る。こうした情報から不定性を減らすことができる。こうしたサイド情報は、利用できる限り利用したほうがよい。ただ、理論はこうした情報がないときでも成立するアルゴリズムを求める。理論家としてはそれは当然であるが、実際の問題に独立成分分析を使うときは、理論家の勝手ばかりを聞く必要はない。応用すべき問題に即してさらにいろいろな工夫ができるし、理論もそのことを考えて発展させていかなければいけない。

3 | 独立成分分析、主成分分析、因子分析

多変量の確率変数 \mathbf{x} を考えよう。これは確率分布 $p(\mathbf{x})$ に従うものとする。 \mathbf{x} が独立信号の混合として得られる場合には、行列 \mathbf{A} を $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ のように列ベクトル \mathbf{a}_i を並べたものとすれば、

$$\mathbf{x} = \mathbf{As} = \sum_{i=1}^n s_i \mathbf{a}_i \tag{5}$$

と書ける。第 i 列のベクトル \mathbf{a}_i は第 i 番目の信号 s_i が、測定信号 \mathbf{x} にどのような割合で影響するかを示す。式(5)は、ベクトル \mathbf{x} を新しい基底 $\{\mathbf{a}_i\}$ で表すものとみなせる。このとき成分 s_i は確率的に独立である。

主成分分析は、 \mathbf{x} が正規分布に従うものと想定することが多い。 \mathbf{x} の期