

心理学者のための線形代数

統計モデルの数学的基礎



小杉 考司

この本は Creative Commons BY-SA(CC BY-SA) ライセンス Version 4.0 に基づいて提供されています。著者に適切なクレジットを与える限り、この本を再利用、再編集、保持、改訂、再頒布 (商用利用を含む) をすることができます。もし再編集したり、このオープンなテキストを変更したい場合、すべてのバージョンにわたってこれと同じライセンス、CC BY-SA を適用しなければなりません。

<https://creativecommons.org/licenses/by-sa/4.0/deed.ja>

This book is published under a Creative Commons BY-SA license (CC BY-SA) version 4.0. This means that this book can be reused, remixed, retained, revised and redistributed (including commercially) as long as appropriate credit is given to the authors. If you remix, or modify the original version of this open textbook, you must redistribute all versions of this open textbook under the same license - CC BY-SA.

<https://creativecommons.org/licenses/by-sa/4.0/>

心理学者のための線形代数

小杉 考司

Last Compiled on 2025.12.3

はじめに

これは、「深層学習は心理学に革新をもたらすか？オープンサイエンスによる理論的・実践的検証」というプロジェクト^{*1}の一環で、2022年夏の数学ワークショップを担当した時の資料です。3日間のブートキャンプ、集中講義のために作られた資料ですので、Day1,2,3という3部建てになっています。初日は**線形代数の計算を身体化すること**を目標にしました。2日目は、**線形代数という数字の世界を空間のイメージに接続することが目標**です。3日目はそれらを踏まえて、**多変量解析を線形代数を通じて理解することが目標**になります。多変量解析は、心理尺度を作成する時や調査研究の文脈で使われることが多く、むしろ数学を苦手としている心理学者が好んで使っている印象があります。よくよく考えるとこれは怖いことで、理屈を知らなくても実践ができてしまうという状況は是正されるべきです。本書がその改善への一助になることを願っています。

数学書と本書の読み方

なんでもそうですが、教える人は教えるべき内容を十分に広く深く理解している必要があり、教壇で教えることができるのはその一部を結晶化したものに過ぎません。線形代数の世界は、数学の様々な領域を広く繋ぎ合わせるようなところがありますから、これを教えるには全ての数学に通じていなければならない。であれば私は明らかに力不足です。だって文系の心理学者なんだから(笑)。

それでも講師のお仕事をお引き受けしたのは、線形代数の「心理学的な」使い方、という観点から説明できるのは、やはり数学者ではなく心理学をやってきた人間だろうと思ったからです。たとえば因子分析などは、数学的に見れば行列のスペクトル分解という一技術に過ぎません。しかし心理学者は、その技術に「潜在変数の意味」を見出したり、また実践的な必要に応じて「因子軸の回転」や「単純構造の原則」などの応用技術を発展させてきたのです。アンケート調査をして心理尺度を作る、といった具体的な研究目的に即した観点で、時には数学的な原理とは違う基準があり、それを説明する観点を持っているのは、やはり心理学出身の人間でないと難しいだろうからです。

その意味で、この本は数学と心理学のグラデーションを埋める中間に位置することを狙っています。半魚人が海でも陸でも半分ずつ不便を被るのと同じように、本書は数学的な観点から見れば不完全で、心理学の観点から見れば不親切に映る、誰にとっても不愉快な一冊かもしれません。これは、数学的なテキストの書き方と、それ以外の一般的な読み物・書き物との方針が異なるからであり、その間を狙うとどっちつかずのものができてしまうからなのです。

今から始めようとするところに、いきなり言い訳から始めるようですが、少しこの「方針の違い」について説明しておきます。

テキスト、とくに数学のテキストは、定義をしてそこからわかったことを証明したり計算したり、と進んでいきます(定義 → 定理 → 証明。竹山(2022)も参照)。そして定義からできあがるものだけで、世界を作り上げます。数学の良いテキストは、その本だけでその本の世界が完成しています^{*2}。

このとき、出発点になる「定義」は必要最小限のことしか書いていないことに、注意が必要です。心理学の研究をしている人は、人間は見たまま、聞いたままを受け取るのは苦手な、認知のバイアス等で言外の情報を読み取ったり、せっかく覚えたことも時間と共に歪ませてしまったり、ということだらけなことをよくご存知のこと

^{*1} <https://sites.google.com/view/dlpsychology>

^{*2} 本によって定義が異なれば、同じテーマであっても、その本の中で違う世界を作っているということもあることにも注意してください。あの本ではAと言っていたのに、この本ではそうっていない、というのは作っている世界が違う可能性もあるのです。

と思います。書いてあることを書いてあるまま読む。書いてないものは読まない。これが意外と難しいのです。数学は必要な情報だけに削ぎ落として定義しますから、ここを歪ませてしまうと誤解してしまいます。この最小限度のピースだけから世界を作り上げていきますから、そっけないとか冷たいという印象をもってしまうのです。逆に、書いてあることだけから、世界を作り上げているんだと思って、その削ぎ落としの技術を味わえるようになると、数学書の読み方もかわってくるでしょう^{*3}。

また、定義は、定義してしまうとそれ以上変化することがありませんから、うまく定義してやると世界の基盤として大変有用な働きをします。つまり、定義することでこういうことが言える、といったメリットがあるのです。気付にくいことですが、定義をする側からすれば、定義することに何かしら狙いがあり、定義できると嬉しいことがあるわけです。我々読者としては、定義は定義として字義通り受け入れ、それを受け入れてやることで何が嬉しいのか、何の役に立つのか、という観点で「読みこなす」といいでしょう。

また、「さんすう」から「数学」に変わったときのように、大学以降で扱う数学は、定義や定理からわかる抽象世界の方が重要で、具体的な事例についての重要度は下がります。算数のときは、みかんが2つとりんごが3つ、あわせていくつ、のように具体的な事象に対応した数の性質を考えていました。みかんとりんごはひっつけられないよ、とか、ジュースにしたらまぜれるよ、といった冗談が有効だったのは、具体的事例を優先するからです。そうではなく、 $2+3$ という数だけの話になり、中学校の数学になると実数、負の数、高校になると虚数まで含めた、「数という世界」の方が優先されていきます。虚数に至っては、もはや具体的な事象と対応がつかなくても、そういうものがあるとして計算すると理論的に世界が整うから有用、ということになっています。

本書で扱う線形代数では、ベクトルという数字のセットを扱います。ベクトルが何なのか（データなのか、位置や力の関係なのか）はさておき、ベクトル空間というものを定義して、そこに含まれるものの性質を考えていく、というのが数学的な発想です。しかし私たち心理学者は、それがデータ、つまり個人や項目の何かを広げていったもの、というイメージをしますから、その空間ってなんなのさ、と想像が行き詰まるのです。抽象 → 具体ではなく、具体 → 抽象で行こうとすると、理屈を紡いでも辿り着けなくなります。抽象的な定義をまず受け入れてその恩恵を味わう、という方針に、いつの間にか説明の向きが変わっていることに気づけば、理解はぐっと進むとおもいます^{*4}。

翻って、数学書でないものはどうなっているか考えてみましょう。小説などは言葉を定義するのではなく、登場人物の関係性を文脈から類推することを楽しみます。なんなら読者ごとに解釈が違う本こそ、味わいがあっていいと評価されたりします。専門書であっても、人文社会科学の本は、読者の経験や実感などにその主張の確からしさを依存したり、著者の主張を批判的に考察し、主張の背景も含めて評価するといった姿勢が必要です。なんということでしょう、数学書の方針と真逆と言ってもいいほどではありませんか。であれば、後者のスタイルで本を読むことに慣れている人は、数学書の揺らぎのないテキストの中で完成された世界は読みにくく、共感しにくいものでしかないでしょう。もちろん逆もまた真なり、です。

つまり、書き方や読み方が違うこと、その違いに気づけないので楽しめない、というのが数学嫌いを生む原因になっていると思われます。本書は文系的な読み方や味わいがわかるように、数学書にしては地の文での解説が多く、意味が不完全にしか伝わらない記述になっているところがあるかもしれません。しかしそれは、本書が「数学書」と「心理学書」の間に位置付けられるように狙っているからであり、数学的な厳密さについては数学書を、心理学的な実践的意義については心理学書を参照していただければと思います。

^{*3} ショートショートというジャンルを開拓した星新一は、「短くならない文章はない」といってストーリーに必要な情報だけに削ぎ落とした文学を作りましたが、その楽しみ方と同じような感じです。

^{*4} この辺りの葛藤は、「数字であそぼ」という漫画（絹田、2019）で楽しく解説されているので、ぜひ一読ください。

言い訳が長くなりました。本書が心理学徒にとって新たな視点を得る機会になることを願いつつ、線形代数の世界にご案内したいとおもいます。

目次

第 I 部	Day 1: 行列の基礎と方程式	7
第 1 章	行列計算の基礎	9
1.1	線形代数とは何なのか	9
1.2	行列の定義	11
1.3	行列の四則演算	14
1.4	その他の行列の操作	20
1.5	練習問題	23
第 2 章	行列とデータ	25
2.1	記述統計の復習	25
2.2	行列を使った計算	27
2.3	線形モデルの行列表現	33
2.4	デザイン行列	34
第 3 章	行列と方程式	39
3.1	方程式を行列で書く	39
3.2	連立方程式を解こう	39
3.3	連立方程式を判定する	45
第 II 部	Day 2: 行列とベクトルの空間的理解	53
第 4 章	ベクトルの空間的理解	55
4.1	ベクトルの可視化	55
4.2	ベクトルとベクトルの関係	59
4.3	データの空間的イメージ	64
第 5 章	ベクトル空間	69
5.1	ベクトル空間	69
5.2	部分空間と直交補空間	72
5.3	ランクと射影	74
第 6 章	データ空間	85
6.1	線形回帰を行列と空間で理解する	85

6.2	分散共分散行列	93
第 III 部 Day 3: 行列と多変量解析		97
第 7 章	固有値と固有ベクトル	99
7.1	固有値と固有ベクトル	99
7.2	固有値と固有ベクトルの幾何学的意味	105
第 8 章	因子分析法	107
8.1	心理学における測定のモデル	107
8.2	因子分析モデルを展開する	110
8.3	因子分析モデルの行列表現	113
8.4	主成分分析と因子分析	117
第 9 章	共分散構造	121
9.1	パスダイアグラムと方程式	121
9.2	構造方程式	123
9.3	測定方程式	125
9.4	複雑なモデルの場合	126
9.5	構造方程式モデルによる未知数の推定	129
第 10 章	まとめにかえて	131
10.1	さまざまな多変量解析	131
10.2	数学書案内	132
付録 A	行列操作の R コード	135
A.1	行列の基本変形	135
A.2	固有値の近似演算	137
付録 B	ギリシア文字一覧	143
引用文献		143
索引		145

第 I 部

Day 1: 行列の基礎と方程式

第 1 章

行列計算の基礎

1.1 線形代数とは何なのか

1.1.1 便利な表現方法としての線形代数

今から線形代数を学んでいくわけですが、はじめに「線形代数とは何で、なぜ (心理学者が) 学ぶのか」について説明しておきましょう。というのも、線形代数の数学書は世の中にたくさんあるわけですが、理由についてはせいぜい、多変量解析をするにあたって行列は必須だ、と書いてあるぐらいだからです。まるで野球の楽しさを知るためにルールブックを読め、と言っているようなものですから、文系の心理学者にとって、これでは面白みが見えてこないでしょう。

このコースの受講生は、心理統計はひととおり学んでいる、統計ソフトウェアは使ったことがあるというぐらいの基礎知識を持っている人を前提にしていますから、データがスプレッドシートに入っているシーンを想像してください、といえバイメージしてもらえますでしょうか。心理学ではデータを取って、一行に一人分のデータを入れていく、変数は列方向に並べていく、というルールが基本です^{*1}。ここまでの説明ですでに「行」「列」という言葉を使ってしまいましたが、このようにデータは**行列 (matrix)** で全体を表現できます。

心理学者はこのデータ行列から、意味のある知見を取り出そうとして、平均や分散の計算をしたり、相関係数や回帰係数の計算をしたりするわけですから、このデータ行列にすべての情報が詰まっているはず。表計算ソフトでは、列毎、行毎の統計量を計算できますが、それはデータの一部分に限定した使い方ですね。もちろん相関係数など複数の行をつかって計算することはありますが、数十の変数列を一気に操作するには表計算ソフトではなく、統計パッケージが必要ですね。その違いは、行列の計算をするか、線形代数の操作をするかどうかだと言ってもいいでしょう^{*2}。つまり、複数の変数を同時に使う分析、たとえば重回帰分析や因子分析などがそうですが、そこには線形代数というツールが裏で働いているわけです。

つまり線形代数は、複数の数字のセットを扱うための計算方法だということです。文字や記号をつかった数学や方程式、関数といえば、 x, y などアルファベット 2, 3 種類が精々だったかもしれませんが、YG 性格検査などは 120 項目あるわけですから 120 文字必要で、そうなるともっと一般化した表現、まとめて扱うツールが必要ですね。それが線形代数なのです。

線形代数は数学の歴史からいくと新しいジャンルだ、と思っていただければ良いかと思います (その成立は 1800 年台半ばのようです)。変数をまとめて、リストにして扱うわけですから、今までとは違う書き方をし、違

^{*1} これはワイド型、あるいは横長型のデータ入力の方法です。これとは別に、ロング型・縦長型のデータという入力方法もあります。またデータの入力規則としては、統計ソフトウェアが計算しやすいデータの持ち方として Wickham (2014) が提唱した**整然データ (Tidy Data)** という考え方も知っておいて損はありません。

^{*2} ええ、もちろんわかっていますよ、分析ツールや VBA をつかえばエクセルでお行列の計算ができるじゃないか、というご意見があるのは。

う計算ルールが必要になります。利点は「まとめて扱う」ことを最初から目的にしていますから、変数3つだろうが4つだろうが、120個あっても同じように表現できるようになっているところ。つまり、データのサイズについて、たかだか有限個にすぎませんから、そのなかでの一般化した話ができるということです。

線形代数の表現に慣れると、変数のリストのまま、スプレッドシート全体像をイメージしたまま、加減乗除の加工ができるようになります。非常にサイズが大きくなることもあり得ますし、なにより新しい表現方法、計算方法に慣れないといけません。線形代数が難しいのでは、と身構える人もいるかもしれませんが、ここでの難しさは、正確に言うと「単純な操作だがたくさん反復しないといけないので、途中でミスが生じやすい」ことによる成功体験の得難さ、つまり「煩雑さ」が原因です。複素関数や測度論などのような「抽象的すぎて理解が困難だ」というタイプの難しさではないのです。逆に言うと面倒だけどシンプルなので、計算のパターンが身につさえすれば、あとは反復計算が得意な計算機に任せてしまって構いません。自転車に乗る練習をするように、最初は根気が必要ですが、後は雰囲気を使うことができる便利なものだと思います。

1.1.2 線形代数を知ると見えてくる世界

線形代数は数字のリストを扱う、という話をしました。今まで習った数字・変数のリストといえば、連立方程式がおもいつくかもしれません。まさに、線形代数は連立方程式を簡単に扱うための表現方法という側面もあります。

もうひとつ、複数の数字をまとめて扱うのは座標などの空間の話です。 $(x, y) = (2, 3)$ のように座標は2つの数字が組になって、ある位置を表すのでした。3次元空間でしたら、 (x, y, z) のように3つ組で表現します。線形代数はこうした空間的表現にも、新しい理解やイメージを与えてくれます。連立方程式の解を関数の交点で求めたように、式を空間で表現することもできるわけです。

さて私たちは120項目からなる1000人分のデータセットを分析する、なんてことをしたりするわけですが、これも120次元、1000次元の空間として表現してもいいですね。もちろん可視化するのは3次元が限界ですが、数学の世界の中では n 次元空間と一般化して考えることができます。このように、データをベクトルとして考え、それがなんらかの空間を持っている、という想像することが可能です。ではその空間において、回帰分析はどういう意味があるのでしょうか？主成分分析や因子分析はどういうことをしているのか、空間のメタファーで考えるとどうでしょう？このように、多変量のデータ分析の空間的なイメージ、幾何学的な想像をはたかせることで、分析モデルの理解が一層進むことが期待されます。たとえば因子分析においては因子数の決定をするときにスクリープロットを見るという方法がありますが、あれは何をみていて、どういう意味があるのかは、線形代数の空間的イメージを経由すると直感的に理解できます。あるいはまた、重回帰分析において多重共線性があると良くないという話がありますが、どうしてダメなのかということについても図式的に理解できます。

データセットを数字の羅列と見て計算することもできるし、データセットを空間にあるベクトルと考えることもできる、この両方の視点を提供してくれるのが線形代数です^{*3}。何事についても、多角的な理解をすること、さまざまな言葉で表現できることが「わかっている」ということの本質だと思います。皆さんも数字のリスト、その作る空間を同時に想像しながら、自らの分析モデルの理解への一助としてください。

^{*3} 一般にデータサイエンティストやコンピュータサイエンティストは、数字の羅列として考えることが多いでしょう。これに対して、物理学では空間における力の表現としてのベクトル、つまり始点・方向・大きさがセットになったものの表現として線形代数を扱います。数学ではこの両方の視点を行き来できるような、一般的な操作を考えています。数字のリストと空間の表現のいいとこ取りをテーマに作られた、Essence of linear algebraというYoutube動画(<https://bit.ly/3abXiFz>)は一見の価値があります。

1.2 行列の定義

1.2.1 行列とベクトル

行列やベクトルは、複数の数字をひとまとめにして扱うためのものです。まずはその基本的な形からみていきます。

ベクトル

複数の数字を一行、あるいは一列にまとめて表現したものを、**行ベクトル (row vector)**、**列ベクトル (column vector)** といいます。

行ベクトルは次のように表します。

$$\mathbf{a} = (a_1 \quad a_2 \quad \cdots \quad a_m)$$

列ベクトルは次のように表します。

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

具体的には、 a_1 とか b_2 のところには数字が入っています。つまり次のような形です。

$$\mathbf{a} = (1 \quad 3 \quad 5), \mathbf{b} = \begin{pmatrix} 2 \\ 4 \\ 6 \\ 12 \\ 8 \end{pmatrix}$$

とくに断りがなければ、本書でベクトルといえば基本は列ベクトルであると思ってください。

ここで今回の $\mathbf{a}$ は 3 つの要素が入っていますので、サイズは 3、同じく $\mathbf{b}$ はサイズが 5 のベクトルです。行列の言い方に合わせて 1×3 の (行) ベクトル、 5×1 の (列) ベクトル、とみることもできますし、それぞれ 3 次元、5 次元空間のベクトルと考えても構いません。ちなみに行は $\rightarrow$ の方向、列は $\downarrow$ の方向です。行という漢字の四画目、列という漢字の最後の二画が方向を表している、と覚えるといいかもしれません^{*4}。ちなみに英語で行は Row、列は Column で、R の下を伸ばしたり、L が縦長なのでそれを意識したりして、方向を覚えるといいかもしれません (図 1.1)。

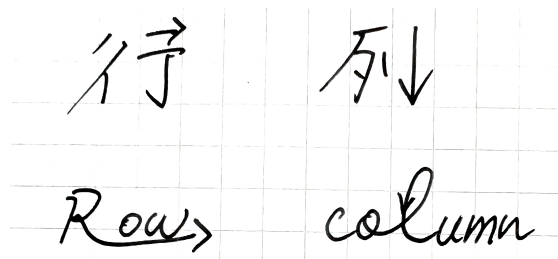


図 1.1 行と列、その覚え方

^{*4} 行の最後の二画、列の最初の二画と逆に覚えられないように!

また、ベクトルやこの後出てくる行列を表すアルファベットは、 a や B のようにただのアルファベットではなく、 $\mathbf{a}$, $\mathbf{B}$ のように太字体で表します。字が太くなっているときはベクトルや行列、すなわち複数の数字から構成されたものだな、と思って見るようにしてください。ちなみに手書きでノートなどに書くときは、どこかの一面を二重線にするなどして表現することが多いです。

閑話休題。このベクトルの中の数字は、とくに関係があるわけではありません。前に入っている数字がえらいとか、横にある方が重要だ、といったことはなく、ただただ数字をまとめて扱っているだけです。本来別々の数字なのです。たとえば身長と体重という2つの変数があったとき、ある人のデータを $\begin{pmatrix} 170 \\ 60 \end{pmatrix}$ と表すことができます。全然別の種類の、別の単位の数字ですね。ただし、一行目の要素に身長、二行目の要素に体重を入れるという規則にしたのであれば、別の人のデータも $\begin{pmatrix} 150 \\ 45 \end{pmatrix}$ のように表さなければなりません。ここで要素の順番を入れ替えることはできないのです。

さて、行数も列数も1であるものつまり行列でない数字は、とくに**スカラー (scalar)** と呼びます。今までは $1 + 2 = 3$ といった計算をしていましたが、この1, 2, 3はすべてスカラーだといえるわけです。「スカラー」という用語は線形代数の専門用語として定着していますが、Scale^{スケール}を扱うもの、という意味があるので「スケーラー」と発音した方がわかりやすいかもしれません。この後、スカラーとベクトル・行列の積の話をするますが、定数倍のように大きさを変えるという操作をします。空間的な表現のイメージとつなげることの利点ですが、そんなところにも現れてくるのです。

行列

行列 (matrix) とは数を長方形に並べたものです。行列として並べられた数を成分といい、成分の横の並びを行、縦の並びを列と呼びます。

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1m} \\ a_{21} & & \ddots & & & a_{2m} \\ \vdots & & & \vdots & & \vdots \\ a_{i1} & \cdots & \cdots & a_{ij} & \cdots & \vdots \\ \vdots & & & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & \cdots & a_{nm} \end{pmatrix}$$

この例は、 n 行 m 列の行列を表しています。この行列の成分を表す文字は、一般に a_{ij} のように、はじめの添え字で行番号、次の添え字で列番号を表します。行列の大きさは行数と列数とによって、 $n \times m$ のように表現します。表計算ソフトなどで見る、シートの中に入っているデータセットを考えると、行が個人・ケース・Observationsを表していて、列が変数を表しているのが一般的ですね。行列を考えると n が数百人ぐらい、 m が数十ぐらいのサイズでイメージすると良いと思います。調査系のデータであれば、 n が3,4桁ぐらい、 m が2桁台なののが一般的ではないかと思います。 $n > m$ の縦に長い長方形を頭に思い描いてみてください。行列では、このサイズ感 (サイズのイメージ) が特に重要です。

■**正方行列** 一方、 n と m が同じ、つまり行数と列数が同じであれば正方形になりますから、これをとくに**正方行列 (Square Matrix)** といいます。正方行列の例を次にあげておきます。

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

■**対称行列** 正方行列の中でも、 i 行 j 列目の値が j 行 i 列目の値と同じである行列 ($a_{ij} = a_{ji}$) のことを**対称行列 (Symmetric Matrix)** といいます。

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 9 \end{pmatrix}$$

この (正方) 対称行列の形は、データ解析の計算途中でよく出てきます。たとえば 3 つの変数 x_1, x_2, x_3 について、その相関係数を考えたいとしましょう。相関係数は 2 つの数字の組み合わせですから、 x_1 と x_2 、 x_1 と x_3 、 x_2 と x_3 について計算でき、それぞれ r_{12}, r_{13}, r_{23} と表したとします。 i と j の相関係数 r_{ij} は、 j と i の相関係数と同じ ($r_{ij} = r_{ji}$) であり、また $r_{jj} = 1.0$ なのは定義から明らかです。これを行列で表すと次のようになります。

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix}$$

$r_{ij} = r_{ji}$ ですから、対称行列になっています。この行列をとくに**相関行列 (Correlation Matrix)** といいます。また相関係数は標準化された共分散でもありました。標準化するまえの相関行列は、**分散共分散行列 (Covariance Matrix)** と言います。その名前の通り、自分自身との共分散が分散になるわけですから、右上から右下にかけての対角線上にある要素 (これをとくに**対角 (diagonal) 要素** といいます) が分散であり、それ以外が共分散になっている行列です。

$$V = \begin{pmatrix} s_1^2 & s_{12} & s_{13} \\ s_{21} & s_2^2 & s_{23} \\ s_{31} & s_{32} & s_3^2 \end{pmatrix}$$

■**対角行列と単位行列** 正方行列の中でもとくに対角要素にのみ値があつて、それ以外はすべて 0 になっている行列のことを**対角行列 (diagonal matrix)** といいます。

$$\begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix} \text{ あるいは } \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}$$

ここで表されているように、対角要素以外は全てゼロですから、ゼロを大きく太字で書いたり、オーを使って O としてここは空っぽだよ、ということを示すことがあります。

また対角行列の中でもとくに、対角項が 1 になっているものは**単位行列 (identity matrix)** と呼びます。単位行列は I とか E で表されます。

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 1 & & O \\ & \ddots & \\ O & & 1 \end{pmatrix}$$

これは後ほど、行列で掛け算をするときに「かけても変わらない行列」として出てきます。この行列は計算過程でよく出てくる行列で、特に式中でサイズを明示しないこともありますが、計算過程を考えるとそのサイズは明らかなので明示しない、ということがよくあります。そのときは「計算できるような適当なサイズ」である、と思ってください。明示するときは、 I_n のように表すことで、サイズ $n \times n$ の単位行列であることを表します。

■**三角行列** さらに、正方行列の特殊なケースで、対角より上、あるいは下の要素がすべて0になっている行列を上三角行列、下三角行列といいます。

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ \mathbf{O} & & a_{nn} \end{pmatrix} \text{ または } \begin{pmatrix} a_{11} & & \mathbf{O} \\ \vdots & \ddots & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

■**ゼロ行列** 全ての要素が0であるような行列は、ゼロ行列 $\mathbf{O}$ といいます。

$$\mathbf{O} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

1.3 行列の四則演算

行列やベクトルが定義できたところで、操作に入っていきます。

1.3.1 行列の加法

■**イコールであるとは** 2つの行列 $\mathbf{A}$ と $\mathbf{B}$ が共に $n \times m$ の大きさに、対応する全ての成分が $a_{ij} = b_{ij}$ であるとき、 $\mathbf{A}$ と $\mathbf{B}$ は等しいといい、 $\mathbf{A} = \mathbf{B}$ で表します。

当たり前のことのようにですが、行列の四則演算は、通常の数字のそれとは異なります。改めて、行列としての加減乗除を定義することになりますので、これまでの数の世界のルールを行列用に定義していく必要があります。その最初が「等しい」の定義なわけです。これで計算結果をイコールで結ぶことができるようになります。

■**行列のの足し算・引き算** まずは行列の足し算(加法)、引き算(減法)から説明します。ベクトルは行列の中でも、行数あるいは列数が1のもので、これで一般的に表現します。これはそれぞれ対応する位置にある成分を加え合わせる(減じる)ことで表されます。

行列 $\mathbf{A}$ の要素を a_{ij} 、行列 $\mathbf{B}$ の要素を b_{ij} とします。このとき行列 $\mathbf{A}, \mathbf{B}$ はいずれも $n \times m$ の同じサイズの行列です。この行列の和 $\mathbf{A} + \mathbf{B} = \mathbf{C}$ は、対応する要素同士の和を要素に持つ行列、すなわち $c_{ij} = a_{ij} + b_{ij}$ になります。

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2m} + b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{nm} + b_{nm} \end{pmatrix}$$

数値例をみておきましょう。

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix}$$

ベクトルでも同様です。

$$\begin{pmatrix} 1 \\ 4 \end{pmatrix} - \begin{pmatrix} 5 \\ 8 \end{pmatrix} = \begin{pmatrix} 1-5 \\ 4-8 \end{pmatrix} = \begin{pmatrix} -4 \\ -4 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 6 \end{pmatrix} = \begin{pmatrix} 2+1 & 3+6 \end{pmatrix} = \begin{pmatrix} 1 & -3 \end{pmatrix}$$

これからわかるように、行列の加法、減法は大きさの等しい行列でないと成り立ちません。サイズが違うものを足そうとすると、演算できない箇所が出てしまうのです。このように行列では、「計算できない」という状態になることがあります。行列のサイズに注意が必要、ということがお分かりいただけるかと思います。

■**行列とスカラーの積** 次に掛け算 (積) の定義にすすみます。

行列 A の λ 倍は、 A の各要素 a_{ij} を λ 倍します^{*5}。

$$\lambda A = A\lambda = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1m} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{n1} & \lambda a_{n2} & \cdots & \lambda a_{nm} \end{pmatrix}$$

数値例でも確認しておきます。

$$2 \times \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 2 \times 1 & 2 \times 2 \\ 2 \times 3 & 2 \times 4 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$$

ベクトルに対しても同様です。

$$2 \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \times 1 \\ 2 \times 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$$

$$3 \times \begin{pmatrix} 2 & 4 \end{pmatrix} = \begin{pmatrix} 3 \times 2 & 3 \times 4 \end{pmatrix} = \begin{pmatrix} 6 & 12 \end{pmatrix}$$

ここまでの話から、交換、結合、分配などの演算法則が成り立つことがわかります。

$$\begin{aligned} A + B &= B + A && \text{交換法則} \\ (A + B) + C &= A + (B + C) && \text{結合法則} \\ (\alpha + \beta)A &= \alpha A + \beta A && \text{分配法則} \\ \lambda(A + B) &= \lambda A + \lambda B && \text{分配法則} \\ \lambda(AB) &= (\lambda A)B = A(\lambda B) && \text{交換法則} \end{aligned}$$

これらは特に証明しませんが、計算のパターンとして頭に留めておいてください。

例題 1

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}, B = \begin{pmatrix} 2 & 1 \\ 4 & 1 \\ 0 & 2 \end{pmatrix} \text{ であるとき, } 2A - 3B \text{ を計算してみましょう。}$$

1.3.2 行列の乗法

続いて積の定義に入ります。ここでは、計算によって行列 (ベクトル) のサイズが変わることになりますので、サイズ感に特に注意を払ってください。

^{*5} 式中にてでくる λ はギリシア文字でラムダといいます。小文字が λ , 大文字では Λ と書きます。

■ベクトルとベクトルの積 行列の積は、前の行列の行成分と、後ろの行列の列成分を順番にかけて足していくことになります。 $AB = C$ とするとき、 A の i 行目と B の j 列目の要素から、 C の $[i, j]$ 要素が算出されるのです。記号で表現すると、次のようになります。

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj} = \sum_{p=1}^k a_{ip}b_{pj}$$

ここで $\sum$ が出てきましたが、これは総和を表す記号です。この記号はギリシア文字の大文字の S で、足し合わせる summation の頭文字をとって $\sum$ 、サムと読みます。この記号には上と下に文字を追加して、次のような意味で使います。

$$\sum_{i=1}^5 x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

ここにあるように、 $\sum$ の下に i という文字があります。この文字 i が 1 から始まって 5 まで変化し総和する、という数学記号です。 $\sum_{i=1}^5 x_i$ の x_i という変数について、 i が 1 から 5 まで変わるわけですから、右辺にあるように添字が変わりながらすべて足し合わせる、という意味の記号です。この $\sum$ も、文脈から考えてどこからどこまで足すのかが明らかな場合、上下の添字は省略されることがあります。今回は 1 から k まで変わりますので、第 4 の記号 p を使って $1, 2, \dots, p, \dots, k$ を表しています。

さてこの計算、要素の位置を想像しながら考えてみましょう。

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ik} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mk} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kj} & \cdots & b_{kn} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1j} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2j} & \cdots & c_{2n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ c_{i1} & c_{i2} & \cdots & c_{ij} & \cdots & c_{in} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mj} & \cdots & c_{mn} \end{pmatrix}$$

積の結果の各要素が、前の行列の行、後ろの行列の列の積和から算出されると考えることになります。

記号ではピンとこないと思いますので、話を簡単にするために、1 列あるいは 1 行だけの行列、すなわち列ベクトル、行ベクトルの積から考えましょう。

$$\begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} d \\ e \\ f \end{pmatrix} = ad + be + cf$$

数値例も確認しておきます。

$$\begin{pmatrix} 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} = 1 \times 3 + 2 \times 4 + 1 \times 2 = 13$$

ここで注意して欲しいのは、両方のベクトルのサイズが同じ ($1 \times n$ ベクトルと、 $n \times 1$ ベクトル、いずれもサイズは n) ということです。サイズが違えば、演算が対応しない要素が出てくるので、計算できなくなるのです。また、ベクトルとベクトルを掛け合わせた結果が 1×1 の行列になりますね*6。

*6 実は 1×1 の行列をスカラーと言っている問題があります。本当は両者は違うものと考えべきなのですが、計算上は違いがないのでスカラーとみなすのが一般的です。

列ベクトルと行ベクトルの積は次のようになります。

$$\mathbf{ab} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} (b_1 \quad b_2 \quad \cdots \quad b_n) = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & & \vdots \\ a_n b_1 & a_n b_2 & \cdots & a_n b_n \end{pmatrix}$$

数値例です。

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} (3 \quad 4 \quad 2) = \begin{pmatrix} 1 \times 3 & 1 \times 4 & 1 \times 2 \\ 2 \times 3 & 2 \times 4 & 2 \times 2 \\ 1 \times 3 & 1 \times 4 & 1 \times 2 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 2 \\ 6 & 8 & 4 \\ 3 & 4 & 2 \end{pmatrix}$$

今度は行列になりました。かける順番が変わるとサイズが変わる (ここでは、上の例では 1×1 のサイズ、下の例では $n \times n$ のサイズ) ことに注意してください。スカラーの計算では順番を入れ替えても、たとえば $2 \times 3 = 3 \times 2$ のように同じ答えになりましたが、行列の場合は必ずしもそうはならないのです ($\mathbf{AB} \neq \mathbf{BA}$)。

積の結果出来上がる行列は、前の行列の行と後ろの行列の列の組み合わせですから、このような結果になるのですね。

例題 2

$\mathbf{a} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 2 & 4 & 1 \end{pmatrix}$ であるとき、 $\mathbf{ab}$ および $\mathbf{ba}$ を計算してみましょう。

■**行列とベクトルの積** 行列とベクトルの積や、行列と行列の積はこの応用になってきます。

まず行列に列ベクトルを右からかける例を見てみましょう。結果は列ベクトルになります。

$$\mathbf{Ab} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^m a_{1j} b_j \\ \sum_{j=1}^m a_{2j} b_j \\ \vdots \\ \sum_{j=1}^m a_{nj} b_j \end{pmatrix}$$

ここでも掛け算なのに足し算のプロセスが入ってきています。注意深く記号を読んでみてください。数値例でも確認しておきます。

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \times 2 + 2 \times 1 \\ 3 \times 2 + 4 \times 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 10 \end{pmatrix}$$

今度は行列に行ベクトルを左からかけましょう。結果は行ベクトルになります。

$$\mathbf{cA} = (c_1 \quad c_2 \quad \cdots \quad c_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} = \left(\sum_{j=1}^n a_{j1} c_j \quad \sum_{j=1}^n a_{j2} c_j \quad \cdots \quad \sum_{j=1}^n a_{jm} c_j \right)$$

$$(1 \quad 3) \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} = (1 \times 1 + 3 \times 2 \quad 1 \times 0 + 3 \times 3) = (7 \quad 9)$$

例題 3

$\mathbf{a} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 2 & 4 & 1 \end{pmatrix}$, $\mathbf{C} = \begin{pmatrix} 2 & 0 & 1 \\ 3 & 1 & 1 \\ 2 & 4 & 2 \end{pmatrix}$ であるとき, $\mathbf{C}\mathbf{a}$ および $\mathbf{b}\mathbf{C}$, そして $\mathbf{b}\mathbf{C}\mathbf{a}$ を計算してみましょう。

■**行列と行列の積** 最後に行列と行列の積にもどって考えてみましょう。行列 $\mathbf{A}$ と $\mathbf{B}$ の積が成立するのは, 前者の列数と後者の行数とが等しいときに限られます。行列 $\mathbf{A}$ のサイズが $m \times k$, 行列 $\mathbf{B}$ のサイズが $k \times n$ とすると, その積は $n \times m$ の行列になります。計算手続きは, 次のようになります,

$$\mathbf{AB} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mk} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kn} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^k a_{1j}b_{j1} & \cdots & \sum_{j=1}^k a_{1j}b_{jn} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^k a_{mj}b_{j1} & \cdots & \sum_{j=1}^k a_{mj}b_{jn} \end{pmatrix}$$

どうにもこれはややこしいかもしれません。足し算や掛け算が入り乱れるし, 計算途中でどの要素を計算しているかわからなくなるからです。ベクトルと行列の積の時のように, 前の行列の要素は左に進み, 後ろの行列の要素は縦に進みますから, 左手と右手で違う図形を描く認知課題のように, そもそも混乱しやすい作業なのです。

しかし2つほど注意をしておくと, 間違いにくくなります。1つは積によって得られる**結果の行列サイズを意識すること**です。先ほど, 前の行列の列数と, 後ろの行列の行数が同じでないと計算ができないといいました。つまり, $n \times k$ 行列と $k \times m$ 行列でないと計算できない (k が同じ) ということです。また, 結果は $n \times m$ 行列になります。前の行列の行数, 後ろの行列の列数が結果のサイズです。ここに注目しておくと, 計算を始める前に, 計算が可能かどうかと結果の行列サイズは想像がつくのです (図 1.2)。計算の前に行列のサイズを見積もっておくと, 計算間違いの可能性が少し減るかと思います。

また, 実際に計算する際は, **前の行列に横の, 後ろの行列に縦の補助線を入れる**とわかりやすいかもしれません。こうすることで, 間違えて計算を進めることがないようになるからです。

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \left(\begin{array}{c|c|c} 0 & 1 & 1 \\ 1 & 0 & 1 \end{array} \right) = \begin{pmatrix} 1 \times 0 + 2 \times 1 & 1 \times 1 + 2 \times 0 & 1 \times 1 + 2 \times 1 \\ 3 \times 0 + 4 \times 1 & 3 \times 1 + 4 \times 0 & 3 \times 1 + 4 \times 1 \\ 5 \times 0 + 6 \times 1 & 5 \times 1 + 6 \times 0 & 5 \times 1 + 6 \times 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 3 & 7 \\ 6 & 5 & 11 \end{pmatrix}$$

これら行列の積については, 次のような計算がなりたちます。 c を実数, $\mathbf{A}, \mathbf{B}, \mathbf{C}$ を行列とすると,

$$\begin{aligned} c\mathbf{A} &= \mathbf{A}c \\ (c\mathbf{A})\mathbf{B} &= c(\mathbf{AB}) \\ (\mathbf{AB})\mathbf{C} &= \mathbf{A}(\mathbf{BC}) && \text{結合法則} \\ (\mathbf{A} + \mathbf{B})\mathbf{C} &= \mathbf{AC} + \mathbf{BC} && \text{分配法則} \\ \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} && \text{分配法則} \\ \text{一般に交換法則は成立しない} \end{aligned}$$

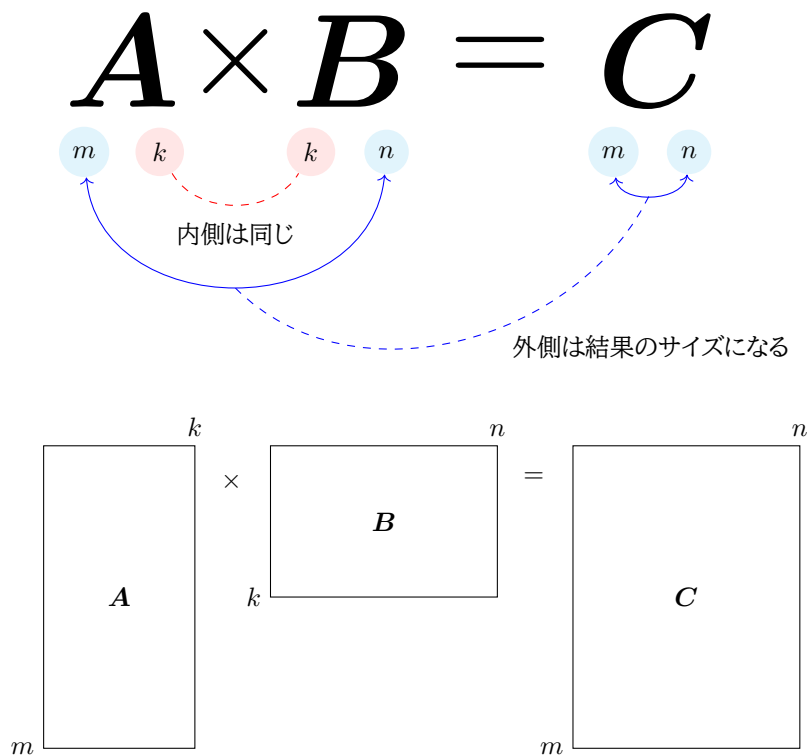


図 1.2 行列のサイズを把握する

また、 $AA = A^2$ と書きます。3 乗は $AAA = A^3$ です。同様に累乗は A^n のように書くことができます。

こうした性質は、スカラーの計算と同じようなところもあり、一部違うところもありますので、注意しておきましょう (特に交換法則)。これらの行列の計算はスカラーの計算ほど身体化されていないことと思います。慣れて自動的にできるようになるには、練習を繰り返すしかありません。少し練習してみましょう。

例題 4

次の計算をそれぞれやってみましょう。

(1)

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(2)

$$\begin{pmatrix} -3 & 2 \\ 5 & 5 \end{pmatrix} \begin{pmatrix} 1 & 3 & 4 \\ 2 & 0 & 1 \end{pmatrix}$$

(3)

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \\ 0 & -2 & 3 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 3 & 0 \\ -1 & -1 \end{pmatrix}$$

(4)

$$\begin{pmatrix} 1 & 5 & -1 \\ -2 & 3 & -1 \\ 3 & 0 & 3 \end{pmatrix} \begin{pmatrix} -2 & 1 & 0 \\ 3 & 0 & 1 \\ 0 & 3 & 4 \end{pmatrix}$$

1.4 その他の行列の操作

最後に、その他の行列操作について幾つかご紹介しておきます。

■**転置 (transpose)** と呼ばれる操作を説明します。これは計算の便宜上、よく使われる行列操作のひとつです。

大きさ $n \times m$ の行列 A における i 行 j 列成分を j 行 i 列成分とする $m \times n$ 行列のことを、元の行列 A の転置とよび、 A' や A^T と表します。行列を転ばせたようなイメージです。

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \text{ のとき, } A' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}$$

ベクトルも転置でき、行ベクトルを転置すると列ベクトルに、列ベクトルを転置すると行ベクトルになります。

$$\mathbf{a} = (a_1 \quad a_2 \quad \cdots \quad a_n) \text{ のとき, } \mathbf{a}' = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

また、転置には以下のような性質があります。

1. $(A')' = A$
2. $(A + B)' = A' + B'$
3. $(AB)' = B' A'$
4. $(cA)' = cA'$

最初のは「転置の転置は元通り」ということです。これは想像に難くないですね。次に転置の分配とでも言いましょうか、足したものを転置したのはそれぞれを転置して足してもよい、ということです。第3のものは少しわかりにくいかもしれません。掛けたものを転置するときは、転置したものの順番を入れ替えてかけたものと同じ、ということです。確認しておきましょう。 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $B = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$ として,

$$(AB)' = \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} \right)' = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + gh \end{pmatrix}' = \begin{pmatrix} ae + bg & ce + dg \\ af + bh & cf + gh \end{pmatrix}$$

ここで $A'B'$ としてしまうと,

$$A'B' = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} e & g \\ f & h \end{pmatrix} = \begin{pmatrix} ae + cf & ag + ch \\ be + df & bg + dh \end{pmatrix}$$

となり、左辺と等しくありません ($(AB)' \neq A'B'$)。前後を入れ替えると合致します。

$$B'A' = \begin{pmatrix} e & g \\ f & h \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} ae + bg & ce + gd \\ af + bh & cf + dh \end{pmatrix}$$

ここだけ注意しておいていただければ、後はスムーズに計算できるのではないのでしょうか。

■トレース 行列のトレース (trace) とは、対角成分の総和のことを指し、 $tr(A)$ のように表記します^{*7}。

$$tr(A) = a_{11} + a_{22} + \cdots + a_{mm} = \sum a_{ii}$$

対角項の特徴ですので、行列の積の計算をしても変わらないという特徴があります。すなわち、 $n \times m$ サイズの行列 A と、 $m \times n$ サイズの行列 B があったとして、

$$tr(AB) = tr(BA)$$

になります。

ところでこの操作に何の意味があるのか、と思う方もいるかもしれませんので、先取りして少し補足します。トレースは正方行列に対して定義される操作ですが、その正方行列が分散共分散行列だったばあい、トレースは分散の総和を計算することになります。分散はその項目から取り出せる意味の大きさを表していますから、トレースはデータ全体からどういう情報が得られるかという値という点で、重要なのです。詳しくはセクション 7.1.2(Pp.104) のあたりで出てきますので、楽しみに。

■分割 行列が大きくなった場合、複数に分割して表現するとわかりやすいことがあります。これを行列の分割とか、分割されたものを小行列と呼ぶことがあります。表記法は次のようなものが一般的です。

$$A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix} = \left[\begin{array}{c|c} P & Q \\ \hline R & S \end{array} \right]$$

例えば 2 つの分割行列,

$$A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}, B = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

があったとき,

$$A + B = \begin{bmatrix} P + E & Q + F \\ R + G & S + H \end{bmatrix}$$

^{*7} 日本語では^{せき}跡といいます。

ですし,

$$\lambda \mathbf{A} = \begin{bmatrix} \lambda P & \lambda Q \\ \lambda R & \lambda S \end{bmatrix}$$

となります。

行列をベクトルに分割することもある、例えば行列 $\mathbf{A} = [a_{ij}]$ を行ごとに,

$$\begin{aligned} \mathbf{a}^1 &= [a_{11}, a_{12}, \dots, a_{1m}] \\ \mathbf{a}^2 &= [a_{21}, a_{22}, \dots, a_{2m}] \\ &\vdots \\ \mathbf{a}^n &= [a_{n1}, a_{n2}, \dots, a_{nm}] \end{aligned}$$

として,

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}^1 \\ \mathbf{a}^2 \\ \vdots \\ \mathbf{a}^n \end{bmatrix}$$

とすることもありますし、列ごとに

$$\mathbf{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix}, \dots, \mathbf{a}_m = \begin{bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{bmatrix},$$

として,

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m]$$

とすることもあります。

特に単位行列を列ベクトルに分割すると,

$$\mathbf{I} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n]$$

となり、この時それぞれの列ベクトルは

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \mathbf{e}_m = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

となります。このようなベクトルは**基本ベクトル**とも呼ばれ、行列を空間で考えるときの基本となる座標ベクトルだと考えられます。

また、ベクトルへの分割の場合、次のような計算が成り立ちます。

列分割と列ベクトルの積

$$\mathbf{A}\mathbf{x} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n$$

行列と列分割の積

$$\mathbf{A}\mathbf{B} = \mathbf{A}[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m] = [\mathbf{A}\mathbf{b}_1, \mathbf{A}\mathbf{b}_2, \dots, \mathbf{A}\mathbf{b}_m]$$

行列と行分割の積

$$AB = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^n \end{bmatrix} B = \begin{bmatrix} a^1 B \\ a^2 B \\ \vdots \\ a^n B \end{bmatrix}$$

行分割と列分割の積

$$AB = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^n \end{bmatrix} [b_1, b_2, \dots, b_m] = \begin{bmatrix} a^1 b_1 & a^1 b_2 & \dots & a^1 b_m \\ a^2 b_1 & a^2 b_2 & \dots & a^2 b_m \\ \vdots & \vdots & & \vdots \\ a^n b_1 & a^n b_2 & & a^n b_m \end{bmatrix}$$

この章では、行列の基本的な定義、名称、演算、操作について説明しました。次回はこれを使ってデータをどのように表現するかをみていきます。

1.5 練習問題

行列 $A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 3 & 1 & 5 \\ 2 & 4 & 6 \end{pmatrix}$, $C = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$, 列ベクトル $x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, 行ベクトル $y =$

$(2 \ 8)$ とするとき、次の計算をなさい。なお、計算が定義されていないものについては「計算できない」と回答しなさい。

1. $A + B$
2. $A - C$
3. AB
4. AC
5. A^2
6. $tr C$
7. $B'B$
8. BB'
9. BI_3
10. $B'A$
11. Ay'
12. AI_2C
13. ACI_2
14. xy
15. xB
16. $x'B'$
17. yx
18. yBx

第 2 章

行列とデータ

線形代数は数字をセットで扱うための表現方法、計算方法ですから、多変量データを分析しようという時には必須の技術になります。今回は線形代数の導入として、行列の計算方法を解説してきましたが、今回はこの計算方法を使って具体的にデータをどのように表現し、どのように計算するのかを見ていくことになります。その前に、基本的な統計量の数式などを確認しておきましょう。

2.1 記述統計の復習

記述統計では、平均、分散、標準偏差、相関係数などをつかってデータの特徴を記述するのです。行列で表現をする前に、基本的な記号や表現について確認しておきます。

総和記号 $\sum$ を使って、変数 x_{ij} の平均値 $\bar{x}_j$ は次のように表現することができます (i はデータ点、個人を表します)。

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

これを使って、各データ x_{ij} が平均からどれくらい離れているかを、 $(x_{ij} - \bar{x}_j)$ であらわし、**平均偏差 (mean deviation)** と言います。平均偏差の二乗の平均のことを**分散 (variance)** というのはでした。

$$s_j^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

ここで二乗してあるのは、そうしておかないと平均偏差の総和はゼロになってしまうからです^{*1}。プラスとマイナスが打ち消し合うからゼロになってしまうわけで、符号の影響をなくすべく二乗しているのだと思ってください。符号をなくすのには絶対値でもいいのですが、絶対値は数式で書けないから使いにくいのです^{*2}。

^{*1} 検算してみよう!

^{*2} 絶対値は $|x|$ と書くじゃないか、と思った人もいるかもしれませんが。おいしい、もう一步。この記号の中身、計算式を考えてみてください。 $0 \leq x$ のとき、 $|x| = x$ です。 $x < 0$ のとき、 $|x| = -x$ です。これ、「...の時」という条件で分岐していますので、1つの式で表せないあたりが使いにくいポイントです。

この分散の式は展開して,

$$\begin{aligned}
 s_j^2 &= \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \\
 &= \frac{1}{n} (x_{ij}^2 - 2x_{ij}\bar{x}_j + \bar{x}_j^2) && \text{平方を展開} \\
 &= \frac{1}{n} \sum x_{ij}^2 - \frac{2}{n} \sum x_{ij}\bar{x}_j + \frac{1}{n} \sum \bar{x}_j^2 && \sum \text{を分配} \\
 &= \frac{1}{n} \sum x_{ij}^2 - 2\bar{x}_j \frac{1}{n} \sum x_{ij} + \bar{x}_j^2 && \text{定数をはり出す} \\
 &= \frac{1}{n} \sum x_{ij}^2 - 2\bar{x}_j\bar{x}_j + \bar{x}_j^2 && \text{平均を計算} \\
 &= \frac{1}{n} \sum x_{ij}^2 - 2\bar{x}_j^2 + \bar{x}_j^2 \\
 &= \frac{1}{n} \sum x_{ij}^2 - \bar{x}_j^2
 \end{aligned}$$

として使われることがあります^{*3}。分散がゼロのデータからは情報が得られませんから、分散とはその変数が持つ情報の大きさを表していると言っているかもしれません。

標準偏差 (Standard Deviation) は分散の正の平方根です。分散は元の単位を 2 乗してしまっているの
で、元のスケールに合わせるためには平方根を取る必要があります。

$$s_j = \sqrt{s_j^2} = \sqrt{\frac{1}{n} \sum (x_{ij} - \bar{x}_j)^2}$$

標準偏差を使ってデータを標準化するというのは、平均偏差を標準偏差で割ることでした。このようにすることで、データ全体の中での相対的な位置付けがわかるようになります。標準化したスコアは z_i で表現され、

$$z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{s_j}$$

と表されます。標準化されたスコア z_{ij} はその平均がゼロ ($\bar{z}_j = 0$) で標準偏差が 1 に標準化されています。

さてここまでは 1 変数についての記述統計でしたが、複数の変数の間の関係を表すには**共分散 (covariance)** を用います。変数 j と k の共分散は次の式で表されます。

$$s_{jk} = \frac{1}{n} \sum (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

これは分散が変数 j と j の共分散、すなわち、

$$s_{jj} = s_j^2 = \frac{1}{n} \sum (x_{ij} - \bar{x}_j)^2 = \frac{1}{n} \sum (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j)$$

であることを思い出して、2 変数バージョンにしたものと理解すれば分かりやすいでしょう。共分散もまた、それぞれの変数の単位を残していますので、標準化することで変数関係と変数関係を相対的に比較することができるようになります。それがピアソンの**相関係数 (Correlation Coefficients)** です。

^{*3} 三段目, $\frac{1}{n} \sum c = \frac{1}{n} nc = c$ がミソ。

$$r_{jk} = \frac{1}{n} \sum (z_{ij} - \bar{z}_j)(z_{ik} - \bar{z}_k) = \frac{1}{n} \sum z_{ij} z_{ik}$$

以後これらの数字, 記号を使って話をしていきます。ここまでの話と違うところは, 変数が 1 つや 2 つではなくたくさんあるので, 一気に表現, 計算する方法が欲しいという点です。まとめて表現できるのが行列の強みなのです。

2.2 行列を使った計算

さて心理統計においてこうした行列表現をするときのねらいは, あくまでも調査研究など多変量データを扱う場面での利用です。変数の数が m 個あるときに, $r_{12}, r_{13}, \dots, r_{1m}, r_{23}, r_{24}, \dots$ と逐一書いていられませんか, 行列として扱うと表現が大変便利なのです。たとえば質問項目が m 個あって, 調査対象者 n 人から回答を得たとすると, データは次のように表現できます。

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

また各反応の平均点は以下のように表現されます。まず, 要素がすべて 1 からなるベクトルを次のように表します。

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

わかりにくいかもしれませんが, この $\mathbf{1}$ は太字でベクトルを表しており, スカラーの 1 とは違うことに注意してください。

さて, 各項目の和はベクトルの掛け算の定義によって次のように表現できます。

$$\mathbf{X}'\mathbf{1} = \begin{pmatrix} \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i2} \\ \vdots \\ \sum_{i=1}^n x_{im} \end{pmatrix}$$

これを使って平均値 (列) ベクトル $\mathbf{m}$ を次のように表すことができます。

$$\mathbf{m} = \frac{1}{n} \mathbf{X}'\mathbf{1} = \begin{pmatrix} 1/n \sum_{i=1}^n x_{i1} \\ 1/n \sum_{i=1}^n x_{i2} \\ \vdots \\ 1/n \sum_{i=1}^n x_{im} \end{pmatrix} = \begin{pmatrix} \bar{x}_{.1} \\ \bar{x}_{.2} \\ \vdots \\ \bar{x}_{.m} \end{pmatrix}$$

ここで $\bar{x}_1$ は 1 つめの添字 i を足し合わせて割ることなくしていますから、 $\bar{x}_1$ のように省略して書くことがあります。このときの 1 は「第一番目の変数」という意味であり、個人の情報がなくなっている変数を意味していることに注意してください。

具体的な数値例で見ておきましょう。次のようなデータセットがあったとします^{*4}。

表 2.1 データセット例

Name	身長	体重	年齢
佐藤	150	50	42
鈴木	160	55	38
高橋	145	40	22
田中	165	65	50
伊藤	175	85	48

数値データだけ取り出すと次のように行列で表現できます。

$$\mathbf{X} = \begin{pmatrix} 150 & 50 & 42 \\ 160 & 55 & 38 \\ 145 & 40 & 22 \\ 165 & 65 & 50 \\ 175 & 85 & 48 \end{pmatrix}$$

データセットは 5 人分ですので、 $\mathbf{1}' = (1 \ 1 \ 1 \ 1 \ 1)$ として、次のような計算ができます。

$$\mathbf{X}'\mathbf{1} = \begin{pmatrix} 150 & 160 & 145 & 165 & 175 \\ 50 & 55 & 40 & 65 & 85 \\ 42 & 38 & 22 & 50 & 48 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 795 \\ 295 \\ 200 \end{pmatrix}$$

例題 5

平均ベクトル $\mathbf{m}$ を求めましょう。ここで $n = 5$ です。

^{*4} 名前は別に ID 番号でもよかったのですが、具体的なイメージがしやすいかなと思って日本の苗字ランキングから上位 5 つの苗字を取り出してみました。

さて、平均からの偏差を要素に持つ行列 V を考えたとします。

$$\begin{aligned}
 V &= X - \mathbf{1}m' \\
 &= X - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} (\bar{x}_{.1} \quad \bar{x}_{.2} \quad \cdots \quad \bar{x}_{.m}) \\
 &= X - \begin{pmatrix} \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \end{pmatrix} \\
 &= \begin{pmatrix} x_{11} - \bar{x}_{.1} & x_{12} - \bar{x}_{.2} & \cdots & x_{1m} - \bar{x}_{.m} \\ x_{21} - \bar{x}_{.1} & x_{22} - \bar{x}_{.2} & \cdots & x_{2m} - \bar{x}_{.m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_{.1} & x_{n2} - \bar{x}_{.2} & \cdots & x_{nm} - \bar{x}_{.m} \end{pmatrix}
 \end{aligned}$$

この行列 V のサイズは $n \times m$ であることに注意してください。

これはまた、次のように表すこともできます。

$$V = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)X$$

ここで I は適切なサイズの単位行列です^{*5}。

^{*5} 適切なサイズってなんだよ、と思いますよね。これは計算に合うようなサイズ、という意味です。具体的に考えてみますと、 I の後ろは $\mathbf{1}\mathbf{1}'$ です。 $\mathbf{1}$ は $n \times 1$ の列ベクトルで、転置したものと掛け合わせますから、 $\mathbf{1}\mathbf{1}'$ のサイズは $n \times n$ です。行列の引き算は同じサイズでないと成立しませんから、ここでの I も $n \times n$ でなければなりません。カッコの中身が $n \times n$ で、それにサイズ $n \times m$ である X をかけますから、計算結果や右辺のサイズは $n \times m$ になります。

この計算過程を具体的に (要素を参照しながら^{*6}) 書き下してみましょう。

$$\begin{aligned}
 \mathbf{V} &= \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) \mathbf{X} \\
 &= \left(\begin{pmatrix} 1 & & \mathbf{O} \\ & 1 & \\ & & \ddots \\ \mathbf{O} & & & 1 \end{pmatrix} - \frac{1}{n} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdots \end{pmatrix} \right) \mathbf{X} \\
 &= \left(\begin{pmatrix} 1 & & \mathbf{O} \\ & 1 & \\ & & \ddots \\ \mathbf{O} & & & 1 \end{pmatrix} - \begin{pmatrix} 1/n & 1/n & \cdots & 1/n \\ 1/n & 1/n & & \\ \vdots & \vdots & \ddots & \vdots \\ 1/n & & & 1/n \end{pmatrix} \right) \mathbf{X} \\
 &= \left(\mathbf{X} - \begin{pmatrix} 1/n & 1/n & \cdots & 1/n \\ 1/n & 1/n & & \\ \vdots & \vdots & \ddots & \vdots \\ 1/n & & & 1/n \end{pmatrix} \mathbf{X} \right) \\
 &= \left(\mathbf{X} - \begin{pmatrix} 1/n(x_{11} + x_{21} + \cdots + x_{n1}) & 1/n(x_{12} + x_{22} + \cdots + x_{n2}) & \cdots & 1/n(x_{1m} + x_{2m} + \cdots + x_{nm}) \\ 1/n(x_{11} + x_{21} + \cdots + x_{n1}) & 1/n(x_{12} + x_{22} + \cdots + x_{n2}) & \cdots & 1/n(x_{1m} + x_{2m} + \cdots + x_{nm}) \\ \vdots & \vdots & \ddots & \vdots \\ 1/n(x_{11} + x_{21} + \cdots + x_{n1}) & 1/n(x_{12} + x_{22} + \cdots + x_{n2}) & \cdots & 1/n(x_{1m} + x_{2m} + \cdots + x_{nm}) \end{pmatrix} \right) \\
 &= \left(\mathbf{X} - \begin{pmatrix} \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \\ \vdots & \vdots & & \vdots \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \end{pmatrix} \right) \\
 &= \begin{pmatrix} x_{11} - \bar{x}_{.1} & x_{12} - \bar{x}_{.2} & \cdots & x_{1m} - \bar{x}_{.m} \\ x_{21} - \bar{x}_{.1} & x_{22} - \bar{x}_{.2} & \cdots & x_{2m} - \bar{x}_{.m} \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_{.1} & x_{n2} - \bar{x}_{.2} & \cdots & x_{nm} - \bar{x}_{.m} \end{pmatrix}
 \end{aligned}$$

例題 6

例題の数字を使って, 平均偏差行列 $\mathbf{V}$ を求めましょう。

これを使って $\mathbf{S} = \frac{1}{n} \mathbf{V}'\mathbf{V}$ を計算してみましょう。

$$\begin{aligned}
 \mathbf{S} &= \frac{1}{n} \mathbf{V}'\mathbf{V} \\
 &= \frac{1}{n} \begin{pmatrix} x_{11} - \bar{x}_{.1} & x_{21} - \bar{x}_{.1} & \cdots & x_{n1} - \bar{x}_{.1} \\ x_{12} - \bar{x}_{.2} & x_{22} - \bar{x}_{.2} & \cdots & x_{n2} - \bar{x}_{.2} \\ \vdots & \vdots & & \vdots \\ x_{1m} - \bar{x}_{.m} & x_{2m} - \bar{x}_{.m} & \cdots & x_{nm} - \bar{x}_{.m} \end{pmatrix} \begin{pmatrix} x_{11} - \bar{x}_{.1} & x_{12} - \bar{x}_{.2} & \cdots & x_{1m} - \bar{x}_{.m} \\ x_{21} - \bar{x}_{.1} & x_{22} - \bar{x}_{.2} & \cdots & x_{2m} - \bar{x}_{.m} \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_{.1} & x_{n2} - \bar{x}_{.2} & \cdots & x_{nm} - \bar{x}_{.m} \end{pmatrix}
 \end{aligned}$$

^{*6} エレメントワイズ element-wise の表現, と言ったりします。

ここで、1 行 1 列目の要素 s_{11} は次のようになります。

$$\begin{aligned}
 s_{11} &= \frac{1}{n} ((x_{11} - \bar{x}_{.1})(x_{11} - \bar{x}_{.1}) + (x_{21} - \bar{x}_{.1})(x_{21} - \bar{x}_{.1}) + \cdots + (x_{n1} - \bar{x}_{.1})(x_{n1} - \bar{x}_{.1})) \\
 &= \frac{1}{n} ((x_{11} - \bar{x}_{.1})^2 + (x_{21} - \bar{x}_{.1})^2 + \cdots + (x_{n1} - \bar{x}_{.1})^2) \\
 &= \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_{.1})^2 \\
 &= s_1^2
 \end{aligned}$$

一般に、この行列 S の対角項は、

$$\begin{aligned}
 s_{jj} &= \frac{1}{n} (x_{1j} - \bar{x}_{.j})^2 + (x_{2j} - \bar{x}_{.j})^2 + \cdots + (x_{nj} - \bar{x}_{.j})^2 \\
 &= \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_{.j})^2 \\
 &= s_j^2
 \end{aligned}$$

となります。つまり、項目 j の分散です。

今度は、1 行 2 列目の要素を計算してみましょう。

$$\begin{aligned}
 s_{12} &= \frac{1}{n} (x_{11} - \bar{x}_{.1})(x_{12} - \bar{x}_{.2}) + (x_{21} - \bar{x}_{.1})(x_{22} - \bar{x}_{.2}) + \cdots + (x_{n1} - \bar{x}_{.1})(x_{n2} - \bar{x}_{.2}) \\
 &= \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_{.1})(x_{i2} - \bar{x}_{.2}) \\
 &= s_{12}
 \end{aligned}$$

これは一般に、

$$\begin{aligned}
 s_{jk} &= \frac{1}{n} (x_{1j} - \bar{x}_{.j})(x_{1k} - \bar{x}_{.k}) + (x_{2j} - \bar{x}_{.j})(x_{2k} - \bar{x}_{.k}) + \cdots + (x_{nj} - \bar{x}_{.j})(x_{nk} - \bar{x}_{.k}) \\
 &= \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_{.j})(x_{ik} - \bar{x}_{.k}) \\
 &= s_{jk}
 \end{aligned}$$

と書けます。すなわち j と k の共分散です。まとめると、

$$S = \frac{1}{n} V'V = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1m} \\ s_{21} & s_2^2 & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_m^2 \end{pmatrix}$$

となり、対角項に分散が、非対角項に共分散が入った行列、すなわち**分散共分散行列**を表したことになります。

ここでもサイズに注目してください。 $V'V$ は、サイズで言うと $m \times n$ と $n \times m$ の積ですから、 $m \times m$ になります。この行列は正方対称行列です。

例題 7

例題の数字を使って、分散共分散行列 S を求めましょう。

また、対角項に各変数の標準偏差 s_j の逆数が入った行列 Q を以下のように定めるとしましょう。次のような行列です。

$$Q = \begin{pmatrix} 1/s_1 & 0 & \cdots & 0 \\ 0 & 1/s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/s_m \end{pmatrix}$$

そうすると、これをつかって $Z = VQ$ を考えてみましょう。

$$\begin{aligned} Z &= VQ \\ &= \begin{pmatrix} x_{11} - \bar{x}_{.1} & x_{12} - \bar{x}_{.2} & \cdots & x_{1m} - \bar{x}_{.m} \\ x_{21} - \bar{x}_{.1} & x_{22} - \bar{x}_{.2} & \cdots & x_{2m} - \bar{x}_{.m} \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_{.1} & x_{n2} - \bar{x}_{.2} & \cdots & x_{nm} - \bar{x}_{.m} \end{pmatrix} \begin{pmatrix} 1/s_1 & 0 & \cdots & 0 \\ 0 & 1/s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/s_m \end{pmatrix} \end{aligned}$$

この行列の1行1列目は

$$z_{11} = \frac{x_{11} - \bar{x}_{.1}}{s_1} + 0 + 0 + \cdots + 0$$

となりますし、1行2列目は

$$z_{12} = 0 + \frac{x_{12} - \bar{x}_{.2}}{s_2} + 0 + \cdots + 0$$

のようになります。一般に、 z_{ij} は

$$z_{ij} = \frac{x_{ij} - \bar{x}_{.j}}{s_j}$$

となり、標準化スコア行列になります。行列のサイズは $n \times m$ 行列と $m \times m$ 行列の積ですから、 $n \times m$ になり、当然ですがデータ行列と同じサイズであることがわかります。

例題 8

例題の数字を使って、標準化スコア行列 Z を求めましょう。

さらに、この Z を使って $R = \frac{1}{n} Z'Z$ とすると、

$$\begin{aligned} R &= \frac{1}{n} Z'Z \\ &= \frac{1}{n} \begin{pmatrix} z_{11} & z_{21} & \cdots & z_{n1} \\ z_{12} & z_{22} & \cdots & z_{n2} \\ \vdots & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{pmatrix} \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{pmatrix} \end{aligned}$$

これは非対角要素 r_{jk} を見ると明らかなように、

$$r_{jk} = \frac{1}{n} (z_{1j}z_{1k}) + (z_{2j}z_{2k}) + \cdots + (z_{nj}z_{nk}) = \frac{1}{n} \sum z_{ij}z_{ik}$$

ですから、相関係数が入った正方行列です。すなわち**相関行列**で、その中身を見ると

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{pmatrix}$$

となります。これは標準化された分散共分散行列と考えることもできます。

例題 9

例題の数字を使って、相関行列 $\mathbf{R}$ を求めましょう。

データのサイズにかかわらず、一般的にこのように表現できるのはとてもわかりやすいですね。

2.3 線形モデルの行列表現

ベクトルや行列の記法をつかうと、**回帰分析**や**重回帰分析**の式もとてもシンプルに表現できます。

回帰分析は従属変数 y_i を独立変数 x_i の線形関数で表現するのです。予測値を $\hat{y}_i$ 、係数を β_0, β_1 、残差を e_i とすると、次のように書けます^{*7}。

$$y_i = \hat{y}_i + e_i = \beta_0 + \beta_1 x_i + e_i$$

さてこの式中の x_i や y_i は観測された個々のデータですので、 $\mathbf{x}' = (x_1, x_2, \dots, x_n)$ 、 $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ というベクトルだと考えることができます。ですから、ベクトルの表記を使えば次のように書けます。

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x} + \mathbf{e}$$

これは、要素を表現しながら書くと次のようになります。

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

列ベクトル $\mathbf{x}$ に β_1 をかけて得られるのは同じサイズの列ベクトル、列ベクトル同士は足しても同じサイズの列ベクトルですから、左辺と右辺はどちらも列ベクトルで、対応関係が取れていることになります。

ところで、 β_0 はベクトルなのに β_1 はベクトルでないのは美しさに欠けますね。ということで、ここで少し表現の工夫をします。説明変数 x のベクトルの左に数字の 1 だけが入った列を作ります。また、係数もまとめてベクトル $\boldsymbol{\beta}$ を次のように用意します。

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

^{*7} 一次関数は最初 $y = ax + b$ で習うので、ここでも $y_i = ax_i + b + e_i$ としてもよかったのですが、カッコつけて β_0, β_1 としてしまいました。ギリシア文字で書くのは標本の数字ではなくて母数であることの表現で、推測統計学での表記の習わしです。回帰分析も標本の線形関係ではなく母集団での線形関係が知りたいはず、というわけです。最小二乗法を使って標本の線形関係を求めるのであれば、 $y_i = ax_i + b + e_i$ でよいのです。ただ、この後に述べるように、重回帰分析になると説明変数が $x_1, x_2, \dots$ と増えていくので、説明変数に合わせて係数も $b_1, b_2, \dots$ とした方がわかりやすいですから、 $y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + e_i$ と書くことになります。切片は説明変数が増えても 1 つの回帰式に 1 つだけですので、 b_0 とゼロをつかいます。記号が変わっただけですが、 a が β に、 b が β_0 になっても急に知らない話になった、と思わないでくださいね。

このようにすると、回帰分析の式は

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

と表すことができます。とても簡単な表現になりました (試しに各行を行列の計算式に則って計算してみてください。うまく表現できていることがわかると思います)。

さらにこの表現はありがたいことに、複数の説明変数がある重回帰分析の時でも同じ形で表すことができます。重回帰分析は、これまでの書き方ですと $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_n X_{ni} + e_i$ というようにしていました。これも先ほどと同じように、説明変数行列の左端に 1 を入れたベクトルを追加し、回帰係数 $\boldsymbol{\beta}$ もセットにして、次のように表現します。

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

とすると、重回帰分析の式を行列表現すると、次のように簡単な表現に変わります。

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

なんと、説明変数が m 個に増えたのに、式の形は (単) 回帰分析のそれと同じではありませんか！

このように、行列表現をすると多くの変数を一気に扱い、表現できるのです。このため、多変量解析ではベクトルの表記が基本になります。サイズを気にせず一般的に表現できるからです。

実際にこれらの式を読む時は、行列のサイズをイメージしながら読むと良いでしょう。たとえば左辺の $\mathbf{y}$ はサイズ n のベクトルなのですから、右辺の $\mathbf{X}\boldsymbol{\beta}$ もサイズ n の縦ベクトルになるはずなのです。実際、 $\mathbf{X}$ は $n \times (m+1)$ の行列で、 $\boldsymbol{\beta}$ は $(m+1) \times 1$ のベクトルですから、計算結果は $n \times 1$ 、つまりサイズ n の縦ベクトルになっています。

2.4 デザイン行列

ところで、心理統計といえば t 検定や分散分析など、平均値の差を見る帰無仮説検定が主流で、回帰分析や因子分析などは多変量解析の領域であって推測統計とは別物だ、と思っている人もいかもしれませんが^{*8}。平均の差を見るために標本平均の計算だけではだめで、母集団に話を一般化するために推測統計の知見を使って、帰無仮説検定やベイズ推定法を駆使するというやつです。この要因計画は実は線形モデルの一環であり、**一般線形モデル (General Linear Model)** と呼ばれています。これを行列で表現することを、ここでは少し考えてみたいと思います。

まずは回帰分析と要因計画は何が同じで何が違うのかを、はっきりさせましょう。同じところは線形モデルであるというところ、違うところは、回帰分析は説明変数も従属変数も連続変数であるのに対し、要因計画では一般に説明変数が離散変数であること、でした。離散変数であるとは、言い方を変えると名義尺度水準の数字だということです。すなわち「統制群」か「実験群」か、という違いを表すのに、0, 1 と言った数字を割り振ったものです。これは別に 3 と 12523, という数字を割り振ったと言ってもいいのです。だって名義尺度水準は、数字と対象が一対一対応していれば良いのですから。

ここでは数学的に話をしやすくするために、統制群を 0, 実験群を 1 とするとしましょう。線形モデルという枠組みは一緒なのですから、従属変数 y_i が説明変数 x_i によって変わるわけですが、ここではこの x が $\{0, 1\}$, というわけです。線形モデルを要素ごとに表現すると次のようになります。

^{*8} 何のことはない、私がそうだったというだけです。学部生の頃は記述統計、推測統計、多変量解析の3領域が別々にあって、それらが繋がっているなんて思ってもみませんでした。

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

この時、 i さんは統制群に割り振られていたとすると、 $x_i = 0$ ですから、この式は次のようになります。

$$y_i = \beta_0 + e_i$$

逆に、 i さんが実験群に割り振られていたとしますと、 $x_i = 1$ ですから、この式は次のようになります。

$$y_i = \beta_0 + \beta_1 + e_i$$

これを見るとわかるように、両群のベースライン β_0 は同じで、そこに効果 β_1 が乗っかるかどうか、が興味の的になります。この式の右辺に i は誤差成分しかなく、誤差を抜きにすると従属変数は β_1 だけ変化するはずだ、というところから、平均値の差を検証しようと言ってることになるのです。

これも x_i が個人ごとに変わるベクトルだと考えると、行列表現では次のようになります。

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

ですからまとめて、次のようにしましょう。

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

おや、これは回帰分析と同じ形ですね！ただし、ここでベクトル $\mathbf{x}$ は、その中身が $\mathbf{x} = (0, 0, 0, 1, 1, 0, 1, 0, \dots)$ のように実験群か統制群かを分けるフラグが入っているだけになります。

以上は実験群と統制群という 2 群の話でしたが、3 群以上になっても基本的なアイデアは同じです。たとえば表 2.2 のようなデータセットがあったとしましょう。

表 2.2 群間要因 (3 水準) のデータセット例

参加者番号	群わけ	従属変数
1	A	3
2	A	3
3	A	4
4	A	4
5	B	6
6	B	7
7	B	8
8	B	9
9	C	7
10	C	6
11	C	5
12	C	4

ここで群ごとの効果を表現したいとすると、次のように書くことになります。

$$\begin{aligned}
 y_1 &= \beta_0 + \beta_1 + e_1 \\
 y_2 &= \beta_0 + \beta_1 + e_2 \\
 y_3 &= \beta_0 + \beta_1 + e_3 \\
 y_4 &= \beta_0 + \beta_1 + e_4 \\
 y_5 &= \beta_0 + \beta_2 + e_5 \\
 y_6 &= \beta_0 + \beta_2 + e_6 \\
 y_7 &= \beta_0 + \beta_2 + e_7 \\
 y_8 &= \beta_0 + \beta_2 + e_8 \\
 y_9 &= \beta_0 + \beta_3 + e_9 \\
 y_{10} &= \beta_0 + \beta_3 + e_{10} \\
 y_{11} &= \beta_0 + \beta_3 + e_{11} \\
 y_{12} &= \beta_0 + \beta_3 + e_{12}
 \end{aligned}$$

添字の対応に注意しながらみてくださいね。群 A の効果は β_1 、群 B の効果は β_2 、群 C の効果は β_3 になります。

この β それぞれを該当するところ (割り当てられた群) だけに対応させつつ、統一的表現形である $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ にするには、次のように書きましょう。

$$\begin{pmatrix} 3 \\ 3 \\ 4 \\ 4 \\ 6 \\ 7 \\ 8 \\ 9 \\ 7 \\ 6 \\ 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \end{pmatrix}$$

このような表記になった時の $\mathbf{X}$ のことをとくに、実験のデザインを表している行列ということで、**デザイン行列 (design matrix)** といいます。デザイン行列は自分で書くと面倒な感じがしますが、とにかくこのような書き方で $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ の統一表現は可能です。

ところでこのデザイン行列、 $\mathbf{X}$ のサイズは今回 $n \times (m + 1)$ になっていますね (m は水準数)。2 水準のときは 2 列で済んだものが、3 水準になると 4 列になるのはおかしくないですか？そうです、1 つ大事なポイントを忘れていました。各群の値はベースライン β_0 からの相対的な違いです。相対的な違いにするというのは、言い換えると全部の水準の和が 0 である必要があるのです。この式は今回の例だと $\beta_1 + \beta_2 + \beta_3 = 0$ であり、これを移項すると明らかなように $\beta_3 = 0 - \beta_1 - \beta_2$ です。つまり総和が決まっているので、自由に大きさを推定できるのは水準数 -1 になります。

これを踏まえてデザイン行列を次のように書き換えることができます。

$$\begin{pmatrix} 3 \\ 3 \\ 4 \\ 4 \\ 6 \\ 7 \\ 8 \\ 9 \\ 7 \\ 6 \\ 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \end{pmatrix}$$

この式は先ほどの式と内容的には同じで、表現の仕方が違うだけです。このように、制約条件も含めた行列の書き方ができるというわけですね^{*9}。

線形モデルの統一的表現、あるいは行列の計算方法に少しは慣れてきたでしょうか。これがさらに多くの変数、未知数を扱うことになると、その利点はよりはっきりしてきます^{*10}。

ここまでで、行列の基本的な形と、これを使って表現することの便利さについて説明してきました。次回以降はいよいよ、行列を使うことの機能的な利点を見ていくことになります。

^{*9} どうして制約条件がいるのか、という話はまた後で出てきますので楽しみに！

^{*10} ここでは触れませんが、被験者内計画・反復測定になっても行列表現はできます。個人差を表すデザイン行列を別途加えることになります。混合計画になると非常に複雑になりますが、それでも一般的な表現は可能です。

第 3 章

行列と方程式

3.1 方程式を行列で書く

線形代数は「便利な書き方」の学問です。便利な書き方をするためにルールが作られていますから、ルールから学ぶと「なんでそんな変な操作をするんだ」という気持ちになるのもわかります。

では何が便利になるのでしょうか。これは方程式を解くことと関係があります。たとえば、以下のような連立方程式があったとしましょう。

$$\begin{cases} x - 2y - 5z = 3 \\ 5x + 4y + 3z = 1 \\ 3x + y - 3z = 6 \end{cases}$$

これは行列で表現すると、次のようになります。

$$\begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix}$$

この左辺を行列とベクトルの式の計算ルールに則って展開してみてください。ちゃんと最初の連立方程式の左辺になることがわかると思います。かけて足して、という面倒な計算ルールは、連立方程式を簡単に表記するためのものだったのですね。

最終的にはこの方程式を解いて、次のように答えを求めます。

$$\begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix}$$

3.2 連立方程式を解こう

さて、ではこの方程式の解はどのようにして求めたのでしょうか。改めてその計算方法を見てみましょう。解を導く方法の 1 つ、消去法は、ひとつの方程式を何倍かして、他の方程式に加えることにより、方程式をどんどん簡単にしていくというものです。まず、第一の式を 5 倍、あるいは 3 倍して、第二、第三の式から x の項を消去します。

$$\begin{cases} x - 2y - 5z = 3 \\ -14y - 28z = 14 \\ -7y - 12z = 3 \end{cases}$$

第二の式の係数を簡単におきましょう。

$$\begin{cases} x - 2y - 5z = 3 \\ y + 2z = -1 \\ -7y - 12z = 3 \end{cases}$$

第二の式を7倍して、第三の式からyを消去します。

$$\begin{cases} x - 2y - 5z = 3 \\ y + 2z = -1 \\ 2z = -4 \end{cases}$$

あとはこれの3行目から $z = -2$ が得られ、芋づる式に $x = -1$, $y = 3$ が得られます。

この操作は、式を一本ずつ、あるいは2つの式を組み合わせ文字を消していく消去法を係数全体に行う操作になっています。実際、ここで操作される係数だけ見ていくと、次のようになっているのです。

■第一段階 ; 一行目の式を使って、二行目以降の二列目の数字をゼロにする。

$$\begin{pmatrix} 1 & -2 & -5 \\ 0 & 1 & 2 \\ 0 & -7 & -12 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 3 \end{pmatrix}$$

■第二段階 ; 二行目の数字を使って、三行目二列目の数字をゼロにする。

$$\begin{pmatrix} 1 & -2 & -5 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ -4 \end{pmatrix}$$

■第三段階 ; 三行目の関係から、対角要素を1にする。

$$\begin{pmatrix} 1 & -2 & -5 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ -2 \end{pmatrix}$$

■第四段階 ; 三行目の関係を使って二行三列目の数字をゼロにする。

$$\begin{pmatrix} 1 & -2 & -5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ -2 \end{pmatrix}$$

■第五段階 ; 二, 三行目の関係を使って、一行二, 三列目の数字をゼロにする。

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}$$

最後の形を見ればお分かりと思いますが、連立方程式を解くこととは、係数の行列を単位行列に置き換えることと同じなのです！

例題 10

次の連立方程式をときましょう。

$$\begin{cases} 2x + 3y - 5z = 3 \\ x + 3y - 2z = 5 \\ 3x + 2y - 4z = 3 \end{cases}$$

この一連の手続きを、行列の言葉で表現し直してみたいと思います。

3.2.1 行の基本変形

行列の行に関して、次の操作を**行基本変形**と言います。

1. ある行に 0 ではない数字を掛ける。
2. ある行に、他の行を何倍かしたものを加える。
3. 二つの行を入れ替える。

連立方程式を解くプロセスは、この 3 つの操作を駆使して行われてきたことは明らかです。これを行列の計算で表せるようにしましょう。

■ある行に 0 ではない数字を掛ける ある行列 A の i 行目を c 倍したいとします。この操作を $P_i(c)A$ とすると、 $P_i(c)$ は次のように表せます。

$$P_i(c) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & c & \\ & & & & 1 \\ & 0 & & & & \ddots \\ & & & & & & 1 \end{pmatrix}$$

単位行列の i 列目だけ c が入っているような行列ですね。数値例で確認しておきます。

数値例として先ほどの連立方程式、

$$\begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix}$$

を行列で表し、

$$Ax = b$$

としましょう。ここで A は**係数行列**と呼ばれます。また、計算に際しては A に右辺のベクトル b を付け加えた行列、すなわち

$$[A, b] = \begin{pmatrix} 1 & -2 & -5 & 3 \\ 5 & 4 & 3 & 1 \\ 3 & 1 & -3 & 6 \end{pmatrix}$$

を考え、一気に操作してしまいましょう。この行列 $[A, b]$ はとくに連立 1 次方程式の**拡大係数行列**と呼ばれます。これを数値例として、行の変形を考えます。

拡大係数行列、

$$A = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & -1 & 3 \\ 3 & 1 & 4 & 1 \end{pmatrix}$$

の 3 行目を 2 倍する場合、

$$P_3(2)A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & -1 & 3 \\ 3 & 1 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & -1 & 3 \\ 6 & 2 & 8 & 2 \end{pmatrix}$$

■ある行に，他の行を何倍かしたものを加える ある行列 A の i 行目に， j 行目の c 倍したものを加えたいと思います (j 行目は変化しない)。この操作を表す行列 $P_{ij}(c)$ は，次のように表せます。

$$P_{ij}(c) = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & \cdots & c & \\ & & & \ddots & \vdots & \\ & & & & 1 & \\ O & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

数値例で確認しておきます。先ほどの行列 A の 3 行目を 2 倍して 1 行目に加えるとしましょう。

$$P_{13}(2)A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & -1 & 3 \\ 3 & 1 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 7 & 4 & 12 & 5 \\ 2 & 1 & -1 & 3 \\ 6 & 2 & 8 & 2 \end{pmatrix}$$

■二つの行を入れ替える ある行列の i 行目と j 行目を入れ替える操作を考えます。これは行列が連立方程式を表していると思えば，連立して成立しているのですから式の順番を入れ替えているだけの操作です。この操作を行列で P_{ij} とするなら，次のように表せます。

$$P_{ij} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & \cdots & \cdots & 1 \\ & & \vdots & 1 & & \vdots \\ & & \vdots & & \ddots & \vdots \\ & & \vdots & & & 1 \\ & & 1 & \cdots & \cdots & 0 \\ O & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

これも数値例で確認しておきます。先ほどの行列 A の 1 行目と 3 行目を入れ替えるとしましょう。

$$P_{13}A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & -1 & 3 \\ 3 & 1 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 4 & 1 \\ 2 & 1 & -1 & 3 \\ 1 & 2 & 4 & 3 \end{pmatrix}$$

この通り，連立方程式を解くための操作もすべて行列演算で表現できることがわかりました。この 3 つの操作をする行列， $P_i(c)$, $P_{ij}(c)$, P_{ij} のことを，**基本行列**といいます。

3.2.2 列の基本変形

少し余談ではありますが，行の基本変形が終わったところで列の基本変形についても言及しておきましょう。行列の列に関して，次の操作を**列基本変形**と言います。

1. ある列に 0 ではない数字を掛ける。
2. ある列に，他の列を何倍かしたものを加える。
3. 二つの列を入れ替える。

これらの操作も基本的に行列で表現でき，興味深いことに行の時に定義した $P_i(c)$, $P_{ij}(c)$, P_{ij} を変形した行列の右側からかけることで実現します。数値例とともに確認しておきましょう。

■ある列に 0 ではない数字を掛ける 先ほどの行列 A の 3 列目を 2 倍する場合の例です。 A が 3×4 行列なので基本行列のサイズが変わっていることに注意してください。

$$AP_3(2) = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & -1 & 3 \\ 3 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 4 & 6 \\ 2 & 1 & -1 & 6 \\ 3 & 1 & 4 & 2 \end{pmatrix}$$

■ある列に、他の列を何倍かしたものを加える このとき $P_{ij}(c)$ は i 行目を c 倍して j 列に加える (i 列目は変化しない) ことを表します。行の基本変形と i, j の意味が変わってることに注意してください。

先ほどの行列 A の 1 列目を 2 倍して 3 列目に加えるとしましょう。

$$AP_{13}(2) = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & -1 & 3 \\ 3 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 6 & 3 \\ 2 & 1 & 3 & 3 \\ 3 & 1 & 10 & 1 \end{pmatrix}$$

■二つの列を入れ替える 先ほどの行列 A の 1 列目と 3 列目を入れ替えるとしましょう。

$$P_{13}A = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & -1 & 3 \\ 3 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 1 & 3 \\ -1 & 1 & 2 & 3 \\ 4 & 1 & 3 & 1 \end{pmatrix}$$

3.2.3 連立方程式を基本行列で解こう

このことを踏まえて、連立方程式の解き方をもう一度くわしく見てみたいと思います。

今から方程式を解くための操作ステップをすべて基本行列で表し、操作のログを取っていくことにします。最初の例と同じ数字ですが、ゴールが係数行列を単位行列にすることだとわかっているのので、途中のステップを少しスピードアップしています。

■第一段階 ; 一行目の式を使って、一列目の非対角要素をゼロにする。

$$\begin{pmatrix} 1 & -2 & -5 & 3 \\ 0 & 14 & 28 & -14 \\ 0 & 7 & 12 & -3 \end{pmatrix} \quad \begin{array}{l} 1. Ab \text{ の } 1 \text{ 行目を } -5 \text{ 倍して } 2 \text{ 行目に加える, すなわち } P_{21}(-5) \\ 2. Ab \text{ の } 1 \text{ 行目を } -3 \text{ 倍して } 3 \text{ 行目に加える, すなわち } P_{31}(-3) \end{array}$$

■第二段階 ; 二行目の数字を使って、二列目の非対角要素をゼロにする。

$$\begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 12 & -3 \end{pmatrix} \quad \begin{array}{l} 3. Ab \text{ の } 2 \text{ 行目を } 1/14 \text{ 倍する. すなわち } P_2(\frac{1}{14}) \\ 4. Ab \text{ の } 2 \text{ 行目を } 2 \text{ 倍して } 1 \text{ 行目に加える. すなわち } P_{12}(2) \\ 5. Ab \text{ の } 2 \text{ 行目を } -7 \text{ 倍して } 3 \text{ 行目に加える. すなわち } P_{32}(-7) \end{array}$$

■第三段階 ; 三行目の数字を使って、三列目の非対角要素をゼロにする。

$$\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{pmatrix} \quad \begin{array}{l} 6. Ab \text{ の } 3 \text{ 行目を } -1/2 \text{ 倍する. すなわち } P_3(-\frac{1}{2}) \\ 7. Ab \text{ の } 3 \text{ 行目を } 1 \text{ 倍して } 1 \text{ 行目に加える. すなわち } P_{13}(1) \\ 8. Ab \text{ の } 3 \text{ 行目を } -2 \text{ 倍して } 2 \text{ 行目に加える. すなわち } P_{23}(-2) \end{array}$$

基本行列による操作は 8 ステップありました。すべてを網羅的につなげて書くと、

$$P_{32}(-2)P_{31}(1)P_3(1/2)P_{23}(-7)P_{21}(2)P_2(1/14)P_{13}(-3)P_{12}(-5)A = I$$

ということになります。係数行列 A の左から右に連なっている点に注意してください。

さてこの操作, 長いので

$$P = P_{32}(-2)P_{31}(1)P_3(1/2)P_{23}(-7)P_{21}(2)P_2(1/14)P_{13}(-3)P_{12}(-5)$$

とまとめて表現してしまいましたが, この操作ログだけに注目してみたいと思います。操作ログだけ取り出すと, 次の関係が成り立ちます。

$$PA = I$$

この P は, 係数行列を単位行列に変えるものであり, いわば行列における逆数のような操作です。

行列でない世界であれば, ある数字 a の逆数を a^{-1} と書きます。ここで

$$aa^{-1} = a^{-1}a = 1$$

です。同様に行列の場合にも, 正方行列 A に対して

$$XA = I \text{ かつ } AX = I$$

な行列があれば, これを特に**逆行列**と呼びます。今回得られた P はまさに A の逆行列なのです。 A の逆行列は A^{-1} と表記します。

方程式を解くことは, 係数行列を単位行列にすることだと言いました。すなわち,

$$Ax = b$$

であるときに,

$$Ix = b^*$$

となれば方程式が解けたといえるわけです。ここで $AA^{-1} = A^{-1}A = I$ ですから,

$$A^{-1}Ax = A^{-1}b$$

とすれば方程式が解けたといえることになります。方程式を解くことは, 逆行列を求めることと同じなのです^{*1}。

また, 逆行列を求める計算方法はいくつかありますが, その1つがこの計算ログを使う方法です。係数行列に行の基本変形を繰り返し, 単位行列に変えることで連立方程式の解を求める方法を, **ガウスの消去法**, あるいは**掃き出し法**といいます。この掃き出し法のログを単位行列に適用すると, その行列の逆行列が求められます。

今回, A に対して $PA = I$ という操作をしたわけですが, この両辺に右から A^{-1} をかけると,

$$PAA^{-1} = IA^{-1}$$

$$PI = A^{-1}$$

となります。つまり, 操作ログ P を単位行列に与えると, 逆行列が求められるわけです。

$$A \xrightarrow{P} I, I \xrightarrow{P} A^{-1}$$

今回の例では, $P = P_{32}(-2)P_{31}(1)P_3(1/2)P_{23}(-7)P_{21}(2)P_2(1/14)P_{13}(-3)P_{12}(-5)$ でしたから,

$$PI = P_{32}(-2)P_{31}(1)P_3(1/2)P_{23}(-7)P_{21}(2)P_2(1/14)P_{13}(-3)P_{12}(-5)I = \frac{1}{28} \begin{pmatrix} 15 & 11 & -14 \\ -24 & -12 & 28 \\ 7 & 7 & -14 \end{pmatrix}$$

^{*1} いいんですよ, ここで感動のあまり立ち上がって拍手しても。

となります。この行列は確かに、

$$\begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix} \frac{1}{28} \begin{pmatrix} 15 & 11 & -14 \\ -24 & -12 & 28 \\ 7 & 7 & -14 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

となっています。

例題 11

次の行列の逆行列を、掃き出し法のログを単位行列に作用させて求めてみましょう。

$$\begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 5 \\ 0 & 1 & 1 \end{pmatrix}$$

3.3 連立方程式を判定する

3.3.1 逆行列

方程式に解けないものがあるように、逆行列も求められない ($XA = AX = I$ のような X が存在しない) ことがあります。存在するとき A は**正則行列 (regular matrix)** であるといいます。レギュラーな行列、ちゃんとしている行列、というような意味です。逆行列が存在するなら、それはただ 1 つに限られます。なぜなら、 A に対して X と Y という 2 つの逆行列があったとしても、

$$X = IX = (YA)X = Y(AX) = YI = Y$$

となるからです。

逆行列の性質として次のものが知られています。正方行列 A, B がいずれも正則であるとする、 A^{-1} , AB , A' はいずれも正則で、以下の関係が成り立ちます。

1. $(A^{-1})^{-1} = A$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $(A')^{-1} = (A^{-1})'$

さて、方程式が解けるというのは逆行列が存在することだ、といいました。逆行列の計算は、方程式の変数が幾つになったとしても、基本行列を使って係数行列を操作するログを取っていけばよいのですが、 2×2 の行列に限っていえば逆行列を簡単に求める公式があります。行列 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ について、 $|A| = ad - bc$ を計算し、 $|A| = 0$ になれば逆行列が存在しないことがわかっています。ここで $|A|$ を行列 A の**行列式 (determinant)** といいます。「式」と言っていますが、計算結果はご覧のとおり、ある「値」ですのでご注意ください。行列式は $\det(A)$ と顯されることもあります。

さてこれを使って、

$$A^{-1} = \frac{1}{|A|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

とすることで逆行列を求めることができます。具体例を見てみましょう。

$$A = \begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix} \text{ とすると,}$$

$$A^{-1} = \frac{1}{1 \times 9 - 2 \times 4} \begin{pmatrix} 9 & -2 \\ -4 & 1 \end{pmatrix} = \begin{pmatrix} 9 & -2 \\ -4 & 1 \end{pmatrix}$$

確認しておきます。

$$AA^{-1} = \begin{pmatrix} 1 & 2 \\ 3 & 9 \end{pmatrix} \begin{pmatrix} 9 & -2 \\ -4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

確かにこの方法で求めたものが、逆行列でした。

行列式は一般に n 次正方行列に対して求めることができますが、サイズが大きくなるとこのような公式で計算できないので、その方法について本書では割愛しています。大事なのは、この行列式は n 次連立方程式に解があるかどうかの判定に用いられるものだ、ということです。サイズが大きくなっても、 $|A| = 0$ であれば逆行列が存在しないといえます。これは 2×2 の場合を考えて貰えばわかりますが、 $|A| = 0$ とすると分母 0 での割り算になるので、計算が成立しないからです。

例題 12

次の行列について、逆行列を持つかどうか判定し、もつならそれを示してください。

1. $\begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix}$

2. $\begin{pmatrix} 2 & -6 \\ -1 & 3 \end{pmatrix}$

3. $\begin{pmatrix} 4 & -6 \\ -2 & 3 \end{pmatrix}$

逆行列の性質

3 つの性質それぞれを証明しておきます。まず逆行列の逆行列は元に戻るというもの ($(A^{-1})^{-1} = A$)。これは逆行列の定義が $AX = I$ かつ $XA = I$ であること、 $AA^{-1} = A^{-1}A = I$ より明らかです。

次に 2 つの正則行列の積の逆行列についてですが、 AB が正則でその逆行列が $B^{-1}A^{-1}$ だとすると、 $AX = XA = I$ ではなく、

$$(AB)(B^{-1}A^{-1}) = ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I$$

$$(B^{-1}A^{-1})(AB) = B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I$$

が確かに成り立っていることから証明できました。

最後に転置した行列の逆行列は、その行列の逆行列を転置したものになるというやつですが、これも逆行列の定義から考えて、

$$A'(A^{-1}) = (A^{-1}A)' = I' = I$$

$$(A^{-1})'A = (AA^{-1})' = I' = I$$

となり、 A は正則で $(A')^{-1} = (A^{-1})'$ が成り立つことが言えました。

逆行列が存在しないと連立方程式が解けません。この「解けない」というのは 2 通りあって、無数の解をもつか、解を持たないのかのどちらかです。さて、無数の解をもつとはどういうことでしょうか？これが次の話題です。

3.3.2 階段行列とランク

連立 1 次方程式が次のようにあったとします。

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n \end{aligned}$$

ここで $n = m$ と断っていないことに注意してください。つまり、変数の数と方程式の数が違う可能性も含めて記述しています。さて係数、未知数、定数が作る行列を考えて、次のように表します。

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

こうすると、連立方程式は

$$Ax = b$$

となります。

さて、行列の基本変形を繰り返すと、係数行列 A を特別な形の行列に変えていくを考えます。どう特別かというと、

1. ある $k (1 \leq k \leq n)$ について 1 行目から k 行目まではどれもゼロベクトルでないし、残りの $n - k$ 行は全てゼロベクトルである。
2. 第 i 行 ($1 \leq i \leq k$) の成分を左から順にみて、最初に出てくる 0 でない数字は 1 で、出てくる場所が q_i 列目だったとすると、 $q_1 < q_2 < \cdots < q_k$ である。
3. 第 q_i 列は基本ベクトル e_i である。

という性質を持っているのです。言葉ですとややこしいですが、要するに矩形行列の左下がゼロでいっぱいになっており、最初に 1 の要素が出てくるところは基本ベクトルになっているわけですから、次のような形をしている行列です。

$$\begin{pmatrix} 1 & 0 & a_{13} & 0 \\ 0 & 1 & a_{23} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

この行列をよく見ると、上記の「特別な形のルール」は守っています。

- 1 行目は 1 列目に 1 があります。そしてその下は全部ゼロ、つまり 1 列目のベクトルは $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ と基本ベクトルになっています。
- 2 行目は 2 列目に 1 があって、左から見て 2 列目までは 0、2 列目だけ見ると基本ベクトルという形。
- 3 行目は 4 列目まで 1 が出てきません。そこまで 0 が続いていて、1 が出てきた列は基本ベクトルになっています。
- 4 行目は 1 がないので特にどの条件にも違反していません。
- 左列から見ていって、1 が出てくる順番は 1,2,3... とそろっています。

このように、左から見て1が出るまで段段になっていること、下はゼロでうめつくされているもので、このような行列を**階段行列**といいます^{*2}。0で階段を作っているように見えるからですね。この例の場合は3段ありますが、このような階段行列におけるゼロベクトルでない行の数のことを、この行列の**階数 (rank)** といい、 $\text{rank}(\mathbf{A})$ で表します。

いくつか階段行列の例を見ておきましょう。次に挙げるのはいずれも階段行列です。

$$\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 4 & 3 \\ 0 & 1 & 5 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 6 & 0 \\ 0 & 1 & 8 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

任意の行列 $\mathbf{A}$ は、適当な基本変形を何回か繰り返すことによって、必ず階段行列 $\mathbf{B}$ に変形できることがわかっています。具体例は既に連立方程式の解のところで見た通りです。単位行列 $\mathbf{I}$ は階段行列であり、連立方程式の係数行列 $\mathbf{A}$ と右辺の定数 $\mathbf{b}$ をあわせた拡大係数行列について係数部分を単位行列にすること、つまり拡大係数行列を階段行列にすることが、方程式を解くということだったのは明らかです。そして拡大係数行列の階数と、係数行列の階数が一致することが、方程式において解が求められるということと同じ意味になりますね。すなわち、

$$\text{rank}([\mathbf{A}, \mathbf{b}]) = \text{rank}(\mathbf{A})$$

となることが、連立1次方程式 $\mathbf{Ax} = \mathbf{b}$ が解をもつための必要十分条件なのです。

連立方程式において、未知数の数 m と階数の関係を係数行列・拡大係数行列との関係から見ていくことにしましょう。

■ $\text{rank}([\mathbf{A}, \mathbf{b}]) > \text{rank}(\mathbf{A})$ の場合 まずは拡大係数行列の階数が係数行列の階数より大きかった場合です。これは方程式の形で表してみればすぐにわかります。

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}$$

つまり係数行列の階数は2です ($\text{rank}(\mathbf{A}) = 2$)。拡大係数行列の階数はどうなるでしょうか。

$$\begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 5 \end{pmatrix} \xrightarrow{\mathbf{P}_{13}(-2/5)\mathbf{P}_{23}(-2/5)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

この通り $\text{rank}([\mathbf{A}, \mathbf{b}]) = 3$ で、係数だけの行列の階数 $\text{rank}(\mathbf{A})$ より大きくなっています。ところがこの場合、もともとの方程式をよくみると、

$$\begin{cases} x & = & 2 \\ y & = & 3 \\ 0 \times z & = & 5 \end{cases}$$

となっていますから、方程式として成立していないですね。これでは解なし、となってしまうのも当然です。

^{*2} 階段行列のなかでも最初に1が出てくるときに基本ベクトルとして出てきており、これを特に簡約化された階段行列、Reduced row echelon form といって区別することもあります。ここでは村上 et al. (2016) に習って簡約化されたものを階段行列としています。いずれにせよ本書の目的では、簡約化された形での利用しかしませんので。

■ $\text{rank}([A, b]) = \text{rank}(A) = m$ の場合 階数が一致していて、未知数の数と同じであれば一組の解が求まるのはもうお分かりだと思いますが、たとえば3つの未知数にたいして4つ以上の連立方程式がある場合はどうなるのでしょうか。階数の観点からみてみましょう。次のような連立方程式があったとします。

$$\begin{cases} x & +y & -z & = & 4 \\ x & +3y & +5z & = & 2 \\ -x & +2y & +3z & = & 0 \\ 2x & +y & +2z & = & 2 \end{cases}$$

拡大係数行列は次のようになります。

$$[A, b] = \begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & 3 & 5 & 2 \\ -1 & 2 & 3 & 0 \\ 2 & 1 & 2 & 2 \end{pmatrix}$$

これに基本操作をしていきます。

$[A, b]$	$\begin{pmatrix} 1 & 1 & -1 & 4 \\ 1 & 3 & 5 & 2 \\ -1 & 2 & 3 & 0 \\ 2 & 1 & 2 & 2 \end{pmatrix}$
$P_{21}(-1)$ $P_{31}(1)$ $P_{41}(-2)$	$\begin{pmatrix} 1 & 1 & -1 & 4 \\ 0 & 2 & 6 & -2 \\ 0 & 3 & 2 & 4 \\ 0 & -1 & 4 & -6 \end{pmatrix}$
$P_2(1/2)$ $P_{12}(-1)$ $P_{32}(-3)$ $P_{42}(1)$	$\begin{pmatrix} 1 & 0 & -4 & 5 \\ 0 & 1 & 3 & -1 \\ 0 & 0 & -7 & 7 \\ 0 & 0 & 7 & -7 \end{pmatrix}$
$P_3(-1/7)$ $P_{23}(-3)$ $P_{43}(-7)$	$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

このように、第3段階で操作が終わってしまいました。拡大係数行列の階数が3、係数行列の階数も3であり、両者が一致するので解が得られます。しかし最後の一行はなんの情報もない行になってしまいました。つまり、元の連立方程式には、一本分冗長な式が含まれていたということです。

■ $\text{rank}([A, b]) = \text{rank}(A) < m$ の場合 最後に、係数行列と拡大係数行列の階数は同じだけど未知数がそれより大きい場合。次のような例で見てみましょう。

$$\begin{cases} x & +2y & +3z & -w & = & 10 \\ x & +y & -2z & +w & = & -2 \\ 2x & -2y & -z & -w & = & -1 \end{cases}$$

拡大係数行列にしてみます。

$$[\mathbf{A}, \mathbf{b}] = \begin{pmatrix} 1 & 2 & 3 & -1 & 10 \\ 1 & 1 & -2 & 1 & -2 \\ 2 & -2 & -1 & -1 & -1 \end{pmatrix}$$

階段行列を目指していきましょう。

$[\mathbf{A}, \mathbf{b}]$	$[\mathbf{A}, \mathbf{b}] = \begin{pmatrix} 1 & 2 & 3 & -1 & 10 \\ 1 & 1 & -2 & 1 & -2 \\ 2 & -2 & -1 & -1 & -1 \end{pmatrix}$
$P_{21}(-1)$ $P_{31}(-2)$	$\begin{pmatrix} 1 & 2 & 3 & -1 & 10 \\ 0 & -1 & -5 & 2 & -12 \\ 0 & -2 & -7 & 1 & -21 \end{pmatrix}$
$P_2(-1)$ $P_{12}(-2)$ $P_{32}(2)$	$\begin{pmatrix} 1 & 0 & -7 & 3 & -14 \\ 0 & 1 & 5 & -2 & 12 \\ 0 & 0 & 3 & -3 & 3 \end{pmatrix}$
$P_3(1/3)$ $P_{13}(-7)$ $P_{23}(-5)$	$\begin{pmatrix} 1 & 0 & 0 & -4 & -7 \\ 0 & 1 & 0 & 3 & 7 \\ 0 & 0 & 1 & -1 & 1 \end{pmatrix}$

これを方程式の形に戻して表現すると、

$$\begin{cases} x & -4w & = & -7 \\ & y & +3w & = & 7 \\ & & z & -w & = & 1 \end{cases}$$

となります。これは結局、

$$\begin{cases} x & = & 4w + 7 \\ y & = & -3w + 7 \\ z & = & w + 1 \end{cases}$$

であり、 w が何になるか定めないと答えが出てきません。このように、変数 1 つ分の自由ができてしまいます。無数の解をもつというのはこのように、未知数の数のほうが拡大係数行列の階数よりも大きい場合なのです。この未知数の数 m と係数行列の階数の差分、 $m - \text{rank}(\mathbf{A})$ のことを**解の自由度**と言います。

ここまでのことをまとめると、以下ようになります。

- 係数行列 $\mathbf{A}$ の階数 $<$ 拡大係数行列 $[\mathbf{A}, \mathbf{b}]$ であれば、解はない。
- 係数行列 $\mathbf{A}$ の階数 $=$ 拡大係数行列 $[\mathbf{A}, \mathbf{b}]$ $=$ 未知数の数 m であれば、1 組の解がある。
- 係数行列 $\mathbf{A}$ の階数 $=$ 拡大係数行列 $[\mathbf{A}, \mathbf{b}] <$ 未知数の数 m であれば、無数の解がある。
- 解の自由度 k は、 $k = m - \text{rank}(\mathbf{A})$ である。

3.3.3 自明な解，終結式，母体

最後に余談になりますが、どうして行列のことを行列というのかについて説明しておきます。いえ、ふざけているわけではなく、日本語で行列というのは行と列からなるので当たり前、適当な名前なのですが、行列の英語である *matrix* とは、という話です。というのも *matrix* を辞書で引くと「母体」という意味が出てきて「なぜ線形代数の名称にこんな意味が？」となること請け合いだからです*3。

*3 これは筆者が映画 *Matrix* を見たときに引っかかったところで、仮想空間の話なのにどうして行列なんて名付けたんだろうと思って、改めて辞書を引いて驚いたからです。映画はこの「母体」という方のニュアンスでつけられているんですね。

このセクションでは方程式の係数を行列で表現し、これを解くためには逆行列が必要で、逆行列があるかどうかは行列式でわかる、という話をしてきましたね。多くの線形代数のテキストもこのような順に説明していきますが、歴史的には逆に行列式が先にあってから、行列の発見に至ったのです(竹内, 2010)。

そもそもは、2つの方程式が同じ解を持つかどうか、という問いです。次のような2つの式があったとします。

$$\begin{aligned} f(x, y) &= ax + by = 0 \\ g(x, y) &= cx + dy = 0 \end{aligned}$$

これは今までやってきた行列の形で表現すると、次のようになります。

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

つまり

$$Ax = 0$$

ですね。

これは定数項のところが0になっている連立方程式のことであり、一般に未知数 m 個の連立一次方程式 $Ax = b$ において、 $b = 0$ の場合を**同次連立一次方程式**と呼びます。

さてこの2つの式 f, g が同じ解 x, y を持つかと言われれば、当然 $x = 0, y = 0$ が思いつきます。これはいつだって成立しちゃいます。でもこれは当たり前すぎて面白くもなんともない。このように $x = 0$ なのは自明な解というのですが、自明でない答えがあるかどうかを知りたいわけです。

たとえば

$$\begin{cases} 2x + y = 0 \\ 6x + 3y = 0 \end{cases}$$

を考えると、 $x = -\frac{1}{2}y$ がどちらにも成立しています。つまり、2つの式が同じ解をもちます。

しかしたとえば、

$$\begin{cases} 2x + y = 0 \\ 3x + 3y = 0 \end{cases}$$

を考えると、上の式は $x = -\frac{1}{2}y$ ですが、下の式は $x = -y$ ですから、同じ解をもちません。

さてどういう時に同じ解を持って、どういう時に同じ解を持たないかを判別したいとします。自明でない答えがあるとして、 $f(x, y)$ と $g(x, y)$ の解が $x = \alpha, y = \beta$ だとしましょう。式 f から、

$$\alpha = -\frac{b}{a}\beta$$

、式 g から

$$\alpha = -\frac{d}{c}\beta$$

という関係が得られます。これが連立するということは、

$$\frac{b}{a} = \frac{d}{c}$$

$$ad = bc$$

$$ad - bc = 0$$

であることが条件、ということになります。

おや、何かみたことがありますか。そう、 2×2 の行列式はこの形でしたね。

係数を並べて行列を作り, 行列式を考えてみると, 確かにそうになっています。

$$R = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

そもそもの問い(「2つの方程式が同じ解を持つかどうか」)はここから, 自明な解以外の共通の解を持つかどうかは, 行列式が0かどうか, というように問い方が変わりました。

行列式より先にこの形が出てきた, ということでしたが, この式($ad - bc = 0$)は**終結式**といいます^{*4}。同様の話は3つの方程式, 4つの方程式の場合...と一般化され, 様々なケースでの終結式を考えていく中で, どうも係数を使った世界, **母体となる解の空間**があるようだぞ, ということから, 行列 *matrix* の形が発見されていったのです。

終結式 $R = 0$ が成立すれば自明でない共通の解があるかどうかを判別できる, いやさこれは係数行列を考えてその解が出るかどうかを決定する式(行列式 *determinant*)を考えよう, これがあれば逆行列が考えられて方程式の答えが出てくる, というのが行列誕生の歴史的経緯です。

- $Ax = 0$ が自明な解 $x = 0$ のみを持つための必要十分条件は, $\text{rank}(A) = m$
- $Ax = 0$ が自明でない解を持つ(無数の解をもつ)ための必要十分条件は, $\text{rank}(A) < m$

ということができます。

行列とベクトル, 連立方程式との関係が密であることがいろいろ見えてきたと思います。次回からはこの話を, 空間的イメージのほうに拡張していきます。

^{*4} ちなみに終結式や行列式に最初に辿り着いていたのは, 日本の数学者関孝和だったと言われています。

第II部

Day 2: 行列とベクトルの空間的理解

第 4 章

ベクトルの空間的理解

4.1 ベクトルの可視化

ここまで代数 (文字と式) としての行列の話をしてきましたが, ここからは空間的なイメージで考えていきたいと思います。何かを理解するときの入り口はいろいろあった方がよいと思います。数字の計算では面白みがわからなかった人も, 空間イメージであればわかりやすいということがあるかもしれません。もちろん数字の羅列が大好きな人はそのままでいいのですが, 空間イメージと合わせて考えることができるという, 数学世界の完璧なまでの整合性に感動するのも一興ではないでしょうか。

心理統計学的な観点からみると, 線形代数は多変量解析に用いるツールということになります。しかし線形代数は, たとえばコンピュータグラフィックスの世界でも必要な知識です。CG というと空間をさまざまな角度から描画するような, 視点の移動や空間の回転などが特徴的かとおもいますが, この空間の回転は行列の仕事です。また, 物理学では物質に働く力, その大きさと向きを考えますが, これはベクトルとして表現されます。線形代数の数学的側面はこのように, 統計学や情報科学, 物理学の間にあってそれぞれの世界を媒介する共通言語なのです。

それではさっそく, ベクトルの空間的イメージをつかむところから始めましょう。

4.1.1 ベクトル空間

話を簡単にするために, 二次元の空間を基本に考えましょう。二次元の列ベクトル, $\mathbf{a} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ を, x 軸が 1, y 軸が 2 の座標を指していると考えerのです。「指している」というと「どこから?」という問いが思い浮かびますが, 基本的にこれは「原点 $o = (0, 0)$ から」と限定しておきましょう^{*1}。

ベクトルは数字として考えても, 空間として考えても良く, 両方の観点を行き来して考えるようにしていこう, というのがここでのテーマです。さてこのベクトル $\mathbf{a}$ は, x 軸上のベクトル $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ と, y 軸上のベクトル $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ を使って, 次のように表現することができます。

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

これはつまり, x 軸と y 軸を表現する基本的な単位となるベクトルがあれば, それらを使って

^{*1} 物理学でベクトルを扱うときは, 「どこから」というのは特に気にせず, 向きと大きさを持っていればそれがベクトルだ, と考えるのですが, 私はそのあたりから抽象的すぎてついていけないので, 地に足をつけるように原点から, と限定して考えることにしています。そのようにしても, 本書の目的からいって問題はありません。ちなみに, 原点を固定したベクトルを位置ベクトル, 固定して考えないものを幾何ベクトル, 数学的に数字のリストだけにこだわる場合は数ベクトルなどと呼んで区別することもあります (村上 et al., 2016)。

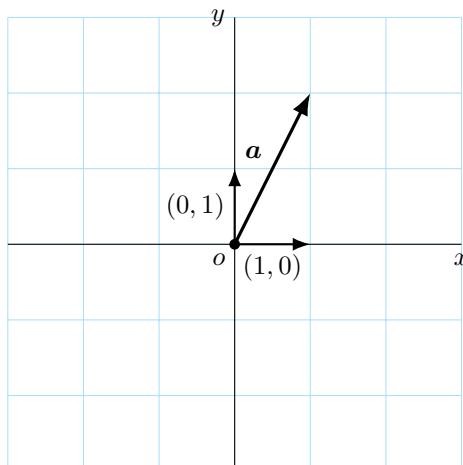


図 4.1 基底ベクトルとその線型結合で表される座標

二次元平面のあらゆる座標を表現することができる、ということでもあります。任意の座標 (x, y) は、

$$\begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

で表せるからです。この式の右辺は**線型結合**とか**一次結合**と呼びます。また、座標の基本となるベクトル、 $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ と $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ を**基底**と呼びます。

このようにベクトルとその空間表現は相互に行き来できるということがわかるかと思います。ここではわかりやすく二次元で描画していますが、三次元であってもいいし、四次元でも n 次元でも、数学的には一般化して扱うことができます。

例題 13

ベクトル $\mathbf{a} = \begin{pmatrix} 1 \\ 9 \end{pmatrix}$ を、ベクトル $\mathbf{b} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$, $\mathbf{c} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ の一次結合で表しましょう。

4.1.2 一次独立と一次従属

基底はベクトルで作る空間の基礎となるベクトル、ということですが、どのようなベクトルであってもいいというわけではありません。例えば次のような式を考えてみましょう。

$$x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

この式が成立するのは、係数 x, y が共にゼロ、すなわち $x = y = 0$ の時だけであることがわかります。このような関係にあるベクトル $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ と $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ であれば、基底として考えることができます。このような関係を特に**一次独立**なベクトル、といいます。

ではなぜ一次独立でないと基底にならないのでしょうか。次のようなベクトルの例で考えてみましょう。

$$x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

この場合ですと、 $x = -2, y = 1$ とすれば数式は成立しますね。このような関係は一次独立ではなく、**一次従属**なベクトルであるといいます。従属している、というのは図を見ればその関係は明らかです。2 本目のベ

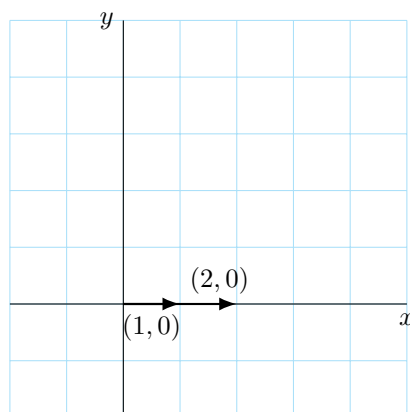


図 4.2 一次従属なベクトルの組み合わせ

クトルは、1 本目のベクトルに従っていますね。言い換えると、 $2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ ですから、2 本目のベクトルは 1 本目のベクトルの一次結合で表現できてしまうのです。これでは結局、2 番目の軸としての情報を持っていないことになりますから、一次従属な関係のベクトルであれば空間を張ることができない、基底となることができない、というわけです。

連立方程式の空間的な意味

方程式 $ax + by = c$ があつたとします。 $a = b = 0$ の場合を除いて、この式は xy 平面に直線として表すことができますね。たとえば $6x - 3y = 9$ とすると、 $y = 2x - 3$ ですから、右上がりの直線で表現できます。またたとえば、 $c = 0$ すなわち $2x - y = 0$ のような方程式であれば、 $y = 2x$ と原点を通る直線になります(図 4.3 参照)。

ちなみに一本目の式 $6x - 3y = 9$ と $x + 3y = 12$ という式が連立していたとすると、その解は交点の座標 $(3, 3)$ になります。つまり、一点で交わる 2 直線ですから解がただ 1 つに決まるというわけです。1 本目と 2 本目は平行線ですから、交点がありません。すなわち、この連立方程式は解無し、ということになります。

では一次従属なベクトルの連立方程式は、ということこれは図 4.2 のようにピッタリ重なった一本ですから、その上の点はどれでもいいことになる、つまり「解が無数にある」という状態になるのです。

これは多次元空間になっても同じことで、「解がない」というのはそれぞれの方程式が張る空間に共通して交わる点がないということ、「解がある」というのはそれぞれの連立方程式が張る空間に共通する一部分があるということ、「解が無数にある」というのはどこかの空間を共有しちゃって定まらない、ということでもあります。

三次元空間の「解なし」「解が 1 つだけ」「解が無数にある」というパターンを可視化したのが図 4.4 です^{*2}。

例題 14

次の 3 つのベクトル、 $\mathbf{a} = \begin{pmatrix} 3 \\ 0 \\ 3 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 4 \\ 1 \\ 4 \end{pmatrix}$, $\mathbf{c} = \begin{pmatrix} -7 \\ 1 \\ 1 \end{pmatrix}$ は一次独立なベクトルか一次従属か判定してください。

^{*2} この素晴らしい可視化のアイデア、出典は村上 et al. (2016) です。

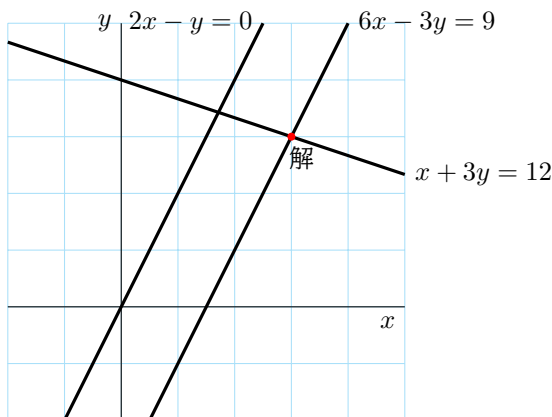


図 4.3 方程式の可視化

ヒント 1

ベクトルが一次独立かどうかを判定するには、 $\lambda_1 \mathbf{a} + \lambda_2 \mathbf{b} + \lambda_3 \mathbf{c} = \mathbf{0}$ を考えてみれば良いでしょう。 $\lambda_1, \lambda_2, \lambda_3$ が全て 0 にしかならないのであれば一次独立ですし、1 つでも 0 でない組み合わせがあれば一次従属なはず。であればこの問題は、

$$\begin{pmatrix} 3 & 4 & -7 \\ 0 & 1 & 1 \\ 3 & 4 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

の方程式の解が 1 組 (自明な解) なのか、無数の解をもつのかで判断できることになります。係数行列に掃き出し法を適用してみましょう!

4.1.3 ベクトルの長さ

ところで、x 軸、y 軸の基底ベクトルは $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ と、 $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ということでしたが、ベクトルは方向がわかっているだけでいいので、これは別に $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$ と、 $\begin{pmatrix} 0 \\ 3 \end{pmatrix}$ などとしても問題ありません。ただ、目盛り 1 単位分にしておいた方が定数倍するときにわかりやすいので、長さを 1 にしておいたにすぎません。

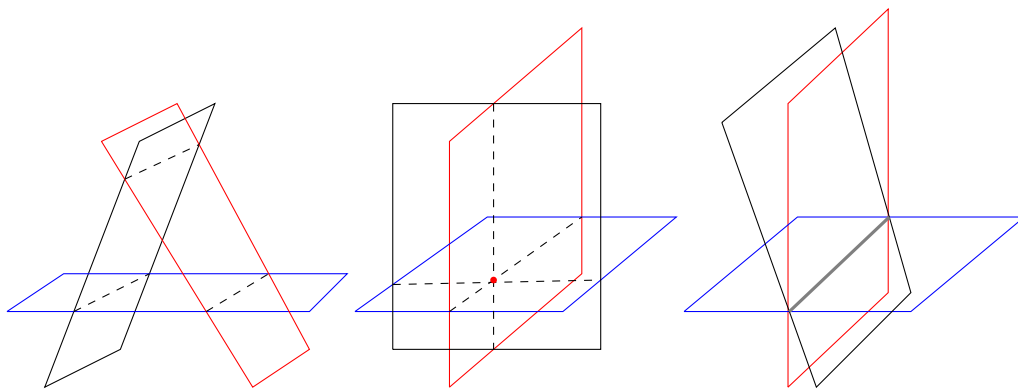


図 4.4 方程式の可視化:3次元の場合。左から順に、解なし、解が1つ、無数の解

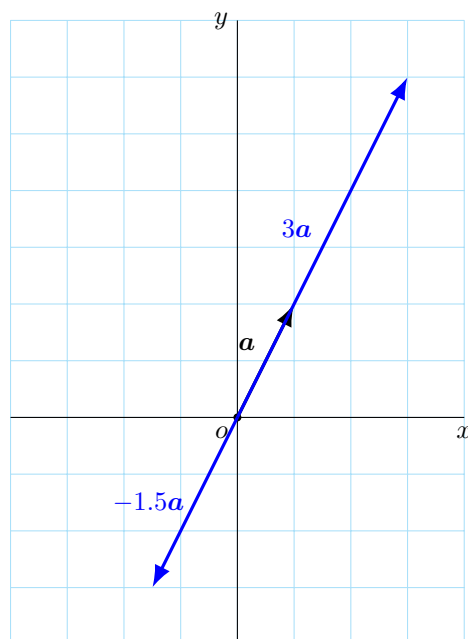


図 4.5 ベクトルのスカラー倍

ベクトルの長さ、という言葉が出てきましたが、先ほどのベクトル $\mathbf{a} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ の長さはどうか、ということも考えることができます。ベクトルの長さのことをノルムといい、 $\|\mathbf{a}\|$ と書きます。これは三平方の定理を使って、

$$\|\mathbf{a}\| = \sqrt{1^2 + 2^2} = \sqrt{5}$$

と計算することができます。一般に、 n 次元ベクトル $\mathbf{a}' = (a_1 \ a_2 \ \dots \ a_n)$ に対して、ノルム $\|\mathbf{a}\|$ は

$$\|\mathbf{a}\| = \sqrt{\sum_{i=1}^n a_i^2}$$

と定義されます。

4.2 ベクトルとベクトルの関係

4.2.1 ベクトルの計算の空間イメージ

さてベクトルで空間のイメージができてきたかとおもいますが、この計算はどのように空間化できるでしょうか。まずはベクトルのスカラー倍を見てみましょう。ベクトル $\mathbf{a} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ を 3 倍すると、

$$3\mathbf{a} = 3 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \times 1 \\ 3 \times 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$$

とそれぞれの要素に掛け込むことができるのでした。空間上は、最初のベクトルと同じ方向に伸ばすことに対応します。同じ方向に長さだけ変える操作をするので、掛ける数字は大きさの調整役を意味します。ただの数字のことをスカラーと言いますが、これは Scaler(スケーラー、スケール担当) から来ているのです。

また、スカラー倍は同じ線上でサイズが変わるだけですが、負のスカラー倍は向きが逆転することを意味します。これは数直線のイメージで考えると、 $+5 - 2 = +3$ という計算は、 x 軸を右に 5 単位行って、そこからさらに -2 単位進むというのは左に進むことと同じ、ということから理解できるかと思います。

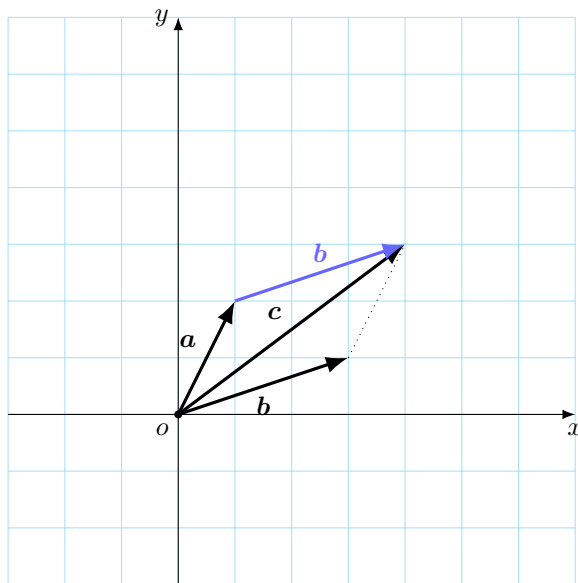


図 4.6 ベクトルとベクトルの和。平行四辺形の対角線,あるいは移動のイメージで

今度はベクトルの和を考えてみましょう。別のベクトル $\mathbf{b} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ があったとして、ベクトルの和はどうなるでしょうか。数字の計算として考えたら、

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1+3 \\ 2+1 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

となりますから、図でいうと「2つのベクトルが作る平行四辺形の対角線が示すもの」と考えることもできますし、「x軸方向に+1, y軸方向に+2移動したところから、x軸方向に+3, y軸方向に+1移動する」と考えても良いかと思います。

4.2.2 内積

ベクトルの和の話が出たので、今度はベクトルの積の話をしてしまおう。ベクトルとベクトルの積は、掛ける順番によってはスカラーになったり行列になったりします。ベクトル、 $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ と $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ について、 $\mathbf{a}$ を転置して行ベクトルにし、行ベクトルと列ベクトルの積にしてみると、

$$\mathbf{a}'\mathbf{b} = (a_1 \quad a_2) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = a_1b_1 + a_2b_2$$

となります。なんだこれ。と、そのまえに、同じことをベクトル $\mathbf{a}$ と $\mathbf{a}$ でやってみますと、

$$\mathbf{a}'\mathbf{a} = (a_1 \quad a_2) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1^2 + a_2^2 = \|\mathbf{a}\|^2$$

となります。これはノルムの二乗ですね。ベクトルの長さの計算は、ベクトルとベクトルの積 (ただし行ベクトル $\times$ 列ベクトルの形にして) だった、ということがわかります。であれば、 $\mathbf{a}'\mathbf{b}$ で算出されるスカラーも、何らかの量を表しているわけですが、これが何なのかを考えてみたいと思います。ちなみにこの積の計算のことをとくに、**内積**と呼びます。

内積は $(\mathbf{a}, \mathbf{b})$ とも表記され、一般に次のように定義されます。

$$(\mathbf{a}, \mathbf{b}) = \sum a_i b_i$$

内積はまた、次の性質を持つことが知られています。

- $(a, b) = (b, a)$
- $(a + b, c) = (a, c) + (b, c)$
- $(\lambda a, b) = \lambda(a, b)$
- $(a, a) \geq 0$ ただし等号になるのは $a = 0$ の場合に限る。

例題 15

次の3つのベクトル、 $a = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $b = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$, $c = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ について、上の内積の性質が全てみたされることを確認しましょう。

さてこの内積、2つのベクトルの長さに関する何らかの量であることはわかりますが、どういう量なのか空間イメージで考えてみたいと思います。

いくつかのベクトルを考えましょう。図 4.7 には、 $a = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$, $b = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$, $c = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$, $d = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$, $e = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ の5つのベクトルがあります。それぞれのノルムや内積は、

$$\|a\| = \|b\| = \|c\| = 5$$

$$\|d\| = \sqrt{20}$$

$$\|e\| = 1$$

$$(a, b) = 12 + 12 = 24$$

$$(a, c) = 15 + 0 = 15$$

$$(a, d) = -8 + 12 = 4$$

$$(a, e) = 3 + 0 = 3$$

などとなっています。結果から考えて、2つのベクトルが近いと内積は大きく、離れると小さくなるようですから、内積は2つのベクトルの関係の近さを表しているようにも見えますが、 $(a, c) > (a, d)$ になっているあたりが微妙です。

実は、内積は2つのベクトルの関係の近さを表していると言えなくもないのですが、そこには2つのベクトルが作る角度が関係しています。角度に注意しながら、ベクトル a と e , a と c を例に見てみましょう。

ベクトル e は x 軸の基本ベクトルで、長さは1です。ベクトル c はその5倍の長さがありますが、 e と方向が同じ x 軸上のベクトルです。それぞれの内積をみると、 (a, c) は (a, e) の5倍ですから、大きさの意味はともかく、ベクトルのノルムに何かかけたものという関係にあることがわかります。

何をかけているのでしょうか。そのヒントがベクトル a との角度です。2つのベクトルの角度が θ だとします。またベクトル a からベクトル e や c に向けて垂線を下ろしたとします。このように、一方のベクトルから他方のベクトルへ垂線を下ろすことを、**射影**する、といいます。光を当てて影を落とすイメージですね。影が落ちる点までの距離は、三角関数を思い出すと $\|a\| \cos \theta$ であることがわかります (図 4.8)。

つまりベクトルの角度のコサインをとったものが、内積 (a, e) , (a, c) に関わってきているといえそうです。 a と e の内積 (a, e) は、ベクトル a を x 軸に射影した時の長さであり、 $(a, e) = \cos \theta \|a\| \times 1 = \cos \theta \|a\| \|e\|$ だったわけです。このように考えると、ベクトル a と c との内積も、 $(a, c) = \cos \theta \|a\| \|c\|$ といえそうです。内積は、一方のベクトルを他方のベクトルに射影した時の長さと、射影されたベクトルの長さ、2つのベクトルの作る角度の $\cos \theta$ からできています。

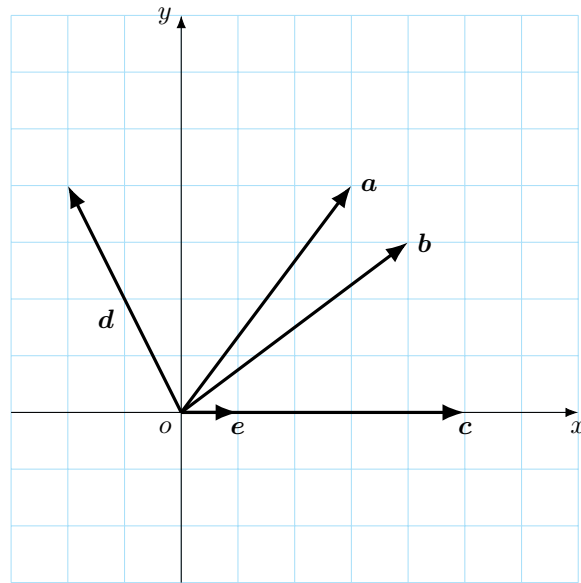


図 4.7 色々なベクトルの例

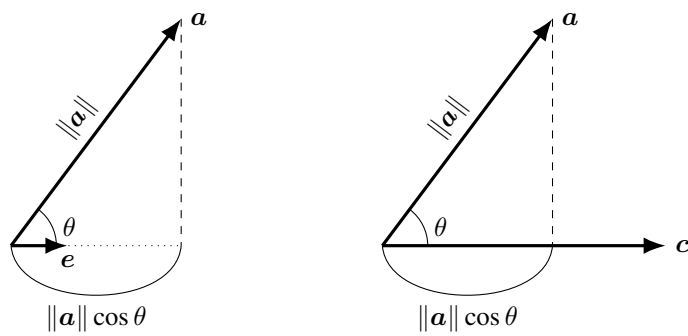


図 4.8 ベクトルのなす角と内積の関係

今回は a を e に、 a を c に射影して説明しましたが、逆でも構いません。例えば c を a に射影し、射影された a のノルム倍したものとして捉えても同じことです (図 4.9)。

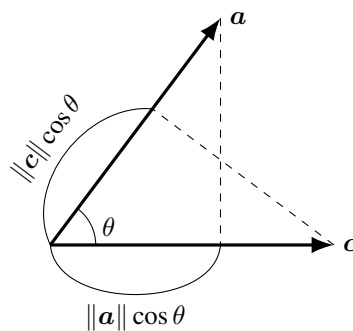


図 4.9 内積と射影関係

ポイントは、内積は2つのベクトルの類似度を表していること、その大きさは両方のベクトルのノルムと角度

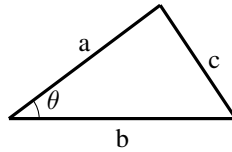
によって決まること、にあります。ちなみに内積の計算から逆に、

$$\cos \theta = \frac{(\mathbf{a}, \mathbf{c})}{\|\mathbf{a}\| \|\mathbf{c}\|}$$

となりますから、 $\cos \theta$ は内積を両方のノルムで割ったものだ、と考えてもいいでしょう。

余弦定理から考える

まさか三角関数が出てくるとはな！と思った人もいるかも知れませんが、ついでに三角関数を使ったうちくをひとつ。 $\cos$ のことを余弦といいます、これを使った定理として三角形の3つの辺の長さ a, b, c と、2つの辺が作る角度 θ から、

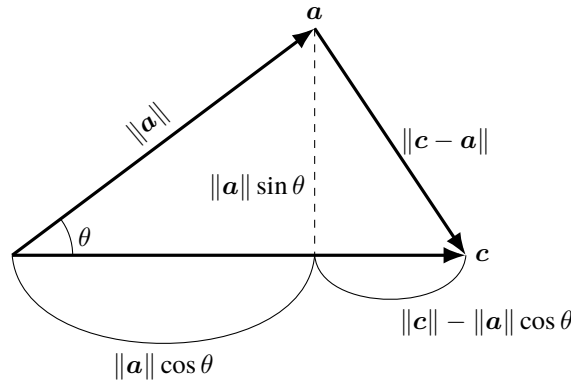


$$c^2 = a^2 + b^2 - 2ab \cos \theta$$

というのがあります。これを今回のベクトルの話に置き換えてみます。

辺の長さはベクトルでいうところのノルムですね。それを踏まえてベクトルの図で考えますと、この式がいつていることは次のように言い換えられます。

$$\|\mathbf{c} - \mathbf{a}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{c}\|^2 - 2\|\mathbf{c}\|\|\mathbf{a}\|\cos \theta$$



三角形の時の第三の辺 c が、ここでは $\|\mathbf{c} - \mathbf{a}\|$ となっているところに注意してください。一旦 c 地点まで移動して、そこから a 地点まで戻るの、マイナスになるという計算です。ここで、ベクトル $\mathbf{a}$ から $\mathbf{c}$ に対して射影することを考えると、ベクトル $\mathbf{c}$ は $\|\mathbf{a}\| \cos \theta$ と残りの部分に分割されます。またこの垂線の高さは $\sin \theta$ を使って表すことができます。この垂線で区切られた右側の領域にある三角形も直角三角形ですから、三平方の定理が成立しますね。

$$\|\mathbf{c} - \mathbf{a}\|^2 = (\|\mathbf{a}\| \sin \theta)^2 + (\|\mathbf{c}\| - \|\mathbf{a}\| \cos \theta)^2$$

ということで、右辺を展開してみますと

$$\begin{aligned} &= \|\mathbf{a}\|^2 \sin^2 \theta + \|\mathbf{c}\|^2 - 2\|\mathbf{c}\|\|\mathbf{a}\| \cos \theta + \|\mathbf{a}\|^2 \cos^2 \theta \\ &= \|\mathbf{a}\|^2 (\sin^2 \theta + \cos^2 \theta) + \|\mathbf{c}\|^2 - 2\|\mathbf{c}\|\|\mathbf{a}\| \cos \theta \\ &= \|\mathbf{a}\|^2 + \|\mathbf{c}\|^2 - 2\|\mathbf{c}\|\|\mathbf{a}\| \cos \theta \end{aligned}$$

ここで $\cos \theta$ の項を別にしてベクトルだけの計算にしてみましょう。移項して、

$$2\|\mathbf{c}\|\|\mathbf{a}\|\cos\theta = \|\mathbf{a}\|^2 + \|\mathbf{c}\|^2 - \|\mathbf{c} - \mathbf{a}\|^2$$

として考えてみます。

右辺はベクトルだけになりましたから、2次元ベクトルの要素として $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, $\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ だとすると、

$$\begin{aligned} 2\|\mathbf{c}\|\|\mathbf{a}\|\cos\theta &= a_1^2 + a_2^2 + c_1^2 + c_2^2 - ((a_1 - c_1)^2 + (a_2 - c_2)^2) \\ &= a_1^2 + a_2^2 + c_1^2 + c_2^2 - (a_1^2 + c_1^2 - 2a_1c_1 + a_2^2 + c_2^2 - 2a_2c_2) \\ &= 2(a_1c_1 + a_2c_2) \end{aligned}$$

ここで両辺2で割って、

$$\|\mathbf{c}\|\|\mathbf{a}\|\cos\theta = (a_1c_1 + a_2c_2) = (\mathbf{a}, \mathbf{c})$$

となり、内積は2つのベクトルのノルムにベクトルのなす角 θ の余弦をかけたもの、という話にたどりつきます。

4.3 データの空間的イメージ

ここまでで、ベクトルは空間的なものとして可視化、イメージできることをみてきました。これと心理統計とはどういう関係があるのでしょうか。

例えば次のようなデータセットがあったとします。ここで身長と体重の散布図を描く、ということになれば、

表 4.1 3人 × 2変数のデータセット

名前	身長	体重
鈴木	160	70
佐藤	150	80
小杉	176	53

いわば行ベクトルごとにプロットしていくわけですからベクトルが個人を表していることになります。たとえばクラスター分析などで回答者を分類するときなどは行ベクトル同士の関係を考えるわけですから、個々人の向きや大きさをに興味関心があるわけです。

ですが、たいていのデータ解析の時は変数間関係が知りたいわけですから、ベクトルを縦に見る、つまり列ベクトルとして考えることになるでしょう。どちらの見方であっても本質的には変わりません。もちろん、列ベクトルとして考える時、今回は3人でしたから可視化できましたが、 n 人になるともう無理です。でも、数学的には n 次元になってもその性質は変わりませんから、イメージとして変数を空間上の矢印として想像していただければと思います。

さて、身長と体重の例のように、変数ベクトルを考えたとします。この時、変数には単位がついていますから、相対的な位置関係を考えるには変数の標準化というプロセスがあるのでした。変数の標準化は、

$$z_{ij} = \frac{x_{ij} - \bar{x}_{.j}}{s_j}$$

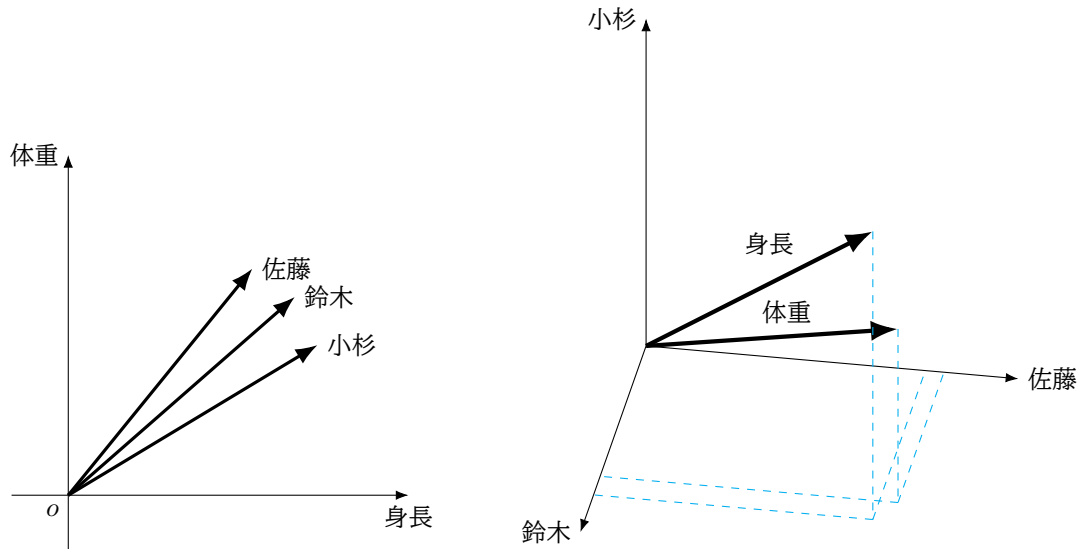


図 4.10 行でみるか列で見るか

とすることでしたが、標準化された変数ベクトルのノルムを考えてみますと、

$$\begin{aligned}
 \|z_j\|^2 &= \sqrt{\sum_{i=1}^n z_{ij}^2} \\
 &= \sqrt{\left(\frac{x_{1j} - \bar{x}_j}{s_j}\right)^2 + \left(\frac{x_{2j} - \bar{x}_j}{s_j}\right)^2 + \cdots + \left(\frac{x_{nj} - \bar{x}_j}{s_j}\right)^2} \\
 &= \sum_{i=1}^N \frac{(x_{ij} - \bar{x}_j)^2}{s_j^2} \\
 &= \frac{1}{s_j^2} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \\
 &= \frac{1}{s_j^2} n s_j^2 = n
 \end{aligned}$$

です。標準化とは、ベクトルのノルムを $\sqrt{n}$ に整えることだったのですね。

これを踏まえて、2つの変数ベクトル z_j と z_k の内積を考えてみましょう。ノルムが $\sqrt{n}$ ですから、

$$(z_j, z_k) = \|z_j\| \|z_k\| \cos \theta = n \cos \theta$$

となります。また、内積の計算をエレメントワイズで考えてみると、

$$(z_j, z_k) = \sum_{i=1}^n z_{ij} z_{ik} = n r_{jk}$$

ですから、相関係数は2つの変数ベクトルの $\cos \theta$ であることがわかります。

$$r_{jk} = \cos \theta$$

さて、余弦 $\cos \theta$ は角度 θ がどんな値であろうと、 $-1 \leq \cos \theta \leq +1$ の範囲に収まりますね。相関係数も $-1 \leq r_{jk} \leq +1$ の範囲に収まるのでした。相関係数の範囲 (絶対値は1以下) であることを証明するには、

シュワルツの不等式などを用いることが一般的ですが、相関係数は変数ベクトルのなす角の $\cos$ だからね、でもいいわけです。素敵。

4.3.1 直交

2つのベクトルの内積が0になるとき、これらのベクトルは**直交**する、といいます。

ベクトルの内積は $(\mathbf{a}, \mathbf{b}) = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta$ ですから、これがゼロになるのは2つのベクトルのノルムが0か、 $\cos \theta$ がゼロかの場合です。 $\cos \theta = 0$ というのは角度でいうと90度のことから、直角に交わるすなわち直交です。

直交していると相関がないというのは、空間イメージですぐに理解できます。図4.11の左の図は、2つのベクトルが直交している例ですが、この場合変数ベクトル $\mathbf{a}$ 上にある値が p から p' に変わったとしても、他方のベクトル $\mathbf{b}$ の値には影響しません。しかし右図にあるように直交でない場合は、一方の変数ベクトルでの変化が他方での変化に影響します。片方を動かすと他方も動くというのが共変動 (共分散, 相関) ということです。直交でなければ相関があることになります。

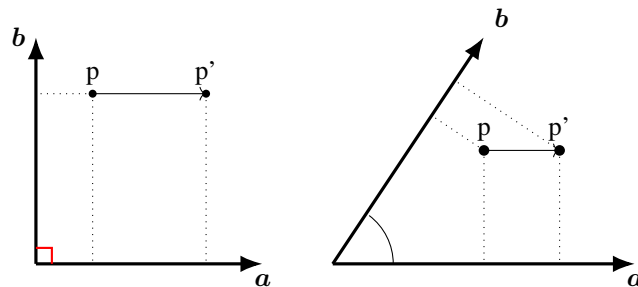


図 4.11 直角の場合は一方の変数の違いが他方に影響しない

さて、あるベクトル $\mathbf{a}$ は、あるベクトルとそれに直交するベクトルに分解することができます。図4.12にはベクトル $\mathbf{a}$ を直交するふたつのベクトルに分解する、いくつかのパターンを描いています。式で表現するなら、

$$\mathbf{a} = \mathbf{b}_1 + \mathbf{b}_2 = \mathbf{c}_1 + \mathbf{c}_2 = \mathbf{d}_1 + \mathbf{d}_2$$

ということです。ここにあるように、分解の仕方は何通りも考えられますが、元のベクトル $\mathbf{a}$ に直交する方向には分解ができません。ベクトルを分解するというのは、あるベクトルから他方のベクトルに向けて垂線を下す、すなわち射影することです。そうすることで、分解したい方向に平行な成分と、それと直交する成分とに分解できることになります。しかし最初から直交している方向に対しては、射影しても影が落ちないのです。分解したい方向に平行な成分がないので、いわば $\mathbf{a} = \mathbf{0} + \mathbf{a}$ のような形にしかならないということです。

このようなベクトルの分解の話は、2次元以上の空間、変数ベクトルが作る空間であっても一般的に成立します。データのベクトルが n 次元空間で表されていたとしても、そのベクトルと直交でない任意のベクトルに垂線を下すことができます。

ところで、空間の元になるベクトルは基底といいましたが、その条件は一次独立であることであって、直交していることではありません。直交していない空間であったとしても、垂線を下ろして座標を計算できますから、基底が作る空間の任意の座標が基底ベクトルの一次結合で表現できることに変わりはないのです。

そうした一般例から考えると、いわゆる x - y 座標は非常に整った形をしています。冒頭に示した図4.1は、基底として $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ と、 $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ を挙げました。これはノルムが1に整えられており (正規化されている)、直交していて、基底になっていますから、特に**正規直交基底**と呼ばれます。長さや角度についての特徴があるので

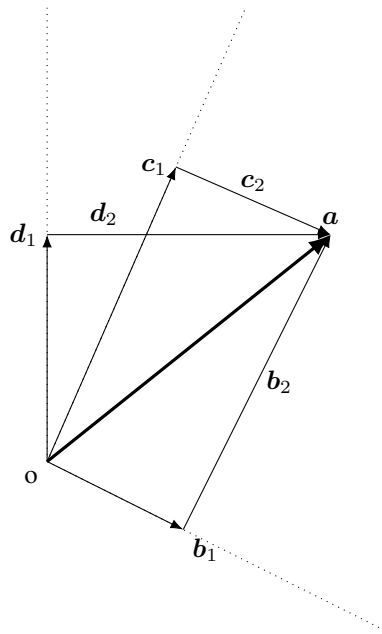


図 4.12 ベクトルを直交分解する

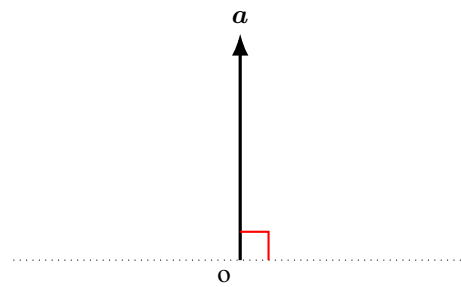


図 4.13 最初から直交している方向には影が落ちないので、分解できない

「整った」という感じがするんですね。

さて、ここまでで、ベクトルを空間的イメージで捉え、長さや角度で表したり、座標をよみとるように基本となる軸に射影する、分解することができるのを見てきました。次回は複数のベクトルが作り出す空間のことを考えていきたいと思います。

第 5 章

ベクトル空間

ベクトルを数字のリストとしてだけではなく、空間としてイメージできるという話をしてきました。ここからは、ひとつひとつのベクトルではなく、ベクトル全体のつくる空間について考えていきたいと思います。ここでのポイントは、行列が空間を作ること、行列は座標を変える変換装置であるということです。早速みていきましょう。

5.1 ベクトル空間

実数を成分とする p 次元ベクトル全体の作る集合を $\mathbb{R}^p$ で表すとします。これを p 次元ベクトル空間、と言います^{*1}。集合の記号で書くと、

$$\mathbb{R}^p = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \middle| x_1, x_2, \dots, x_p \in \mathbb{R} \right\}$$

ということになります。二次元平面であれば $\mathbb{R}^2$ 、三次元空間であれば $\mathbb{R}^3$ のように書けばよいですね。

この p 次元ベクトル空間 $\mathbb{R}^p$ において、一次独立な p 次元ベクトルは p 個までしか選ぶことができません。基底の話を思い出しましょう。 p 次元ベクトルのうち、第 i 要素が 1 で他が全て 0 であるようなベクトルのことを**基本ベクトル**と呼びますが、次の二つの基本ベクトルは二次元の正規直交基底です。

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

この 2 つのベクトルさえあればその一次結合の形で、空間上のあらゆる座標を作ることができるのでした。

言い換えると、任意のベクトル $\mathbf{x}$ は、

$$\mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2$$

で表現できるので、この $\mathbf{e}_1, \mathbf{e}_2$ 以外に別のベクトル $\mathbf{x}$ を用意しても新しい基底となることはできません。

ところで、正規直交基底以外のベクトルで空間を張ることも可能です。例えばベクトル $\mathbf{z}_1 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ と $\mathbf{z}_2 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$ は内積がゼロではないので $((\mathbf{z}_1, \mathbf{z}_2) = -1)$ 直交していませんが、基底の条件は一次独立であればいいので、空間を作ることはできます。

^{*1} 厳密には実数ベクトル空間です。複素数を成分とするベクトルからなる空間 $\mathbb{C}^p$ は複素ベクトル空間と言い、区別します。心理学データ解析の文脈ではほとんどの場合実数だけで話が成立しますが、非対称多次元尺度法の中には複素ベクトル空間に拡張されたモデルがあります。詳しくは千野 et al. (2012) を参照してください。

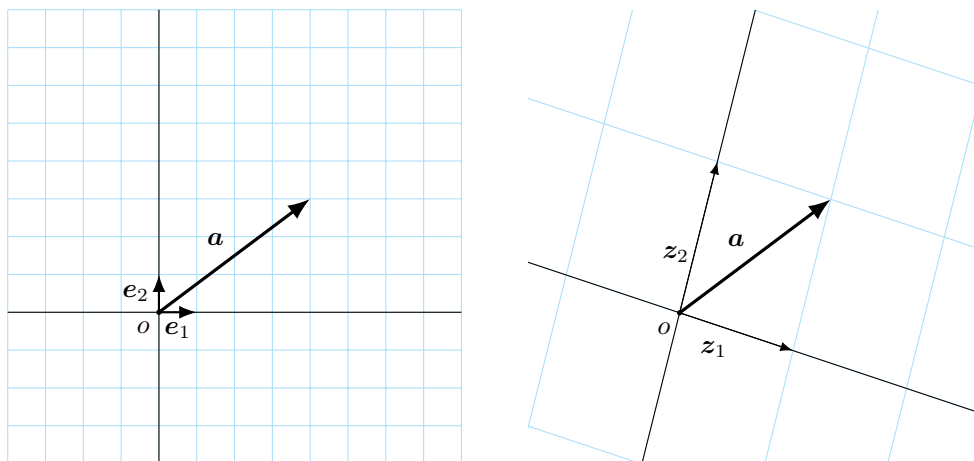


図 5.1 基底の変換

具体的に考えてみましょう。正規直交基底からなる二次元平面上に、ベクトル $\mathbf{a} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ があったとして、これを先ほどの z_1 と z_2 で表すと、

$$\begin{pmatrix} 4 \\ 3 \end{pmatrix} = 1 \begin{pmatrix} 3 \\ -1 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ 4 \end{pmatrix} = 1z_1 + 1z_2$$

で表せます。基底の取り方を変えたので、座標が変わったのです。図 5.1 にはこの座標の変換の例を示しています。図中のベクトル $\mathbf{a}$ は原点からの向きと大きさが同じものですが、図の左はいわゆる x, y 座標によるプロットをしており、そこでの座標は $\begin{pmatrix} 4 \\ 3 \end{pmatrix}$ ですが、右側の座標では $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ になっていますね。

このように、基底はいろいろな取り方ができますし、今回の例のように直交でなくてもかまいません^{*2}。図 5.1 のような座標であっても二次元に変わりはなく、基底の一次結合で空間全てを表現することはできるので、基底の数は一定です。一般に、ベクトル空間 $\mathbb{R}^p$ に含まれる基底の数 p は**次元 (dimension)** と呼び、 $\dim(\mathbb{R}^p) = p$ のように表します。

5.1.1 グラム・シュミットの正規直交化法

ここで示したあたらしい基底 z_1, z_2 は直交していませんが、応用上は直交した基底のほうが色々便利なので、基底を直交化したい、というときもあるかと思います。そこで直交していない基底から、直交した基底を作る方法である**グラム・シュミットの正規直交化法**について説明します。

直交だけでなく、ついでに正規化もおきましょうね。正規化とは長さを 1 にすることですから、

$$\mathbf{u}_1 = \frac{\mathbf{z}_1}{\|\mathbf{z}_1\|} = \frac{\begin{pmatrix} 3 \\ -1 \end{pmatrix}}{\sqrt{3^2 + 1^2}} = \begin{pmatrix} \frac{3}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} \end{pmatrix}$$

とします。ここで $\mathbf{u}_1$ はこれから作る正規直交基底のひとつになります。

また、セクション 4.3.1 でみたようにベクトルはあるベクトルとそれに直交するベクトルに分解することができるのでした。なので、 z_2 のほうを z_1 あるいは $\mathbf{u}_1$ の方向とそれに直交するベクトル $\mathbf{u}_2$ の方向に分解して

^{*2} 内積 $(\mathbf{z}_1, \mathbf{z}_2) = \|\mathbf{z}_1\| \|\mathbf{z}_2\| \cos \theta = -1$ ですから、 $\sqrt{10}\sqrt{17} \cos \theta = -1$ より 116° ぐらいです。計算して $\cos \theta = -\frac{1}{\sqrt{170}}$ から、 $\mathbb{R}$ の $\arccos$ 関数で求めてもいいですし、Wolfram alpha に「{3,-1}と{1,4}のなす角」と聞いてもいいかもしれません。

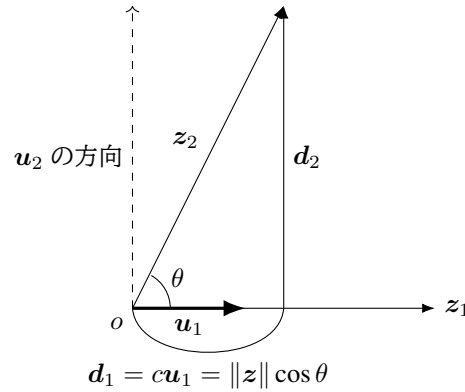


図 5.2 新しく作る直交基底の方向と大きさ

やりましょう。分解するベクトルを、ここでは d_1, d_2 とおきましょう。 d_1 は正規化された新しい基底 u_1 の定数倍ですので $d_1 = cu_1$ となります。

$$z_2 = d_1 + d_2 = cu_1 + d_2$$

$$d_2 = z_2 - cu_1$$

さて d_1 ベクトルの長さ c はどれぐらいでしょう。図 5.2 に少し角度を調整して図示してみましたが、この長さは三角関数を使って $c = \|z_2\| \cos \theta$ で表せそうですね。ここでは z_2 と u_1 の内積を考えると、 u_1 はノルム 1 に整えているのですから

$$(z_2, u_1) = \|z_2\| \|u_1\| \cos \theta = \|z_2\| \cos \theta$$

となります。ここから $c = \|z_2\| \cos \theta = (z_2, u_1)$ ということになります。

まとめると、

$$\begin{aligned} d_2 &= z_2 - d_1 \\ &= z_2 - cu_1 \\ &= z_2 - (\|z_2\| \cos \theta) u_1 \\ &= z_2 - (z_2, u_1) u_1 \end{aligned}$$

となります。実際に数値を入れて考えてみましょう。

$$d_2 = \begin{pmatrix} 1 \\ 4 \end{pmatrix} - \left(\begin{pmatrix} 1 & 4 \end{pmatrix} \begin{pmatrix} \frac{3}{\sqrt{10}} \\ -1 \end{pmatrix} \right) u_1 = \begin{pmatrix} 1 \\ 4 \end{pmatrix} - \frac{-1}{\sqrt{10}} \begin{pmatrix} \frac{3}{\sqrt{10}} \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix} - \begin{pmatrix} -0.3 \\ 0.1 \end{pmatrix} = \begin{pmatrix} 1.3 \\ 3.9 \end{pmatrix}$$

となります。この d_2 を正規化すれば u_2 が作れますから、

$$u_2 = \frac{d_2}{\|d_2\|} = \frac{\begin{pmatrix} 1.3 \\ 3.9 \end{pmatrix}}{\sqrt{1.69}} = \begin{pmatrix} 0.316 \\ 0.948 \end{pmatrix}$$

が得られます。この u_1, u_2 は z_1 の方向を基軸とし、直交する正規化された基底になります。この手続きを一般化しておきます。

■グラム・シュミットの直交化法 $a_1, a_2, \dots, a_p$ からなる p 次元空間について、次の手順で正規直交基底 $b_1, b_2, \dots, b_p$ を作ることができる。

1. a_1 を正規化して $b_1 = a_1 / \|a_1\|$ をつくる。
2. $d_2 = a_2 - (a_2, b_1)b_1$ として a_1 に直交するベクトルを求める。

3. d_2 を正規化して $b_2 = d_2 / \|d_2\|$ をつくる。
4. 以下同様に, $d_k = a_k - (a_k, b_1)b_1 - (a_k, b_2)b_2 - \cdots - (a_k, b_{k-1})b_{k-1}$ のように, k 番目のもとのベクトル a_k から, それまでに得られた直交する次元の要素を取り除いて, k 番目の直交するベクトルを求める。
5. d_k を正規化する。
6. ステップ 4 と 5 を $k = p$ になるまでワンステップずつ繰り返す。 $k = p$ になれば終了。

計算は面倒ですが^{*3}, ここで「次元の分解はいく通りもある」という感覚を掴んでおくと, 因子分析など多変量解析を理解する時にイメージがしやすくなるでしょう。

例題 16

ベクトル $a_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, $a_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, $a_3 = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$ から $\mathbb{R}^3$ の正規直交基底を作ってみましょう。

5.2 部分空間と直交補空間

あるベクトルの集合から, 任意に 2 つのベクトル x_1 と x_2 を選んだ時, 任意のスカラー b_1, b_2 をつかって表現される $b_1x_1 + b_2x_2$ がもとの集合に含まれているのであれば^{*4}, このベクトルの集合を**部分ベクトル空間**といいます。部分ベクトル空間は, 必ずゼロベクトルを含みます ($b_1 = b_2 = 0$ の場合)。

2 つの部分ベクトル空間に対して, 次のような集合を考えましょう。

$$M_1 \cap M_2 = \{x \mid x \in M_1, x \in M_2\}$$

これはどちらの部分空間にも属するベクトルからなる集合で, **積空間**と言います。また,

$$M_1 + M_2 = \{x \mid x = x_1 + x_2, x_1 \in M_1, x_2 \in M_2\}$$

はどちらかの空間に入っていればよいという集合で, **和空間**と言います。これらはいずれも部分ベクトル空間です。 $\mathbb{R}^p$ に 3 つの部分空間 W, W_1, W_2 があって, $W = W_1 + W_2$ かつ, $W_1 \cap W_2 = \{0\}$ であれば, W は W_1 と W_2 の**直和**であるといい, $W = W_1 \oplus W_2$ と表します。

これを踏まえて, ベクトルの分解のように空間の分解を考えます。

$\mathbb{R}^p$ の任意の部分空間 W があったとき, $\mathbb{R}^p$ のベクトル a が, W に属する全てのベクトルに直交するなら, a は W に直交する, といい $a \perp W$ と書きます。このベクトル全体の集合 $W^\perp$ をとくに**直交補空間**といいます。直交補空間もまた $\mathbb{R}^p$ の部分空間です。

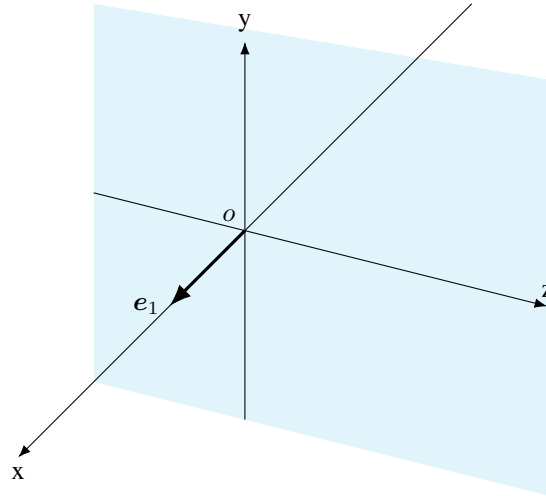
直交補空間もまた部分空間になる

確認しておきましょう。 $u, v \in W^\perp$ とします。この u, v が加法とスカラー倍しても $W^\perp$ に入っていれば良いのです。ここで $x \in W$ を用意して考えると, 内積の性質から

$$(x, u + v) = (x, u) + (x, v) = 0 + 0 = 0$$

^{*3} Wolfram 先生に `gram schmidt{...}` と聞いたり, R 先生に `matlib::GramSchmidt(X)` などと聞く前に, 一度は自分で計算してみましょう

^{*4} 「加法とスカラー倍に関して閉じている」と言ったりします

図 5.3 e_1 ベクトルが作る空間 (1 次元の直線) と直交補空間

ですから, $u + v \in W^\perp$ が言えました。スカラー倍についても同様に, $\lambda \in \mathbb{R}, u \in W^\perp$ として, $x \in W$ と内積の性質から

$$(\lambda u, x) = \lambda(u, x) = \lambda \cdot 0$$

となり, $\lambda u \in W^\perp$ が言えました。また $x \in W$ に対してゼロベクトルとの内積, $(0, x) = 0$ であることは明らかですから, 直交補空間もゼロベクトルが含まれている ($0 \in W^\perp$) が確認できます。

イメージで考えてみましょう。 $\mathbb{R}^3$, つまり三次元空間において, ある基本ベクトル $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ が作る部分空間 W は, x 軸だけの一次元空間 (直線) です。この直交補空間を考えてみます。 e_1 と直交する三次元ベクトル $v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ を考えると, これは e_1 と直交する, つまり内積が 0 になるわけですから,

$$(e_1, v) = x + 0 + 0 = 0$$

つまり $x = 0$ であればよく, 直交補空間 $W^\perp$ は

$$W^\perp = \left\{ \begin{pmatrix} 0 \\ y \\ z \end{pmatrix} \middle| y, z \in \mathbb{R} \right\}$$

となります。これは y 軸, z 軸について任意ですからこの 2 つの軸で作る平面空間が直交補空間ということになります。図 5.3 に空間イメージで書いてみました。3 次元空間が 1 次元空間と 2 次元空間に分解されていることがわかるかと思います。あるいは 1 次元空間の残りを補うようなのが補空間だということです。

数式的にも, 次のようにいうことができます。 $\mathbb{R}^p$ の部分空間 M とその直交補空間 $M^\perp$ を考えると, $M + M^\perp = \mathbb{R}^p, M \cap M^\perp = \{0\}$ なので, 任意のベクトル $x \in \mathbb{R}^p$ は

$$x = x_1 + x_2 (x_1 \in M, x_2 \in M^\perp)$$

と表すことができ, また

$$\dim(M) + \dim(M^\perp) = p$$

が成り立ちます。

例題 17

W を $\mathbf{a} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}$ という 2 つのベクトルによって生成される, $\mathbb{R}^3$ の部分空間とします。このとき, $W^\perp$ の基底を見つけましょう。

ヒント 2

3 次元空間のうち 2 つの基底ベクトルが見つかったので, 探すべきベクトル (次元) はあと 1 つです。このベクトルを $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ と置いて, W の 2 つの基底との内積がゼロになることから考えれば...

5.3 ランクと射影

5.3.1 ランクのさまざまな性質

以前, 連立方程式と行列の話をした時に, **階数 (rank)** という話が出てきました (→ セクション 3.3.2, Pp.47)。また, ベクトル同士が一次独立かどうかを判断する時に, ベクトルが作る行列の階数を求め, ベクトルの数と階数が一致するようであればそれぞれのベクトルでしか表せない要素があるから独立である, と判断できるのでした (→ セクション 4.1.2, Pp.56)。このように, 方程式とベクトル, ベクトルと行列, 行列の独立な成分の数と空間の次元の数は密接な関係があります。

ここでは行列のランクとその性質を見てみましょう。

まず n 個の p 次元ベクトル, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ を並べた行列 $\mathbf{A}$ を考えます。サイズは $p \times n$ です。

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$$

このとき, 掃き出し法によって $\mathbf{A}$ の階数を求めることができますが, これは $\mathbf{A}$ という空間が持つ次元数でもあるわけです。

具体例で考えてみます。3 個の 2 次元ベクトル, $\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 5 \\ 1 \end{pmatrix}$ からなる行列 $\mathbf{A} = \begin{pmatrix} 1 & 2 & 5 \\ 1 & -1 & -1 \end{pmatrix}$ を考えると, 掃き出し法によって次元数が求められます。

$$\begin{pmatrix} 1 & 2 & 5 \\ 1 & -1 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 5 \\ 0 & -3 & -6 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 5 \\ 0 & 1 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

この場合, 行列 $\mathbf{A}$ のランク・次元は 2 になります。

ここで n 次元ベクトル $\mathbf{x} = (x_1, x_2, \dots, x_n)' \in \mathbb{R}^n$ に行列 $\mathbf{A}$ を左からかけると, $p \times n$ と $n \times 1$ の行列の積ですから $p \times 1$, つまり p 次元ベクトルが得られます。例えば, 次のような計算です。 $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ とすると,

$$\mathbf{Ax} = \begin{pmatrix} 1 & 2 & 5 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 11 \\ -1 \end{pmatrix}$$

となりますから, 3 次元ベクトルが 2 次元ベクトルになりました。このように, 行列 $\mathbf{A}$ は $\mathbb{R}^n$ 次元から $\mathbb{R}^p$ 次元への**写像**をする「座標変換装置」だと考えることができます。一般的に書くと,

$$\mathbf{y} = \mathbf{Ax} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n$$

ということですね。この写像されたベクトル $\mathbf{y}$ が張る p 次元空間全体を像 (Image) といって $Im\mathbf{A}$ と表します。

$$Im\mathbf{A} = \{\mathbf{y} : \mathbf{y} = \mathbf{A}\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n, \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^p$$

この $Im\mathbf{A}$ は行列 $\mathbf{A}$ の列ベクトルが張る空間であり、行列 $\mathbf{A}$ の階数を k とすると、 $Im\mathbf{A}$ の階数も k です。

$$\dim(Im\mathbf{A}) = \text{rank}(\mathbf{A}) = k$$

$\mathbf{A}$ の階数が k というのは、 k 個の列ベクトル $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$ が一次独立で、残りの $\mathbf{a}_{k+1}, \mathbf{a}_{k+2}, \dots, \mathbf{a}_n$ は一次独立なベクトルの一次結合で表されるわけですから、

$$Im\mathbf{A} = \{\mathbf{y} : \mathbf{y} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_k\mathbf{a}_k\}$$

とすることができます。この $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$ は $Im\mathbf{A}$ の基底と考えることができます。

これは先ほどの例で考えればわかりやすいですね。 $\mathbf{A} = \begin{pmatrix} 1 & 2 & 5 \\ 1 & -1 & -1 \end{pmatrix}$ は、 $1\mathbf{a}_1 + 2\mathbf{a}_2 = \mathbf{a}_3$ でしたから、 $\mathbf{a}_1, \mathbf{a}_2$ だけが $Im\mathbf{A}$ の基底なのです。

また、移した先がゼロベクトル $\mathbf{0}$ になってしまうような元のベクトル集合のことを核 (Kernel) といいます。

$$Ker\mathbf{A} = \{\mathbf{x} \in V \mid \mathbf{A}\mathbf{x} = \mathbf{0}\}$$

例の行列 $\mathbf{A}$ のカーネルは、 $1\mathbf{a}_1 + 2\mathbf{a}_2 = \mathbf{a}_3$ という関係から、あるいは

$$\begin{pmatrix} 1 & 2 & 5 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

という関係から、

$$\begin{cases} x_1 + 2x_2 + 5x_3 = 0 \\ x_1 - x_2 - x_3 = 0 \end{cases}$$

が得られますから、これを解いて $x_1 = 1, x_2 = 2, x_3 = -1$ より、 $\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ を基底とする部分ベクトル空間、

$$Ker(\mathbf{A}) = \{\mathbf{x} : \mathbf{x} = c \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}\}$$

と表すことができます。

さて元の n 次元ベクトル $\mathbf{x}$ が作る空間 $V \in \mathbb{R}^n$ のうち、 $Ker\mathbf{A}$ は潰れて無くなるわけですから、 $Im\mathbf{A}$ の次元数はもとの次元 n から $Ker\mathbf{A}$ の次元数 $\dim(Ker\mathbf{A})$ を引いたもの、すなわち

$$\dim(Im\mathbf{A}) = \text{rank}(\mathbf{A}) = n - \dim(Ker\mathbf{A}) \leq p$$

となります*5。

このように、行列を座標変換装置だと考えること、変換装置の次元数を考えることは、多変量解析を考える時に関わってくるアイデアですので、是非しっかりとイメージしておいてください。その際、次元数がどうなるかということを考えるために、ランクが重要な発想になってくるのです。

ランクはいろいろな意味や性質があります。ここまでみてきたように、行列 $\mathbf{A}$ の一次独立なベクトルの最大本数、ということもできますし、 $Im\mathbf{A}$ の次元数でもあります。他にも次のようなことが言えます。

*5 移される先の空間 $\mathbf{W}$ は p 次元空間ですが、 $\mathbf{x}$ 以外のところから別の写像で $Im\mathbf{A}$ のそとに落ちることについては特に決まり事はありません。言い換えれば、ある $\mathbf{y} \in \mathbb{R}^p$ に対して $\mathbf{y} = \mathbf{A}\mathbf{x}$ となる $\mathbf{x} \in \mathbb{R}^p$ が存在しない可能性もあるからです。

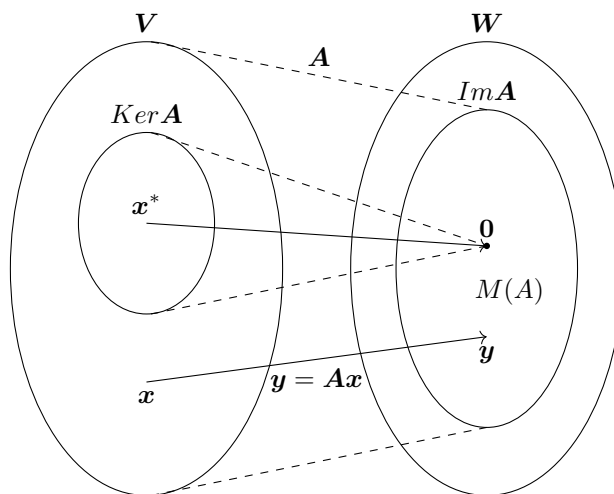


図 5.4 ベクトルによる線形写像

■ランクは行あるいは列の数より小さい 言い換えると, $n \times m$ 行列 A に対して, $\text{rank}(A) \leq n$ かつ $\text{rank}(A) \leq m$ ということになります。

$n \times m$ 行列 A で $n \leq m$ であった場合, 1 次独立なベクトルは n 個までなので $\text{rank}(A) \leq n \leq m$ です。逆に $m < n$ であれば $\text{rank}(A) \leq m < n$ なので, 行数・列数の小さいほう以下の次元数しかありません。ちなみに $\text{rank}(A) = \min(n, m)$ であるばあい, フルランクである, という言い方をします。正方行列 ($n = m$) の場合は $\text{rank}(A) = n$ のときにフルランクであるといい, $\text{rank}(A) < n$ のときに「ランク落ちが生じた」などといいます。

また, 正方行列がフルランクであることは, 逆行列の存在を意味します。すなわち, m 次の正方行列 A に対して, $\text{rank}(A) = m \Leftrightarrow A$ が正則ということです。

正則とは逆行列があるということ, つまり $AA^{-1} = I$ になるような A^{-1} が存在することでした。逆行列があれば方程式が解けますから, $Ax = 0$ のような式があっても両辺に A^{-1} をかけて

$$\begin{aligned} A^{-1}Ax &= A^{-1}0 \\ x &= 0 \end{aligned}$$

すなわち自明な解しかない, ということになります^{*6}。さてこのとき, A を列ベクトル $a_1, a_2, \dots, a_m$ と考えると, $Ax = 0$ というのは

$$a_1x_1 + a_2x_2 + \dots + a_mx_m = 0$$

であり, 自明な解しか持たないので $x_1 = x_2 = \dots = x_m = 0$ であること, つまりこのベクトルの組は線形独立であり, ベクトルの数だけ独立した次元があるのですからフルランクであるということになります。フルランクであること, 全てのベクトルが線形独立であること, 同次連立一次方程式に自明な解しかないこと, 逆行列があること, がつながりました。

■行列を転置してもランク数はかわらない つまり $\text{rank}(A') = \text{rank}(A)$ です。方程式の話の時, ランクとは階段行列に変形して行ったときの段の数, つまり線形独立な行の数ということでした。またこのセクションでは行列を列ベクトルの組として考え, 線形独立な列の数 (空間の基底の数) ともいいました。転置行列と

^{*6} 自明な解ってなんだっけ, というひとはセクション 3.3.3, Pp.52 を参照。係数行列が行ごとに別々の方程式だったとして, それが同じ解を持つかどうか, というお話の流れで, 全部 0 なら当たり前 (自明) だから面白くないよね, というやつでした。

は行と列を入れ替えたものですから、 $\text{rank}(A')$ とは A' に含まれる線形独立な行の数であり、 A に含まれる線形独立な列の数です。つまり、同じことだというわけです。

■行列の積のランクは、積を構成する行列のランク以下 つまり $\text{rank}(AB) \leq A$ かつ $\text{rank}(AB) \leq B$ ということです。行列が写像するもの、座標変換装置であることを思い出すと、 A を使った変換は A 以上の次元になることはありませんね。写像は影を落とすようなものですから、落ちた先の空間以上の次元は持てないのです。 AB は、 B で写像したものを A で写像する、と考えればどちらかのランク以上に広がることはないな、と直感的に理解できるのではないのでしょうか。

丁寧に証明しておきましょう。まず $\text{rank}(AB) \leq A$ からいきます。 $A = (a_1, a_2, \dots, a_n)$ である $m \times n$ 行列、 $B = (b_1, b_2, \dots, b_l)$ である $n \times l$ 行列であるとし、 $C = AB$ とすると、 C のサイズは $m \times l$ になりますから、

$$C = (c_1, c_2, \dots, c_l)$$

と考えることができます。この C の 1 列目のベクトル、 c_1 がどうなっているか、行列の積のプロセスをよく思い出してみてください。最初 (1 行目) の要素 c_{11} は、 $a.$ の各 1 行目要素と b_1 すなわち B の 1 列目の要素をかけて足して行ったものです。 c_{22} は $a.$ の各 2 行目要素と b_1 すなわち B の 1 列目の要素の積和。つまり、 c_{jk} は A の j 行目と B の k 列目の積和になっているわけです。

C の 1 列目のベクトル、 c_1 が $a.$ の線型結合で出来上がっており、全体でも $c_1, c_2, \dots, c_l$ は $a_1, a_2, \dots, a_n$ の線型結合で出来上がっていることがわかります。ということは、 C の列に含まれる線形独立なベクトルの数が、 A に含まれる線形独立なベクトルの数を超えることはありません。この「列に含まれる線形独立なベクトルの数」とはもちろんランクのことですから、 $\text{rank}(C) = \text{rank}(AB) \leq \text{rank}(A)$ となります。

次に $\text{rank}(AB) \leq B$ はどうでしょうか。これは転置行列を考えて、

$$\text{rank}(AB) = \text{rank}(AB') = \text{rank}(B'A')$$

が成り立つので、先の議論から

$$\text{rank}(B'A') \leq \text{rank}(B')$$

が成り立ちます。つまり

$$\text{rank}(AB) \leq \text{rank}(B')$$

であり、転置してもランクは変わらないのですから、

$$\text{rank}(AB) \leq \text{rank}(B)$$

です。ということで、左右の行列いずれものランクを超えることはないことが証明できました。

ちなみにかける行列が正則 (=逆行列がある) なら、次のことも言えます。

$$\text{rank}(AB) = \text{rank}(CA) = \text{rank}(A)$$

ここで B, C は正則です。

先ほどの積の関係から、 $\text{rank}(BA) \leq \text{rank}(A)$ であることは言えますし、 B が逆行列を持っているので

$$\text{rank}(A) = \text{rank}(IA) = \text{rank}(B^{-1}BA)$$

と言えます。また再び、積のランクはかけた行列のランクを超えることはないのです、

$$\text{rank}(B^{-1}BA) \leq \text{rank}(BA)$$

となります。 $\text{rank}(A) \leq \text{rank}(A)$ で $\text{rank}(A) \leq \text{rank}(BA)$ ですから,

$$\text{rank}(BA) = \text{rank}(A)$$

となります。

逆から掛けた場合も同様に,

$$\text{rank}(AC) \leq \text{rank}(A)$$

$$\text{rank}(AC) = \text{rank}(AI) = \text{rank}(ACC^{-1})$$

$$\text{rank}(ACC^{-1}) \leq \text{rank}(AC)$$

$$\text{rank}(AC) = \text{rank}(A)$$

となります。

5.3.2 正射影と射影行列

正射影

さて空間を分解できることがわかりました。任意のベクトル $a \in \mathbb{R}^p$ は互いに直交する部分空間 $W, W^\perp$ に属するベクトル $b \in W, c \in W^\perp$ をつかって

$$a = b + c$$

の形に一意に表すことができます。このとき、 b を a の W への**正射影**といいます^{*7}。

p 次元ベクトルをこのように分解した時、 a に b を対応させる行列 P を考えることができます。話を簡単にするために、二次元空間で考えてみましょう (図 5.5)。射影は a の空間に落とした影のようなものですから、 v が p に変換されたわけです。

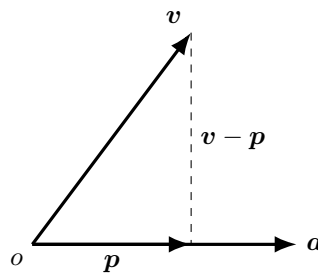


図 5.5 二次元ベクトルを一次元に射影

ここで p は a と同じ向きを持ったベクトルですから、定数倍で $p = ka$ と考えることができます。また垂線の部分、すなわち $v - p$ は a に垂直、すなわち角度が直角、つまり内積 0 ですから、

$$a'(v - p) = a'(v - ka) = 0$$

です。ここから、

$$a'v - ka'a = 0$$

$$k = \frac{a'v}{a'a}$$

^{*7} 同様に c は $W^\perp$ への正射影です。

となります。つまり、ベクトル $p = \frac{a'v}{a'a}a$ であることがわかります。

このようにして射影先のベクトルがわかりますが、これだと a を変換して p にしているようですね。もとは v を変換して p にする、という話だったので、式を少し書き換えましょう。

この k はスカラーですから、 $p = ka = ak$ より

$$\begin{aligned} p &= a \frac{a'v}{a'a} \\ &= \frac{aa'v}{a'a} \\ &= \frac{aa'}{a'a}v \end{aligned}$$

となり、 v に $\frac{aa'}{a'a}$ をかけることで p ができることがわかりました。ここで行列のサイズに注意してほしいのですが、 aa' は 2×1 の列ベクトルと 1×2 の行ベクトルの積なので行列です。分母はスカラーですからサイズに影響はありません。ともかく、行列は座標変換装置であり、射影のはたらきをする行列のことを射影行列といいます。

具体例で考えてみましょう。先ほどの例から、 $a = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ とします。これは x 軸を構成するベクトルですね。これを使って射影行列 P を考えますと、

$$\begin{aligned} P &= \frac{aa'}{a'a} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

となります。この行列を使うと、二次元空間にあるベクトル $v = \begin{pmatrix} x \\ y \end{pmatrix}$ はどんなものでも、

$$Pv = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}$$

となり、 x 軸の要素だけ取り出されていることがわかると思います。

今度は平面に分解することを考えてみましょう。空間イメージは図 5.6 のような感じです。これは 3 次元ベクトルを、それより小さい 2 次元の空間におけるベクトルに分解するしている例です。図ではちょうど上の方向からライトを当てて、平面空間に影を落としたような感じになっていますよね。

今回 p は b_1 と b_2 が作る空間のベクトルですから、

$$p = \alpha_1 b_1 + \alpha_2 b_2$$

という線型結合で表されることになります。 b はベクトルですから、

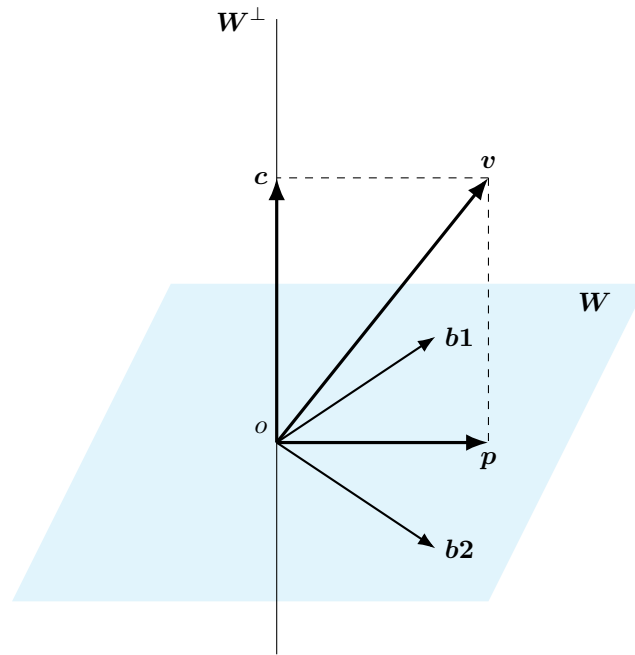
$$p = [b_1 \quad b_2] \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = Ax$$

のように表せます。

今回もベクトル v から下ろした垂線と、 b_1, b_2 はそれぞれ垂直、内積が 0 なので、

$$b_1'(v - Ax) = 0$$

$$b_2'(v - Ax) = 0$$

図 5.6 $\mathbf{b}$ が作る空間 (2 次元平面) への射影と $\mathbf{c}$ がつくるその直交補空間

という条件のもとで α_1, α_2 を求めることになります。ここで

$$\mathbf{A}' = \begin{pmatrix} b'_1 \\ b'_2 \end{pmatrix}$$

ですから、まとめて

$$\mathbf{A}'(\mathbf{v} - \mathbf{A}\mathbf{x}) = \mathbf{0}$$

としましょう。

$$\mathbf{A}'\mathbf{v} - \mathbf{A}'\mathbf{A}\mathbf{x} = \mathbf{0}$$

$$\mathbf{A}'\mathbf{A}\mathbf{x} = \mathbf{A}'\mathbf{v}$$

であり、 $\mathbf{A}'\mathbf{A}$ は正方行列です。この正方行列を構成していた要素 b_1, b_2 が基底であれば一次独立で、フルランクになりますから逆行列が存在します。それを使って、

$$\mathbf{x} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{v}$$

が得られます。正射影ベクトル $\mathbf{p}$ は

$$\mathbf{p} = \mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{v}$$

ということですから、射影行列は

$$\mathbf{P} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$$

だということになります。

具体例で見てみましょう。 $\mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ と考えると、これは x, y 軸が張る二次元平面になりますね。ここに三次元ベクトル $\mathbf{a} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ から射影することを考えると

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$P = A(A'A)^{-1}A' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

となりますから,

$$Pa = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

です。つまり z 軸方向の次元がなくなって、 x, y 平面に影を落とすことになるわけです。

ちなみに W に直行する空間 $W^\perp$ に属するベクトル $c \in W^\perp$ とともに

$$a = b + c$$

と表せるということでしたが、今回 c は z 軸方向の 1 次元ということになりますね。

一般に、 W への射影行列を P_W 、 $W^\perp$ への射影行列を $P_{W^\perp}$ と表したとします。つまり

$$b = P_W a \in W$$

$$c = P_{W^\perp} a \in W^\perp$$

ですが、このとき空間のランクを考えると、 P は $p \times p$ 行列で p 次元ある行列ということになります。元の空間 W の次元を $\dim(W) = k$ とすると、 $\dim(W^\perp) = p - k$ という関係が成り立ちます。

射影行列

射影行列を一般的に定義すると、 $P^2 = P$ を満たす正方行列、ということになります。 P^2 は PP ということであり、また P はその正方行列の持つ基底空間以外の次元を削ぎ落とすことを意味するのです。何回削ぎ落としても、その射影行列の次元分は維持されることでもって、射影行列は定義されるのです。

ちなみに正射影は直交射影ともいい、 P が特に対称行列であるときのことを指します。先ほどの例にあった

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

は対称行列です。定義通り、

$$P^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}^2 = P$$

という関係が成り立っていますね。

直交でない例は、非対称行列を考えれば良いのですが、例えば

$$P = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

は $P^2 = P$ なので射影行列です。これがどんな射影をするかというと、

$$P \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + y \\ 0 \end{pmatrix}$$

ですから、図にすると図 5.7 のようになります。確かに y 軸の要素をなくすような射影になっているのですが、影の落とし方が斜めからになっているわけですね。

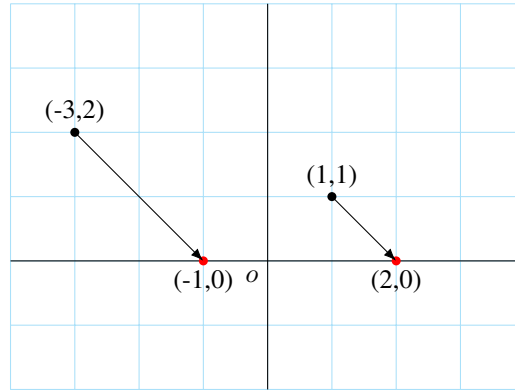


図 5.7 斜交射影の例

■射影行列の相補性 射影行列 P があると、単位行列から P を引いた行列 $I - P$ もまた射影行列になります。特に P が直交行列であれば、 $I - P$ も直交行列です。この性質のことを射影行列の相補性と言います。

このことは次のように、簡単に確かめられます。

$$(I - P)^2 = I^2 - 2P + P^2 = I - 2P + P = I - P$$

また

$$(I - P)' = I' - P' = I - P$$

■直和分解 このように、射影行列 P があれば $I - P$ もすぐにわかりますから、任意のベクトル a を Pa と $(I - P)a$ に分けることができます。これはまた空間の像、核の言葉を使っていうと次のようになります。 $I - P = Q$ と表すとして、

$$\mathbb{R}^p = \text{Im}P \oplus \text{Im}Q$$

また

$$\text{Ker}P = \text{Im}Q$$

であり、

$$\mathbb{R}^p = \text{Im}P \oplus \text{Ker}P$$

でもあります。

P の核が Q の像に等しいというのは、まず核のことを思い出さねばなりません。核とは $Px = 0$ になるような、つぶれてしまうところの話です。ここで $x = Px + Qx$ のように分解できるのですから、 Px が核であるなら $x = Qx = \text{Im}Q$ というのと同じです。

直和であることを示すために、 $\text{Im}P \cap \text{Ker}P = 0$ であることを導きましょう。まず射影された空間は部分空間なので、 $\text{Im}P \cap \text{Ker}P \supset 0$ です。またこの集合 $\text{Im}P \cap \text{Ker}P$ の要素を x とすると、像 $\text{Im}P$ をつくったもとなるもの、つまり $x = Py$ となる y が存在します。またこの y は $\text{Ker}P$ の要素でもあるので、 P を作用させて $P(Py) = P(x) = 0$ です。射影行列の定義からいって $P^2 = P$ なので、 $P(Py) = Py = 0$ であり、この要素 $x = Py = 0$ より、 $\text{Im}P \cap \text{Ker}P = 0$ となります。

さて、ベクトルを空間、行列を座標変換装置とみる、ということと、それにまつわる数学的なお話は一旦ここで終わりにしておきましょう。次からは、こうした特徴や考え方を、心理統計の文脈でどのように扱うのかという話に入っていきます。

第 6 章

データ空間

ベクトルを空間的なイメージとして理解できたところで、これをデータ解析に応用していきましょう。私たちが考えたいのはデータのベクトルであり、データの空間だったわけです。これに具体的なモデルを入れて考えていきましょう。

6.1 線形回帰を行列と空間で理解する

セクション 2.3(Pp.33) でみたように、回帰分析はデータベクトルと係数ベクトルを使って、

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

と表せるのでした。具体例として、まずは説明変数が 2 つある重回帰分析を考えてみましょう。予測値 $\hat{y}_i$ は次のように表現できます。

$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

これを空間のイメージで考えると、右辺は説明変数 $\mathbf{x} = (\mathbf{1}, x_1, x_2) \in \mathbb{R}^3$ に係数ベクトル $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' \in \mathbb{R}^3$ をかけてたことによる空間が広がっていますね。 $\mathbf{X}$ が張る空間を $W(\mathbf{X})$ とすると、次のように表すことができます。

$$W(\mathbf{X}) = \{\beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 \mid \boldsymbol{\beta} \in \mathbb{R}^3\}$$

$\hat{\mathbf{y}}$ はこの空間の中にあるベクトルで、実際のデータベクトル $\mathbf{y}$ はこの右辺が張る空間に射影されたものとみることができます (図 6.1)。

セクション 5.3.2(Pp.81) でみたように、射影行列 $\mathbf{P}$ は基底ベクトルを使って $\mathbf{P} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ と表されるということでしたから、回帰分析の場合は

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

として求められることになります。

ちなみに誤差 $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ は、被説明変数 $\mathbf{y}$ から空間 $W(\mathbf{X})$ に射影した時の垂線の長さです。垂線ですから、説明変数が張る空間とは直角に交わっています。すなわち、誤差は説明変数との相関がゼロである、という回帰分析の特徴も一気に理解することができます。

また垂線ですので、被説明変数ベクトルと説明変数が張る空間との距離が最短になっていることもわかります。一応証明しておきましょうか。平面 $W(\mathbf{X})$ 上の別のベクトル $\mathbf{P}$ を考えて $\mathbf{P}$ と $\mathbf{y}$ との距離を考えるとします。

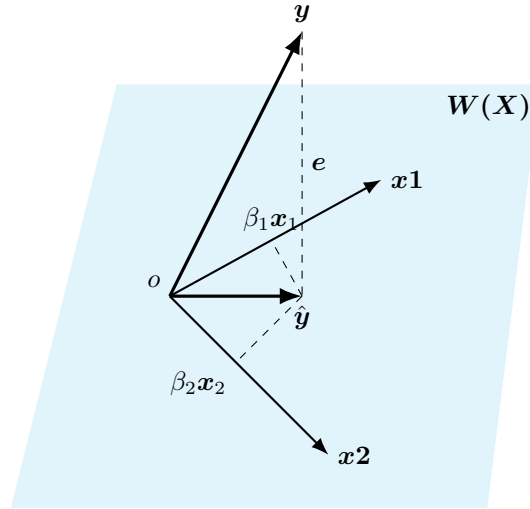


図 6.1 説明変数の空間に被説明変数ベクトルを射影する

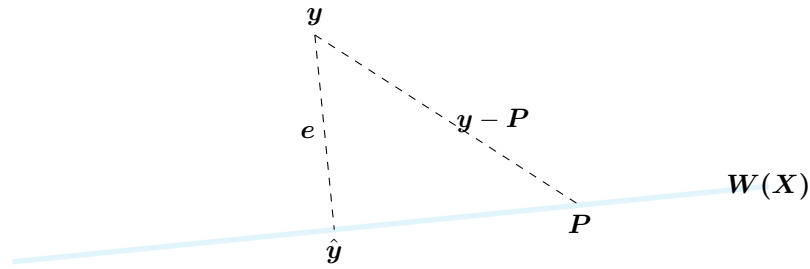


図 6.2 垂線が最短距離であることを考える

距離の式を展開していくと、次のような関係が得られます。

$$\begin{aligned}
 \|y - P\|^2 &= \|y - \hat{y} + P - \hat{y}\|^2 \\
 &= \|y - \hat{y}\|^2 + \|P - \hat{y}\|^2 + 2e'(P - \hat{y}) \\
 &= \|e\|^2 + \|P - \hat{y}\|^2
 \end{aligned}$$

つまり、別のベクトル P との距離はどうやっても $e = y - \hat{y}$ 以上になるので、 e が最短だということになります。この e のノルムを最小にするのが**最小二乗法**ですが、距離を最小にするので二乗なんだな、と考えることができます。

具体的な係数ベクトル β は、空間 $W(X)$ と残差 $e = y - \hat{y}$ が直交することから、

$$X'e = 0$$

$$X'(y - X\beta) = 0$$

$$X'y - X'X\beta = 0$$

$$X'X\beta = X'y$$

となりますから、 β だけにするには $X'X$ に逆行列があればよく、

$$\beta = (X'X)^{-1}X'y$$

ということになります。そもそも $\hat{y} = X\beta$ だったわけですから、射影行列との関係 $\hat{y} = X(X'X)^{-1}X'y$ とも整合的であることがわかりますね。

6.1.1 多重共線性の問題

重回帰分析において**多重共線性 (multicollinearity)** に気をつけなければならない、というのは聞いたことがあると思います。

多重共線性は、説明変数同士の相関が高いことが原因でおこることで、なんか良くないんだ、というイメージはお持ちかと思います。さて何が悪いのでしょうか？これを考えるときに、空間的なイメージからアプローチしてみましょう。説明変数同士の相関が高いこと、というのがありましたが、相関が高いことを空間のイメージでいうと、ベクトルの $\cos \theta$ が 0 ではないということです。相関が最高潮に高い場合は、ベクトルがぴたりと重なってしまいますから、これは一次独立でなくなり、空間が潰れてしまうことになります。

今回の例では 2 つの説明変数が作る空間があって、そこに被説明変数ベクトルの影を落としたのでした。図 6.1 にもあるように、説明変数ベクトルの方向を何倍するかという、倍率が偏回帰係数になります。このとき、2 つの説明変数が実は一次従属関係にあったとすると、2 本のベクトルではなく 1 本のベクトル空間になります。そんな状況で 2 つの回帰係数を求めることはできないですね。

重回帰分析が実践的に使われるシーンでは、2 つと言わずもっと多くの説明変数が使われることがあります。 n 人分のデータ、 p 個の説明変数があつたとして、行列 X は $n \times (p + 1)$ になりますが、説明変数のなかに一次従属な変数があるとこの行列 X のランクが $\text{rank}(X) < (p + 1)$ になっていることになります^{*1}。行列の積とランクの関係から、

$$\text{rank}(X'X) \leq \text{rank}(X)$$

ですが、 $p + 1 \times p + 1$ の正方行列 $X'X$ が逆行列を持つ条件は $\text{rank}(X'X) = p + 1$ であることで、ランク落ち $\text{rank}(X'X) < p + 1$ が生じていると逆行列が持てません。行列 $(X'X)$ は射影するときに逆行列にして使いますので、ここで計算上のエラーが生じます。

逆行列の計算は、行列のサイズが大きくなっていくと非常に計算が大変になって、近似的に解いていくことになります。解析的には答えがないはずの行列に対して、無理矢理近似的な数値を算出させたところで、その数字が信用できるはずもないことはおわかりいただけると思います。もちろん実際のデータで、変数同士が一次従属になることは稀です^{*2}が、尺度の合計得点と因子得点といった、相関がごく高い変数の両方を使用変数にしてしまうと、ランク落ちの様相を呈してきます。もちろん偽相関など見せかけ的に高い関係があってもいけません。

R でやってみて考える

実際に数値例をつかってみてみればいいのですが、逆行列の計算は面倒なので、R をつかって計算させてみたいと思います。R で逆行列を求めるには、`solve` という関数があるので、これを使います。

code : 6.1 フルランクの行列

```
1 A <- matrix(c(1,3,5,2,4,3,3,1,2,2,1,6,3,5,4),ncol=3)
2 M <- t(A) %*% A
3 solve(M)
```

^{*1} 行列のランクは行数、列数の小さい方であり、一般にデータ数 n のほうが変数の数 p よりも大きい ($p \ll n$) ので、ここでのランクは $p + 1$ で考えます。ちなみに $\ll$ は「十分に小さい」という意味です。

^{*2} 変数を生成して説明変数に加えると、こうした問題が生じがちです。例えば身長を単位を cm から m に変えて両方とも投入してしまうといった人為的なミスというケースです。

まず適当な行列 A を作り, $M = A'A$ を構成します。今回の例では,

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 3 & 3 & 6 \\ 5 & 1 & 3 \\ 2 & 2 & 5 \\ 4 & 2 & 2 \end{pmatrix}, M = \begin{pmatrix} 55 & 29 & 60 \\ 29 & 27 & 42 \\ 60 & 42 & 87 \end{pmatrix}$$

であり, 逆行列は

$$M^{-1} = \begin{pmatrix} 0.0734 & -0.0003 & -0.0504 \\ -0.0003 & 0.1487 & -0.0715 \\ -0.0504 & -0.0715 & 0.0808 \end{pmatrix}$$

となりました。

ここでわざとランク落ちを生じさせます。 A の 3 列目を, A の 2 列目で上書きしてしまうのです。

code : 6.2 ランク落ちの行列

```
1 A[,3] <- A[,2]
2 M <- t(A) %*% A
3 solve(M)
```

こうすると, R はエラーを返します。

`solve.default(M)` でエラー:

Lapack routine dgesv: システムは正確に特異です: U[3,3] = 0

だそうです。

さて, ではランク落ちかけ, という程度ならどうでしょう。 A の 3 列目を A の 2 列目で上書きするときに, 少し数字をいじってしまいます。

code : 6.3 ランクおちかけ

```
1 A[,3] <- A[,2]+rnorm(5)*0.001
2 M <- t(A) %*% A
3 solve(M)
```

今回は微小な誤差を追加しましたので, 行列は

$$A = \begin{pmatrix} 1 & 3 & 3.0005855 \\ 3 & 3 & 3.0007095 \\ 5 & 1 & 0.9998907 \\ 2 & 2 & 1.9995465 \\ 4 & 2 & 2.0006059 \end{pmatrix}, M = \begin{pmatrix} 55.00000 & 29.00000 & 29.00368 \\ 29.00000 & 27.00000 & 27.00408 \\ 29.00368 & 27.00408 & 27.00816 \end{pmatrix}$$

です。ほとんど変わらない大きさですが, この逆行列は,

$$M^{-1} = \begin{pmatrix} 0.0430067 & -36.95954 & 36.90777 \\ -36.9595434 & 1260304 & -1260074 \\ 36.9077732 & -1260074 & 1259844 \end{pmatrix}$$

という無茶苦茶な大きさになってしまいました。

多重共線性が問題になるシーンは, 計算がうまくいっていないかも, という数値計算の話だけではありません。データ解析の多くのシーンでは, 求めたい回帰係数は母集団のそれであり, 毎回のサンプルごとどれぐら

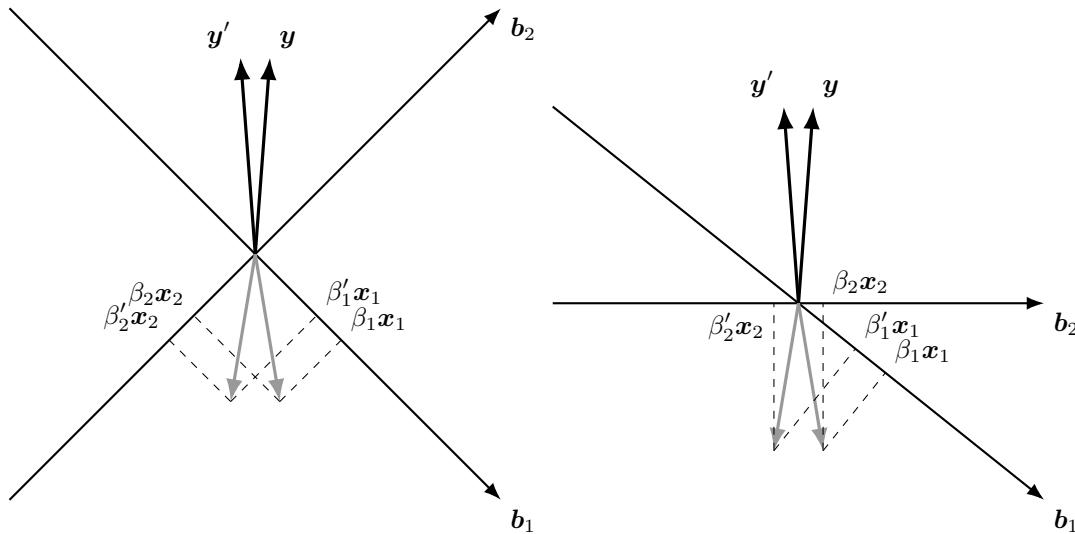


図 6.3 左; サンプルが変わると少し係数がぶれる。右; サンプルが変わっただけで β_2 の符号も変わってしまう。

い変わるかということをも、標準誤差で表現します。多重共線性が生じて計算上の不確かさがあると、これが標準誤差の大きさに跳ね返ってきます。

図 6.3 には、被説明変数が y と y' の 2 本描かれています。違うサンプルを取ったことで少し違う値が得られたというイメージです^{*3}。説明変数ベクトルが張る空間に落とした影から、説明変数ベクトルに垂線を引いたところが係数になります。右図にあるように、説明変数同士に相関があると、ちょっと影がずれるだけで係数の符号まで変わってしまいます。このような不安定さが簡単に生じるので、多重共線性があった場合は係数が信用ならないわけです。

6.1.2 分散分析による空間の分解

線形モデルの特殊例として、分散分析を考えることができます。分散分析はデータ全体の平方和 (Sum of Squares) を、効果の平方和と誤差の平方和に分解するもので、誤差に比べて効果がどれだけ大きいかを判断するものです^{*4}。これを行列と空間のメタファで考えてみたいと思います。

分散分析は被験者を実験デザインに割り当てて、データを取ります。この「デザインに割り当て」というのは、デザイン行列がつくる空間にデータを射影する、と考えることができます。たとえば実験群・統制群それぞれに 2 人ずつ割り当てて、データ $y' = \begin{pmatrix} 2 & 3 & 4 & 5 \end{pmatrix}$ を得たとしましょう。この時のデザイン行列は、

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

と書くことができます。この行列 X が張る空間に、データ Y を射影します。射影行列は

$$P_A = X(X'X)^{-1}X'$$

になるのです。データ全体 $\mathbb{R}$ は、この射影によって部分空間 $W(X)$ と $W(X)^\perp$ に分解されます。相補性

^{*3} この素晴らしい幾何学的な説明のアイデアは、南風原 (2002) を参考にしました。

^{*4} 平方和をデータの数で割れば分散になりますが、分散分析の場合は自由度 (degrees of freedom) でわって平均平方 (Mean Squares) にし、その比が F 分布に従うことから有意性の有無を判定する、というのが全体の流れでしたね。

の考え方から $I - P_A$ が $W(X)^\perp$ に対応しています。実験計画の場合、デザイン行列で把握しきれない変動は誤差だと考えるわけですから、 $W(X)^\perp$ が誤差ベクトルの作る部分空間であり、そこに射影する射影行列は $P_e = I - P_A$ だと考えることができます。

具体的な数値例で見てみます。射影行列は、

$$\begin{aligned}
 P_A &= X(X'X)^{-1}X' \\
 &= \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \left(\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}
 \end{aligned}$$

となります^{*5}。これの直交補空間への射影行列は、

$$\begin{aligned}
 P_e &= I - P_A \\
 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix} \\
 &= \begin{pmatrix} 0.5 & -0.5 & 0 & 0 \\ -0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & -0.5 & 0.5 \end{pmatrix}
 \end{aligned}$$

となります。

データベクトルをこれらの空間に射影するとどうことが起きるか、計算してみましょう。まずはデザイン行列の空間へ射影しましょう。

$$P_A y = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 2.5 \\ 2.5 \\ 4.5 \\ 4.5 \end{pmatrix}$$

ここにあるように、実験群・統制群それぞれへの平均値が計算されました。また、 P_e を使った射影の結果は次のようになります。

^{*5} 2×2 の逆行列は、セクション 3.3.1, Pp.45 を思い出してください。

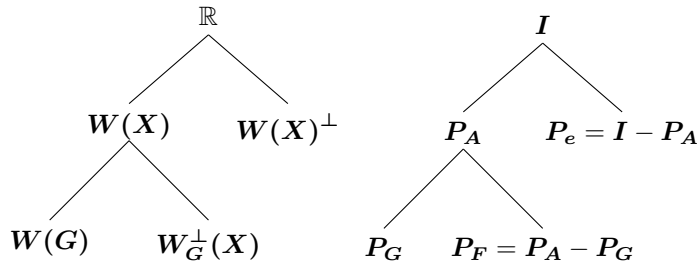


図 6.4 分散分析による空間の分解

$$P_E y = \begin{pmatrix} 0.5 & -0.5 & 0 & 0 \\ -0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & -0.5 & 0.5 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} = \begin{pmatrix} -0.5 \\ 0.5 \\ -0.5 \\ 0.5 \end{pmatrix}$$

これは一瞬分かりにくいかもしれませんが、それぞれの群平均からの差を表していますね。群平均からの個々の値の差を誤差と考えるわけですから、まさにその通りになっています。

さて今回のデザイン行列には、全体平均 (切片) が含まれていませんでしたが、デザイン行列が群平均を取り出す写像であることを考えると、その空間の中に全体平均の要素も含まれているはずです。つまり $W(X)$ はさらに分割することができて、全体平均 $W(G)$ と、 $W(X)$ の中であって $W(G)$ でない部分、 $W_G^\perp(X)$ に分解できるのです。

全体平均は全て同じ大きさを持つベクトル $G' = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}$ が張る空間だと考えることができます。この射影行列を考えます。

$$\begin{aligned} P_G &= G(G'G)^{-1}G' \\ &= \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix} \end{aligned}$$

ここにデータを射影してみると、

$$P_G y = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 3.5 \\ 3.5 \\ 3.5 \\ 3.5 \end{pmatrix}$$

となります。

図 6.4 には、空間の分解と射影行列を併記してみました。左図はデータ空間の分解で、右図はそれに対応する射影行列の分解パターンです。そこには要因の効果を取り出す射影行列 P_F がありますが、これは誤差を取り除いた空間から、全体平均を写像で潰した残りの空間として表現されています。射影行列の関係から、

$$I = P_G + P_F + P_e$$

であることがわかります。

さて、データ全体の平方和は $y'y = y'Iy$ と書くことができますから、

$$y'Iy = y'P_G y + y'P_F y + y'P_e y$$

という分解ができることがわかります。

表 6.1 射影行列による分散分析表の構成

source	SS	df	MS
Factor	$\mathbf{y}'\mathbf{P}_F\mathbf{y}$	$\text{tr}(\mathbf{P}_F)$	$\mathbf{y}'\mathbf{P}_F\mathbf{y} / \text{tr}(\mathbf{P}_F)$
Error	$\mathbf{y}'\mathbf{P}_e\mathbf{y}$	$\text{tr}(\mathbf{P}_e)$	$\mathbf{y}'\mathbf{P}_e\mathbf{y} / \text{tr}(\mathbf{P}_e)$
Total	$\mathbf{y}'\mathbf{I}\mathbf{y} - \mathbf{y}'\mathbf{P}_G\mathbf{y}$		

これを踏まえて考えると、分散分析表の構成は、次のように表現できます。

表 6.1 中に、 $\text{tr}(\mathbf{A})$ というのがありますが、これは**トレース (trace)** という行列の演算で、正方行列 $\mathbf{A} = a_{ij}$ に対して

$$\text{tr}(\mathbf{A}) = \sum a_{ii}$$

として計算される値です。対角成分の総和ですね。

分散分析は分散を分解していく、つまり足し算の形に分離変形をしていくことなのですが、空間で説明するなら部分空間への分割であるともいえるのです。

またデザイン行列 (セクション 2.4, Pp.34) のところで制約条件について触れましたが、そのこととこの射影行列がつながってきます (Pp.*9 の脚注にかいた伏線の回収です!)。デザイン行列 $\mathbf{X}$ に、はじめから切片をいれ、また制約条件を入れない書き方をしてみましょう。すなわち、

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

のように考えたとします。このとき、計算途中で含まれる逆行列の計算、 $(\mathbf{X}'\mathbf{X})^{-1}$ が計算不能になってしまうのです。この行列を掃き出し法などで計算すると分かりますが、階数が 2 しかないのです。つまり作れる空間は 2 次元までであり、切片をそのひとつと数えるのであれば、残りは効果があるかないかという要素しか残らないからです。

例題 18

4 人の被験者を A,B 群に 2 人ずつ振り分けて、群間差を見る分散分析をしたいと思います。データベクトル $\mathbf{y} = (1 \ 3 \ 4 \ 6)'$ とデザイン行列の関係は、

$$\begin{pmatrix} 1 \\ 3 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}$$

というようにかけるでしょう。ここから、 $\mathbf{P}_G, \mathbf{P}_F, \mathbf{P}_e$ を構成し、また分散分析表を計算してみましょう。

6.2 分散共分散行列

セクション 2.2(Pp.27) でみたように、データ行列 $\mathbf{X}$ を変形して平均偏差行列 $\mathbf{V}$ を作ることができるのでした。例えば 2 変数の場合、

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix}$$

に対して

$$\begin{aligned} \mathbf{V} &= \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) \mathbf{X} \\ &= \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 \\ \vdots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 \end{pmatrix} \end{aligned}$$

という計算ができます。皆さんはすでにこれを見ると、2 次元ベクトル空間のイメージができていないかと思います。各点 x_i までの距離は $d_i^2 = (x_i - \bar{x}_1)^2 + (x_i - \bar{x}_2)^2$ ですから、全ての点に対してこれを総和したもの、すなわち

$$\sum_i d_i^2 = \sum_i ((x_i - \bar{x}_1)^2 + (x_i - \bar{x}_2)^2)$$

はそれぞれの分散の n 倍を足し合わせたものだということをすることができます。図 6.5 にはこの全員分の分散 (の n 倍) を薄く色付けして表現しています。

行列で考えると、 $\frac{1}{n} \mathbf{V}'\mathbf{V}$ が分散共分散行列を表しますから、

$$\sum_i d_i^2 = \text{tr}(\mathbf{V}'\mathbf{V})$$

とすることができます。

さて、これをベクトルや行列の空間的イメージとして見ると、原点さえ同じであれば必ずしもそれぞれの変数を作る空間にこだわらなくとも、分散の総和 (を n 倍したもの) は同じであることがわかりますね。大事なのはこの空間上にあるデータの散らばりであって、座標 (基底) はどのようにとっても良いことです。

ここで、ある行列 $\mathbf{T}$ を考えます。これは次のような要素からなります。

$$\mathbf{T} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

行列は座標変換装置なので、この行列も元の行列の座標を変えてしまいます。しかし、各ベクトルのノルムは 1 で、しかも 2 つの列ベクトルとして見ると互いに直交していることがわかります。確認してみましょう。

$$\left\| \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \right\|^2 = (\cos \theta)^2 + (\sin \theta)^2 = \cos^2 \theta + \sin^2 \theta = 1$$

$$\left\| \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \right\|^2 = (-\sin \theta)^2 + (\cos \theta)^2 = \sin^2 \theta + \cos^2 \theta = 1$$

$$\left(\begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}, \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \right) = \cos \theta \times (-\sin \theta) + \sin \theta \times \cos \theta = 0$$

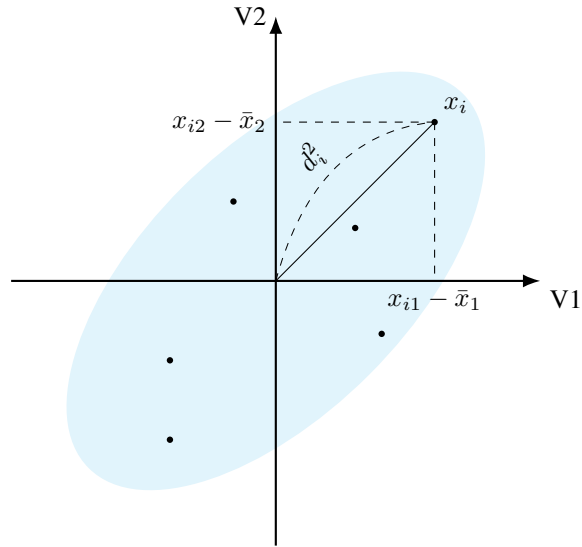


図 6.5 平均偏差行列をつかった空間

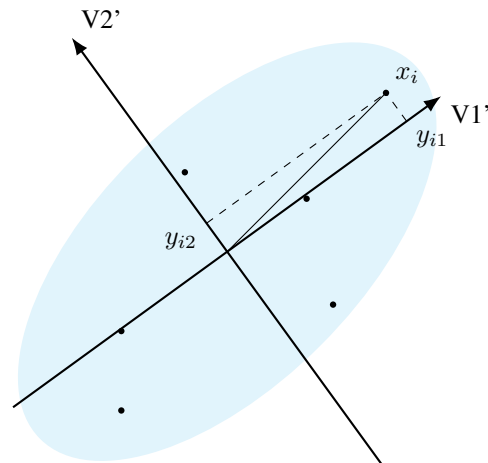


図 6.6 座標を回転させる

このように、直交している次元で、かつノルムを変えない変換をする行列のことを**直交行列**といい、特に回転角度 θ で座標を変えるものを**回転行列**といいます。

直交行列はノルムが 1 で互いに直交しますので、 $T'T = I$ であり、 T の逆行列が転置行列になっているようなものです。回転行列で考えると、逆行列はちょうど逆回転の操作に対応します。

さて分散の総和 (を n 倍したもの) という特徴は変えずに、座標だけ切り替えるためにこの行列を使います。適当な角度 θ を決めて、もともっている座標をグイッと回転させたところで、各データ点の原点からの距離に変化はありません。つまり、データ行列全体の持つ情報の総量は変化していないのです。新しくなった座標

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = T \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}$$

をつかって表現すると、

$$d_i^2 = (x_{i1} - \bar{x}_1)^2 + (x_{i2} - \bar{x}_2)^2 = (y_{i1} - \bar{y}_1)^2 + (y_{i2} - \bar{y}_2)^2$$

であり、新しい第一次元の分散 $(y_{i1} - \bar{y}_1)^2$ がもとの第一変数の分散よりも大きく、新しい第二次元の分散

$(y_{i2} - \bar{y}_2)^2$ はもとの第二変数の分散よりも小さくなるように、ウェイトが再配分されているわけです。

もともとは、第一変数と第二変数のどちらが大事か、ということは決まっていませんでしたから、それぞれの変数の重みは $1 : 1$ でしたが、より強調が効いている 2 つの重み $w_1 : w_2$ の比率 ($w_1 > w_2$) で表現しなおしたわけです。

どうしてそんなことをするのかって? なぜなら新しく作る第一次元の重みが圧倒的に大きければ ($0.7 : 0.3$ とか $0.9 : 0.1$ とか $0.99 : 0.01$ とか...!) とか、もはや第二次元のことを無視して第一次元だけで考えても本質的には変わりませんよね。もちろん $0.99 : 0.01$ のような超効率の良い回転はできないと思いますが (もしできるとしたらもとの 2 つの変数はほとんど同じ情報しか持っていないことになります), もっとも効率が良くなるような回転を探す, というのはありそうです。

実はこの「もっとも情報圧縮効率の良い軸を探す」ということが、多変量解析における潜在変数の見つけ方につながってくるのです。ということで、空間のイメージと多変量解析の世界へ進んでいくことになります。

第 III 部

Day 3: 行列と多変量解析

第 7 章

固有値と固有ベクトル

7.1 固有値と固有ベクトル

今回は正方行列にみられるおもしろい特徴である、固有値 (eigenvalue) と固有ベクトル (eigenvector) についての話から始めます。ある正方行列 A 、列ベクトル x 、スカラー λ が次のような関係にあった時、 λ を固有値、 x を固有ベクトルと言います。

$$Ax = \lambda x$$

一見すると、 x が両辺に入っていますから、 A が λ に置き換わった等式のようにです。しかし一方は行列で、他方はスカラーです。こんな奇妙なことが本当にあるのでしょうか？具体的な数値例をみてみましょう。

$$A = \begin{pmatrix} 1 & 6 \\ 2 & 5 \end{pmatrix}, x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

を例にします。この時次の関係が成り立ちます。

$$Ax = \begin{pmatrix} 1 & 6 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 7 \end{pmatrix} = 7 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 7x$$

確かに成立する組み合わせがありますね。この行列 A に対して、7 が固有値、 $(1, 1)$ が固有ベクトルになっています。また、実はこの行列 A については、-1 も固有値であり、そのときの固有ベクトルは $(-3, 1)$ になります。

この固有値分解こそ、因子分析を元とする多変量解析の中心的な数学原理なのです。多変量解析の世界においては、分散共分散行列やそれを標準化した相関行列など、変数同士の関係を分析のスタートにおくのでした。これらの行列は正方行列ですから、その固有値や固有ベクトルを計算することで正方行列の特徴を別の視点から分解して考えられるようになります。

7.1.1 固有値と固有ベクトルを求める

では正方行列からどのようにして固有値と固有ベクトルを見つけ出したら良いのか、みていきましょう。元の式を書き換えて次のような方程式を考えます。

$$(A - \lambda I)x = 0$$

固有ベクトルは $x = 0$ すなわち全部ゼロであれば当然成り立ちますから (自明な解)、これは除外することにします ($x \neq 0$)。

さて行列の表現は連立方程式の解を求めることと同じなものでした。 $A - \lambda I$ を連立方程式の係数行列だと考えれば、それが**逆行列**を持つと左辺にそれをかけてしまえば全部ゼロの答えになってしまいますから、そうでない答えを求めるには、この係数行列が逆行列を持たないことが重要です。そして逆行列を持たないようにするというのはつまり、行列式がゼロであれば良いことになります。

2×2 の係数行列、 $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ の逆行列は $ad - bc$ で求められるのでした。 $ad - bc$ が 0 でなければ方程式は解けるのですが、今回の場合は解けると自明になってしまうので困ります。今回は $ad - bc = 0$ でなければならないのです。つまり一般的に書くと次のようになります。

$$|A - \lambda I| = 0$$

$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ とすると、この式は次のようになります。

$$\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = 0$$

$$(a - \lambda)(d - \lambda) - bc = 0$$

この方程式をとくに**固有方程式**といいます。これを解いてやれば良いことになりますね。具体的に $\begin{pmatrix} 1 & 6 \\ 2 & 5 \end{pmatrix}$ の例で計算してみましょう。

$$\begin{vmatrix} 1 - \lambda & 6 \\ 2 & 5 - \lambda \end{vmatrix} = 0$$

$$(1 - \lambda)(5 - \lambda) - 12 = 0$$

$$\lambda^2 - 6\lambda - 7 = 0$$

$$(\lambda - 7)(\lambda + 1) = 0$$

ここから $\lambda = 7, -1$ が得られますね。

では固有ベクトルはどうなるでしょうか。固有値 7 の例で計算してみます。

$$\begin{pmatrix} 1 & 6 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 7 \begin{pmatrix} x \\ y \end{pmatrix}$$

$$x + 6y = 7x \rightarrow 6x = 6y \rightarrow x = y$$

$$2x + 5y = 7y \rightarrow 2x = 2y \rightarrow x = y$$

あれっ？ なんじゃこりゃ？ と思った人もいるかもしれません。でもこれ、間違いではありません。実は固有ベクトルは大きさが定まっておらず、今回の例で言えば $x = y$ つまり $x : y = 1 : 1$ の関係であればあらゆる値が成立してしまうのです。固有ベクトルが $(1, 1)$ や $(2, 2)$, $(100, 100)$ などであっても、 $Ax = \lambda x$ の関係に影響しませんから、計算機で算出する場合などは、ベクトルのノルムを 1.0 に規格化するという条件を追加することが一般的です。

例題 19

行列 $\begin{pmatrix} 8 & 1 \\ 4 & 5 \end{pmatrix}$ の固有値・固有ベクトルを求めてみましょう。

さてここでみたように、 2×2 の方程式であれば手計算でも固有方程式を解くことはできるのですが、行列のサイズがどんどん大きくなると一般的に解けなくなっていくことは想像にかたくないと思います。実際我々は正方行列として、項目の情報が詰まった分散共分散行列とか相関行列を使いますから、それが2項目しかないなんてことはなくて、もっともっと大きなサイズになります。そうすると計算機を使って近似的に答えを求めて行くことになります。

7.1.2 固有値と固有ベクトルの性質

ではここで、 n 次正方行列 A の固有値と固有ベクトルに関する性質をみていきましょう。

■固有値の重複度 固有値は固有方程式から求められますが、固有方程式は n 次方程式であり、解は n 個あることになります。これは代数学の基本定理^{*1}が示すことでもあります。

ただし、同じ解が重複して出ること(重解)もあることに注意です。正確には「 n 次正方行列の固有ベクトルは、重複度を入れて n 個ある」ということです。

■固有ベクトルは一次独立 行列 A の異なる固有値に対する固有ベクトルは、一次独立です。証明してみましょう。 A の固有値 $\lambda_1, \lambda_2, \dots, \lambda_p$ がそれぞれ異なり、対応する固有ベクトルが $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ だったとします。もし $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ が一次独立でないとすると、あるベクトル $\mathbf{x}_i$ はほかの固有ベクトルの一次結合で表せることになりますね。たとえば $\mathbf{x}_1$ が残りの一次独立な固有ベクトル $\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_p$ で表せたとします。つまり、

$$\mathbf{x}_1 = c_1 \mathbf{x}_2 + c_2 \mathbf{x}_3 + \dots + c_{p-1} \mathbf{x}_p$$

ということですね。これに行列 A を左からかけると、

$$A\mathbf{x}_1 = A(c_1 \mathbf{x}_2 + c_2 \mathbf{x}_3 + \dots + c_{p-1} \mathbf{x}_p)$$

より

$$\begin{aligned} \lambda_1 \mathbf{x}_1 &= c_1 A\mathbf{x}_2 + c_2 A\mathbf{x}_3 + \dots + c_{p-1} A\mathbf{x}_p \\ &= c_1 \lambda_2 \mathbf{x}_2 + c_2 \lambda_3 \mathbf{x}_3 + \dots + c_{p-1} \lambda_p \mathbf{x}_p \end{aligned}$$

となります。さて、一次結合の式に λ_1 をかけてやると

$$\lambda_1 \mathbf{x}_1 = c_1 \lambda_1 \mathbf{x}_2 + c_2 \lambda_1 \mathbf{x}_3 + \dots + c_{p-1} \lambda_1 \mathbf{x}_p$$

ですから、両辺それぞれ引き算すれば、

$$\mathbf{0} = c_1(\lambda_2 - \lambda_1)\mathbf{x}_2 + c_2(\lambda_3 - \lambda_1)\mathbf{x}_3 + \dots + c_{p-1}(\lambda_p - \lambda_1)\mathbf{x}_p$$

となります。固有値はすべて異なるということでしたから、 $c_1 = c_2 = \dots = c_{p-1} = 0$ となります。さてそうすると、 $\mathbf{x}_1 = \mathbf{0}$ となってしまいますが、固有ベクトルはゼロではないのでこれは矛盾です。したがって、 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ は一次独立です。

■対称行列の固有値は実数 一般に固有値や固有ベクトルを考える時は、 n 次方程式の根ですから複素数まで広げて考える必要があるのですが、幸いにして対称行列の場合はすべて実数で得られることがわかっています。

行列 A が実対称行列であるとし、この行列の固有値 λ に対する固有ベクトルを $\mathbf{x}$ としましょう。ここで λ や $\mathbf{x}$ は複素数だとします。

^{*1}「 p 次方程式は p 個の根を持つ」ことが代数学の基本定理ですが、証明には複素数も含めた議論が必要ですのでここでは省略します。ごめんなさい！

複素数については本書の範疇を超えますが、少しだけその性質を紹介しておきます。 a, b を実数, $i = \sqrt{-1}$ として $a + bi$ のことを複素数と言います。複素数 $\alpha = a + bi$ に対して $\bar{\alpha} = a - bi$ を共役複素数と言います。その性質として,

- α が実数 $\Leftrightarrow \alpha = \bar{\alpha}$
- $\overline{\alpha \pm \beta} = \bar{\alpha} \pm \bar{\beta}$
- $\overline{\alpha\beta} = \bar{\alpha}\bar{\beta}$
- $\alpha\bar{\alpha} = a^2 + b^2$

などがあります。

A は対称行列ですから, $A = A'$ ですし, $A = \bar{A}$ でもあります。さてこの固有値と固有ベクトルを使って, 次のような計算を考えてみましょう。

$$\begin{aligned}
 \lambda \bar{x}'x &= \bar{x}'\lambda x && \text{スカラーなので} \\
 &= \bar{x}'Ax && \text{固有値の定義} \\
 &= \bar{x}'\bar{A}x && \text{行列は実数なので} \\
 &= \bar{x}'\bar{A}'x && \text{行列は対称なので} \\
 &= (\bar{Ax})'x && \text{転置のルールから} \\
 &= (\bar{\lambda x})'x && \text{固有値に戻す} \\
 &= \bar{\lambda}\bar{x}'x
 \end{aligned}$$

ということで,

$$\lambda \bar{x}'x = \bar{\lambda}\bar{x}'x$$

という関係が得られましたが, これを移項するなどして

$$(\lambda - \bar{\lambda})\bar{x}'x = 0.$$

であることがわかります。ここで, $x \neq 0$ ですから, $\bar{x}'x \neq 0$ より $\lambda = \bar{\lambda}$ であることになります。ある値 λ が複素数であるとしても, その共役複素数 $\bar{\lambda}$ と等しいことから, λ が実数であることが示しました。

例題 20

対称行列 $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$ の固有値と固有ベクトルを求めましょう。固有値が実数になることを確認してください。

■対称行列の固有ベクトルは直交する さてこんどは同じく対称行列の, 固有ベクトルに関する性質です。対称行列 A の異なる固有値, λ_1, λ_2 に対して, それぞれ固有ベクトル x, y があるとします。

$$Ax = \lambda_1 x, Ay = \lambda_2 y$$

です。先ほどの式展開とほとんど同じなのですが、

$$\begin{aligned}
 \lambda x' y &= (\lambda x)' y && \text{スカラーなので} \\
 &= (Ax)' y && \text{固有値の定義} \\
 &= x' A' y && \text{転置の計算} \\
 &= x' A y && \text{対称行列なので} \\
 &= x' \lambda_2 y \\
 &= \lambda_2 x' y
 \end{aligned}$$

となり、

$$(\lambda_1 - \lambda_2) x' y = 0$$

が得られます。 $\lambda_1 \neq \lambda_2$ なので、 $x' y = 0$ 、内積が0なので直交していることがわかります。固有ベクトルを求めることは、(正規) 直交基底を求めることでもあるのです。

例題 21

さきほどの対称行列 $A = \begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}$ の固有ベクトルが直交していることを確認してください。

■対角化とスペクトル分解 固有値と固有ベクトルがわかると、行列をおもしろい形に変えることができます。

A を n 次正方対称行列、つまり $n \times n$ サイズの対称行列だとすると、 n 個の固有値が求められます。これを $\lambda_1, \lambda_2, \dots, \lambda_n$ として、対応する固有ベクトルを $x_1, x_2, \dots, x_n$ とします。ここで各固有ベクトルのノルムが1であるとしましょう。行列と固有値・固有ベクトルの関係から、

$$Ax_i = \lambda_i x_i$$

となりますが、このベクトルを並べた行列 $X = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$ を考えると、

$$AX = X\Lambda$$

と書くことができます。ここで Λ は

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

のような行列です。

対称行列の異なる固有値に対する固有ベクトルは直交するので、 X は直交行列です。さてこの関係から、

$$AXX' = X\Lambda X'$$

を考えれば、 $XX' = I$ ですから、

$$A = X\Lambda X'$$

となることがわかります。正方行列は、自身の固有ベクトルからなる行列 X と、固有値からなる行列 Λ と、固有ベクトルの転置行列 X' の積に分解できることがわかりました。この分解のことを**スペクトル分解**とよぶことがあります。この話はまた後で出てきますので、どうぞ楽しみに。

ついでに、さらにこの行列を固有ベクトルからなる行列で挟んで、

$$\mathbf{X}'\mathbf{A}\mathbf{X} = \mathbf{X}'\mathbf{X}\mathbf{\Lambda}\mathbf{X}'\mathbf{X} = \mathbf{\Lambda}$$

とすることもできます。 $\mathbf{\Lambda}$ は固有値が対角に入った対角行列ですから、このように変形することを**対角化**といいます。

この操作の何が楽しいかというと、行列の累乗計算がずいぶん楽になるところです。 $\mathbf{A}^n$ を考えると、

$$\begin{aligned}\mathbf{A}^n &= \mathbf{A}\mathbf{A}\mathbf{A}\cdots\mathbf{A} \\ &= \mathbf{X}\mathbf{\Lambda}\mathbf{X}'\mathbf{X}\mathbf{\Lambda}\mathbf{X}'\mathbf{X}\mathbf{\Lambda}\mathbf{X}'\cdots\mathbf{X}\mathbf{\Lambda}\mathbf{X}' \\ &= \mathbf{X}\mathbf{\Lambda}\mathbf{\Lambda}\cdots\mathbf{\Lambda}\mathbf{X}' \\ &= \mathbf{X}\mathbf{\Lambda}^n\mathbf{X}' \\ &= \mathbf{X}\begin{pmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^n \end{pmatrix}\mathbf{X}'\end{aligned}$$

となるからです。行列 $\mathbf{A}$ が座標の変換であつたり、何らかの作用を表現した行列であるとする、それを用いさせ続けた場合を考える時に累乗計算になりますが、その行列計算が一気に求められることになります。「何らかの作用」をハミルトン演算子と見るか、ページランクとみるか^{*2}、隣接行列^{*3}と見るかによって見える景色は違って来るかもしれませんが、数学的には固有値問題といういみで同じなのです。興味がある人は長沼 (2011) の電子版付録などもみてみるといいでしょう。

■総和はトレースに一致する さて今度は、 n 個の固有値を $\lambda_1, \lambda_2, \dots, \lambda_n$ とすると

$$\sum_i^n \lambda_i = \text{tr}(\mathbf{A})$$

ということを示します。これは対角化できることを考えると、

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{X}\mathbf{\Lambda}\mathbf{X}')$$

となります。トレースは行列の積に対して順序の入れ替えを許しますから (→Pp.21),

$$= \text{tr}(\mathbf{\Lambda}\mathbf{X}\mathbf{X}') = \text{tr}(\mathbf{\Lambda}) = \sum_i^n \lambda_i$$

です。

例題 22

さきほどの対称行列 $\mathbf{A} = \begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}$ の固有値の和がトレースと一致することを確認してください。

^{*2} Google の検索エンジンが採用しているページの評価法。Google の検索システムは、ページ同士のリンク関係を行列で表現し、スペクトル分解して重要度の高いページを重みづけて表現する、というアルゴリズムをとっており、「検索したいものに効率よく辿り着ける」ことで人気を博したからです。

^{*3} 数理社会学などでは対人的なネットワークの関係を行列で表現し、このように呼びます。ネットワーク科学一般でも用いられている用語と言えるでしょう。

7.2 固有値と固有ベクトルの幾何学的意味

固有値, 固有ベクトルについて, 今度は違う側面から見直してみましょう。

ある正方行列から固有値 λ と固有ベクトル $\mathbf{a}$ が得られたとします。このベクトル $\mathbf{a}$ のすべての要素を定数 c 倍したベクトル $\mathbf{b} = c\mathbf{a}$ を考えると, これもやはり同じ関係が成り立ちます。

$$A\mathbf{b} = \lambda c\mathbf{a} = \lambda \mathbf{b} \quad (7.1)$$

先ほどの計算でも明らかになりましたが, 固有ベクトルの値は絶対的なものではなく, 要素間の相対的大きさを反映しているに過ぎないのでしたね。

さてこれを幾何学的に, 図形として考えてみましょう。要素が 2 つのベクトルは, 2 次元座標に表現できます。ベクトル $\mathbf{x} = (x, y)$ という座標を表しているというわけです。固有ベクトルも要素が 2 つであれば, 座標で表現できます。先ほどの, 要素を c 倍しても固有ベクトルとしての性質は変わらない, という話は, 「固有ベクトルは大きさに意味はなく, 方向を表したもの」ということになります。では何の方向を指し示しているのでしょうか。

ここで行列の空間イメージで出てきた, 「行列は座標変換装置」ということを思い出してください。今回の例では, 行列 A には次のような性質があります。

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (7.2)$$

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (7.3)$$

そう, お気づきのように, これは固有値・固有ベクトルです。この行列の固有値・固有ベクトルはそれぞれ $\lambda_1 = 2, \mathbf{x}_1 = (1, 0)$, $\lambda_2 = 3, \mathbf{x}_2 = (0, 1)$ であることがわかります。この固有値, 固有ベクトルの組み合わせをじっとみていると, おもしろい特徴がわかって来ます。

今回の行列から得られた 2 つの固有ベクトル, $(1, 0)$ と $(0, 1)$ は, 2 次元平面の正規直交基底であることにお気づきのことと思います。つまり, 正方行列 A から得られる固有ベクトルは, その正方行列がもつ基底だといえるのです。

では固有値はどうでしょうか? 今回は x 座標を 2 倍, y 座標を 3 倍に引き延ばす変換をしたわけですが, この座標の歪み (重み) が固有値に対応していますね。つまり, 固有値と固有ベクトルは新しい座標に変換する, その変換先の空間的性質を表していることになります。元の座標空間は $(1, 0), (0, 1)$ で作られる空間ですが, 変換先の空間は $(2, 0), (0, 3)$ で作られている空間, ということになります。

つまり, 正方行列は空間を変換するものであり, 正方行列がもっている固有ベクトルを基底とした空間を通して変換されるということになります。

固有値と固有ベクトルの関係は, 次にお話する因子分析の話に直結します。多変量解析の文脈で行列がどのように応用されるのか, みていきましょう。

第 8 章

因子分析法

8.1 心理学における測定のモデル

8.1.1 古典的テスト理論

心理学における測定は、物理学などハードサイエンスと違って誤差が大きいことを見越して考えなければなりません。知能検査、性格特性などを測定するために考え出された古典的テスト理論のモデルは、

$$X = t + e$$

で表されます。すなわち、テストの点数 X は真のスコア t と誤差 e に分割できるというものです。非常に単純なモデルですが、測定したものには誤差がついているという考え方、言い換えると目に見えるものだけが真実ではないという考え方がしめされています。この考え方は、ソフトサイエンスの領域においては重要なことです。

またこの古典的テスト理論から、いくつかの重要な考えを読み取ることができます。ひとつは誤差についての考え方です。このモデルを $X_j = t_j + e_j$ のように、ある個人の変化しない属性について、 j 回測定したとします。この時、測定の平均は次のように計算できます*1。

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{j=1}^n X_j \\ &= \frac{1}{n} \sum_{j=1}^n (t_j + e_j) && \text{定義より} \\ &= \frac{1}{n} \sum_{j=1}^n t_j + \frac{1}{n} \sum_{j=1}^n e_j && \text{分配して} \\ &= \bar{t} + \bar{e}\end{aligned}$$

さて古典的テスト理論では、誤差に関して次のことが仮定されます。

- 誤差の平均はゼロ。つまり誤差が出現するときは、「正に偏る」「負に偏る」といった一貫した傾向がないと考える。

*1 この式は、ある測定を多くの個人 i に対して行ったものとして、 $X_i = t_i + e_i$ と考えることもできますが、添字が異なるだけで式の展開に違いはありません。

- 真のスコアと誤差との相関はゼロ。つまり誤差は真のスコアに関係なく影響してくるもので、真のスコアと共変動するようであれば偶然によるものとはいえない。
- 異なる測定誤差間の相関はゼロ。誤差同士がなにか意味のある変動をしているのであれば、それはもう偶然によるものとはいえない。

これらはいずれも、誤差が「偶然によって現れる影響で、制御不可能なもの」という考え方からは自然な仮定だといえるでしょう。より詳しくいえば、誤差は測定に応じて毎回一定の傾向で生じる**系統誤差 (systematic error)**と、まったく傾向のつかめない**偶然誤差 (random error)**に分けて考えられますが、系統誤差は測定に際して工夫して取り除くべき問題であり、ここでは、まったくの偶然による誤差についての議論だからです。

さて誤差の平均がゼロ、つまり $\bar{e} = 0$ ですから、 $\bar{X} = \bar{t}$ となって、いつかは誤差がなくなって真のスコアを得ることができるようになる、ということが示されます。

また平均は 0 ですが分散はゼロではありません^{*2}。このテストの分散を考えると、次のようなことがわかります。

$$\begin{aligned}
 Var(X) &= \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 && \text{定義より} \\
 &= \frac{1}{n} \sum_{j=1}^n ((t_j + e_j) - (\bar{t} + \bar{e}))^2 && \text{X をテスト理論のモデルに} \\
 &= \frac{1}{n} \sum_{j=1}^n ((t_j - \bar{t}) + (e_j - \bar{e}))^2 && \text{同じ記号でまとめる} \\
 &= \frac{1}{n} \sum_{j=1}^n ((t_j - \bar{t})^2 + 2(t_j - \bar{t})(e_j - \bar{e}) + (e_j - \bar{e})^2) && \text{展開する} \\
 &= \frac{1}{n} \sum_{j=1}^n (t_j - \bar{t})^2 + \frac{1}{n} \sum_{j=1}^n 2(t_j - \bar{t})(e_j - \bar{e}) + \frac{1}{n} \sum_{j=1}^n (e_j - \bar{e})^2 && \sum \text{を分配} \\
 &= Var(t) + 2Cov(te) + Var(e)
 \end{aligned}$$

ここで Cov とは共分散を表しています。第二項の $2Cov(te)$ は真のスコアと誤差との共分散 (を 2 倍したもの) を表していることになりませんが、共分散 (相関) がそもそも線形関係を表す指標であったことを思い出してください。相関係数は共分散を標準化したものだったわけですが、そういう意味ではこの $Cov(te)$ というのは真のスコアと誤差との相関関係を表しているようなものです。さて、ここでも誤差の仮定から、相関はゼロです。すなわち、誤差とはどのような傾向もなく出現するもの、という考えられているのです。どのような傾向もないわけですから、当然なにかと相関関係にあるはずがない、すなわち $Cov(te) = 0$ であるとなります。

これを踏まえると、 $Var(X) = Var(t) + Var(e)$ のように、テストの分散が真のスコアの分散と誤差の分散に完全に分割されました。ここから、信頼性の定義は $\frac{Var(t)}{Var(X)} = \frac{Var(t)}{Var(t) + Var(e)}$ と表すことができます。言葉で言えば、**信頼性**の定義は「全分散中にしめる真のスコアの分散の割合」ということになります。

^{*2} ガウスの考えた誤差論から、誤差は確率**正規分布 (Gaussian Curve)**に従うと考えられます。

8.1.2 単因子モデル

因子分析 (Factor Analysis) は古典的テスト理論の拡張であり、もっとも簡単な 1 因子モデルは次のように表されます。

$$z_{ij} = a_j f_i + e_{ij}$$

ここで Z_{ij} は個人 i の項目 j に対する反応を標準得点で表したもので、 a_j は項目 j の**因子負荷量 (factor loading)**、 f_i は個人 i の**因子得点 (factor score)**、 e_{ij} は個人 i と項目 j の組み合わせた時に生じた誤差と呼ばれます。

因子負荷量 (factor loading) とは、因子というこのテストで測定したい特性と、項目との関係の強さを表しているものです。**因子得点 (factor score)** とは、因子というこのテストで測定したい特性と、個人との関係の強さを表しているもので、その人のスコアだということができます。

記号について添字に注目してください。添字 i は個人を、添字 j は項目を表していますが、因子負荷量は a_j と表されています。つまり項目によって変わる変量です。因子得点は f_i と表されています。つまり人によって変わる変量です。古典的テスト理論をこの添字を使って表現するならば、 $X_{ij} = t_i + e_{ij}$ となりますが、これと比べてみると t_i が $a_j f_i$ に変わったのが因子分析モデルだということになります。 t_i は個人についての真のスコアなのですが、古典的テスト理論の場合、テストの点数は個人のこの能力だけを反映していると考えられていたことになります。もしテストの問題が難しすぎて、まったく答えることができれば、その人の能力はゼロということになるわけです。しかし中には悪い問題というものもあって、たとえば小学生に高校で習う知識が必要な問題を解かせるような問題があれば、誰だって解けないかもしれません。解けない問題を出して「学力が低い」と結論づけるのはやや暴力的ですらありますね。このように、古典的テスト理論は項目の良し悪しといったものが評価できないモデルだったのです。

因子分析モデルはこれを改良し、 $a_j f_i$ としています。すなわち、ある項目に対する反応 z_{ij} は、その項目が測定したい特徴を十分に反映しているかどうか (a_j) と、その人がその特徴を有しているかどうか (f_i) の両方が成立している必要があるわけです。掛け算ですから、一方がゼロであれば結果もゼロになります。すなわち測定したい特徴を反映していない項目 ($a_j = 0$) であれば、どれほどそれについての能力 (f_i) が高くても反応できないのです。たとえば美的センスに非常に秀でた人がいても、数学のテストでその能力を反映させることはできませんよね。これは数学のテストというのが数学力を測定するものであって、美的センスを測定するものではないからです。

8.1.3 多因子モデル

さて先ほどは一因子、あるいは単因子ともいいますが、測定したい特徴が 1 つだけのモデルでした。しかし、知能検査や性格検査の場合は潜在的な特性がいくつあるのかすら議論のたねになります^{*3}。そこで因子の数が 1 つではなく、複数ある多因子モデルを考えることにします。多因子モデルは次のように表現されます。

^{*3} Spearman, C.E. は知能の 2 因子説を唱え、一般因子 general factor と特殊因子 specific factors に区分しました。g 因子のような優れた才能が 1 つあれば、そのほかの能力もそれに影響されて現れてくるという考え方で、二次因子分析的モデルです。これに対して Thurstone, L.L. は多因子説を取ります。Reasoning ability, Verbal comprehension, Word Frequency, Numerical facility, Memory, Spatial Orientation, Perceptual Speed という要素を見出しています。他にも知能についてのモデルはいろいろありますが、1940 年代から因子分析モデルによってデータを説明しようという流れができてきたようです。

$$z_{ij} = a_{j1}f_{i1} + a_{j2}f_{i2} + a_{j3}f_{i3} + \cdots + a_{jm}f_{im} + d_j u_{ij} \quad (8.1)$$

記号の意味は単因子モデルと基本的には同じです。 z_{ij} は個人 i の項目 j に対する反応を標準得点で表したものであり、 a_{jm} は第 m 因子の**因子負荷量**、 f_{im} は第 m 因子の**因子得点**を表しています。因子の数が複数あるモデルですから、 $a_{j1}, a_{j2}, \dots, a_{jm}$ とか $f_{i1}, f_{i2}, \dots, f_{im}$ のように因子の番号と項目・個人の添字の組み合わせになっていることを確認してください。ここでは因子を m 個あるとしています。この因子はどの項目にも共通して働くので、**共通因子 (common factor)** と呼ばれます。また最後の e_{ij} が $d_j u_{ij}$ となっていますが、これは誤差についても他の因子と形を同じくし、項目に依存するものとそれ以外に区別しているだけです。この誤差因子については、

$$d_j u_{ij} = b_j S_{ij} + e_j E_{ij}$$

とさらに分割することがあります。ここで $b_j S_{ij}$ は特殊因子、 $e_j E_{ij}$ は誤差因子です。共通因子に対する誤差の中にも、テスト特有の特殊な要因と、全くの誤差の部分に分けて考えることがあります。とはいえ、ここでは線型結合のように表現できることが重要ですので、 $d_j u_{ij}$ のままで話を進めることにします。

8.2 因子分析モデルを展開する

因子分析モデルも古典的テスト理論のように、式の展開から何が見えてくるか考えてみましょう^{*4}。

左辺の z_{ij} は観測されたデータから算出できるものですが、右辺の因子負荷量、因子得点はいずれも未知数です。この未知数だらけの方程式を解くために、因子について以下のような条件を置きます。

- 共通因子の因子得点、独自因子の因子得点は、標準化されている。すなわち、いずれの因子得点も平均点は0であり、分散は1である。
- 独自因子は共通因子、他の独自因子と相関しない。

この他に、状況に応じて因子同士の間に関係を仮定します。

- 共通因子同士の相関を認めないのを「直交因子モデル」、認めるのを「斜交因子モデル」と呼ぶ。

その上で、2つの変数、 j と k の標準得点から、

$$r_{jk} = \frac{1}{n} \sum_{i=1}^n z_{ij} z_{ik} \quad (8.2)$$

のように、相関係数が算出されることをふまえ、先ほどの因子分析の基本代数式 (式 8.1) をこの式に代入してみましょう。

$$\begin{aligned} r_{jk} &= \frac{1}{n} \sum_{i=1}^n z_{ij} z_{ik} \\ &= \frac{1}{n} \sum_{i=1}^n (a_{j1}f_{i1} + a_{j2}f_{i2} + \cdots + a_{jm}f_{im} + d_j u_{ij})(a_{k1}f_{i1} + a_{k2}f_{i2} + \cdots + a_{km}f_{im} + d_k u_{ik}) \end{aligned} \quad (8.3)$$

これは代数の計算としてやっていくと、非常に煩雑で間違いが起きやすそうです。そこで、少し視覚化してわかりやすくしてみましょう。多項式の掛け算は、各項目の総当たり戦ですので、列方向に z_{ij} 、行方向に z_{ik} の各要素を置いて、要素同士の組み合わせ表を作ります (図 8.1)。

^{*4} 以下このセクションは小杉 (2018) の原稿を再構成したものです。

		$z_{ij} =$				
		$a_{j1}f_{i1} +$	$a_{j2}f_{i2} +$	$a_{j3}f_{i3} +$	$\cdots +$	$a_{jm}f_{im} + d_j U_{ij}$
$z_{ik} =$	$a_{k1}f_{i1}$					
	$+$					
	$a_{k2}f_{i2}$		①		②	
	$+$					
	$a_{k3}f_{i3}$					③
	$+$					
$\vdots$						
	$+$					
	$a_{km}f_{im}$					
	$+$					
	$d_k U_{ik}$			③		④

図 8.1 項目同士の総当たりを考える

図 8.1 に示されたのは個人 i についてのものであり、これが人数分ある、すなわち $\sum_{i=1}^n$ をつけないといけないことに注意してください。さて図を軽く色分けしてあるのですが、ここにあるように計算すべき領域を 4 つに分けて考えていきましょう。

- ①の領域 因子 p と p の積和部分 (同じ因子同士の掛け合わせ)
- ②の領域 因子 p と q の積和部分 (異なる因子同士の掛け合わせ)
- ③の領域 因子 $p(q)$ と独自因子の積和部分
- ④の領域 独自因子同士の積和部分

この各パートを順に計算していきましょう。まず①ですが、たとえば第一因子については

$$\frac{1}{n} \sum_{i=1}^n a_{j1} a_{k1} F_{i1}^2 \quad (8.4)$$

となることがわかります。ここで、 a_{j1} と a_{k1} は n には関係がない ($\sum$ は i が 1 から n まで変化することを意味しているが、係数 i はどちらにも入っていない) ので、総和して割る意味がないことに気づきます。そうすると、必然的にこの式は、

$$\frac{1}{n} \sum_{i=1}^n a_{j1} a_{k1} F_{i1}^2 = a_{j1} a_{k1} \frac{1}{n} \sum_{i=1}^n F_{i1}^2 \quad (8.5)$$

となります。この F_{i1} は因子得点であり、上の仮定より標準化されたものだということになります。さて、標準得点と標準得点の積和平均は相関係数になることをもう一度思い出してください！そうすると、これは自分自身との相関係数を表していることになりますから、当然 $F_{i1}^2 = 1.0$ であることがわかります。

結局、①のエリアは

$$\frac{1}{n} \sum_{i=1}^n a_{j1} a_{k1} F_{i1}^2 = a_{j1} a_{k1} \frac{1}{n} \sum_{i=1}^n F_{i1}^2 = a_{j1} a_{k1} \quad (8.6)$$

と、とてもあっさり書き下すことができるのです。

つづいて②を見てみましょう。ここは異なる因子がかけ合わさった部分ですね。落ち着いて、第一因子と第二因子を例にして考えてみましょう。この箇所で見られるのは、

$$\frac{1}{n} \sum_{i=1}^n a_{j1} F_{i1} a_{k2} F_{i2} = a_{j1} a_{k2} \frac{1}{n} \sum_{i=1}^n F_{i1} F_{i2} \quad (8.7)$$

ということになります。ここで、 F_{i1} および F_{i2} はそれぞれ第一、第二因子における個人 i の因子得点を意味しています。因子得点は標準化されていることをもう一度思い出すと、これは第一因子と第二因子の相関係数になります。ここで、この因子分析が直交因子モデルだと考えますと、因子同士に相関がないわけですから、数字としては 0 で消えてしまいます。するとこの部分は、

$$\frac{1}{n} \sum a_{j1} F_{i1} a_{k2} F_{i2} = a_{j1} a_{k2} \frac{1}{n} \sum F_{i1} F_{i2} = 0 \quad (8.8)$$

となることがわかりました。つまり、この領域②は、すべて 0 になってしまうのです。

続いて③の部分について考えてみましょう。これはある共通因子と独自因子の積の部分です。例によって標準得点同士の関係から、相関係数を算出することになりますが、独自因子は共通因子と無相関であることを考えると、

$$\frac{1}{n} \sum a_{i1} F_{i1} d_j U_{ij} = a_{ik} d_j \frac{1}{n} \sum U_{ij} F_{i1} = 0 \quad (8.9)$$

とこのように、この箇所もすべて 0 になってしまいます。

最後の④に至っては、独自因子と独自因子の積和ですから、これも

$$\frac{1}{n} \sum d_j d_k U_{ij} U_{ik} = d_j d_k \frac{1}{n} \sum U_{ij} U_{ik} = 0 \quad (8.10)$$

のように 0 になります。結局、消えて無くなるのがほとんどで、残るのは①の部分だけであり、 r_{jk} を考えるときはそこだけ考慮すればよいことになります。

整理すると、

$$r_{jk} = a_{j1} a_{k1} + a_{j2} a_{k2} + \cdots + a_{jm} a_{km} \quad (8.11)$$

ということがわかります。つまり、項目 j と項目 k の相関係数は、項目 j の因子負荷量と項目 k の因子負荷量を、すべての因子について総和したものであるということです。因子分析の基本モデルから導出されるこの定理を、とくに**因子分析の第二定理**と呼びます。

ここで同じ項目同士の相関を考えてみましょう。項目 j と項目 j の相関係数は、もちろん 1.0 になりますね。これを因子分析の基本式で表すと、次のように表現できます。

$$r_{jj} = a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2 + d_j^2 = 1.0 \quad (8.12)$$

さて、この式が意味するのはなんでしょうか。意味を考えてみると、ある項目それ自身との相関係数は、因子負荷の二乗和からなっている、ということがわかります。これこそ**因子分析の第一定理**と呼ばれるものであり、解けるはずのなかった方程式を解くための鍵となる式なのです。

8.2.1 因子分析の定理

数式の展開はいったんここまでにして、第一定理は次のような形をしているのでした。

$$a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2 + d_j^2 = 1.0$$

ここで共通因子部分を、 $a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2 = h_j^2$ のようにすると、この式は単に $h_j^2 + d_j^2 = 1.0$ となります。この h_j^2 のことをとくに**共通性 (communality)** といいます。この式は共通性と独自因子の二乗和が 1.0 になることを意味しています。言い換えると、全体を 100% とした比率で共通性と誤差を比較できるということです。共通性は因子負荷量の二乗和で、共通因子はそのテストの背後にある共通の要因、すなわちテストで測定したいものだったわけです。古典的テスト理論では、モデル式の展開から信頼性を全分散中に示る真のスコアの割合と定義しましたが、因子分析モデルはこのように 1 つの項目 j における共通因子の割合

を算出し、項目の信頼性を考えることができます。因子分析モデルにおける信頼性は、1 項目の中に含まれる共通因子の大きさだとも言えるわけです。逆に $d_j^2 = 1 - h_j^2$ は**独自性 (uniqueness)** は、当該項目がそのテストで測っていないものの大きさを表しており、この割合があまりにも大きいと「この項目は全然関係ないものを測っちゃってるんじゃないか？」と疑われることになります。多因子モデルにおいては、多角的に対象を切り分けるために多くの質問を投げかけるわけですが、独自性の高い項目は回答者に負担をかけるだけの邪魔なものですから、実践上はこうした項目を除外することが少なくありません。因子分析には**単純構造の原則 (principle of simple structure)** と呼ばれるものがあり、項目は該当する因子を適切に反映し、かつ、他の因子と関係ないことが美しいとされます。尺度構成段階では、共通性 (独自性) をみて項目の良し悪しが判断されるのです。

次に第二定理を見てみましょう。第二定理は次のような形をしているのです。

$$r_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \cdots + a_{jm}a_{km}$$

2 つの項目 j と k の相関係数は、それぞれの因子負荷量の積和の形で表される、というものです。ここに誤差の話は入ってこず、共通因子だけで話ができています。

左辺の相関係数は、2 つの項目がどれほど同じものを測定しているかの指標です。相関係数 (の絶対値) が高ければ高いほど、2 つの項目は同じものを指し示しているわけです。逆に相関係数が低いということは、2 つの項目に関係がないことを表します。ここで右辺に目をやりますと、右辺の各項目は因子負荷量の積の形になっています。左辺の値が小さくなる 1 つの理由は、ある共通因子 m が項目 j, k に対して、異なる方向で寄与しているからだと考えられるでしょう。そしてそのパターンが一貫していないという状況です。そもそも相関係数が小さいところからは因子を見つけ出すのは難しいのですが、そうした状況があるのはある項目ペアについて因子同士の向きがバラバラに影響しているからだと言えます。そのような状況は、測定がきちんとできているかどうか怪しいですね。**測定の一義性**とも言われますが、そのような尺度は妥当性が低いといえるでしょう。

このように、因子分析モデルは第一定理で信頼性を、第二定理で妥当性をあらわすものになっているのです。

8.3 因子分析モデルの行列表現

さて、行列とベクトルを学んできた皆さんは、このようなエレメントワイズの計算はうんざりで、行列でまるっと表現できるはずなのに、とうとうずしていることと思います。ということで、因子分析モデルの代数的表現ですが、これも行列を使って表現すると非常にシンプルに表現できるということを説明していきましょう。

内容はまったく同じですが、確認しておきましょう。標準得点行列 Z を因子負荷行列 A と因子得点行列 F をつかって、次のように表します。

$$Z = FA' + UD \quad (8.13)$$

ここで、各行列の要素のサイズ感をつかんでおきましょう。まず R というのは相関行列ですから、 m 個項目があるのでサイズは $m \times m$ の正方行列になります。次に F ですが、これは因子得点の行列です。得点は人数分ありますから行は n 、因子の数が列になるのでこれを p とすると $n \times p$ です。 A は因子負荷行列。因子負荷行列は因子の数と項目の数の組み合わせだけあるわけですから、 $m \times p$ になりますね。 U は独自因子得点です。得点ですから人数分、独自性分は各項目にありますから、サイズとしては $n \times m$ になります。最後に D ですが、これは独自因子の負荷量です。項目の数だけあるのですが、列ベクトルや行ベクトルで表現すると計算の時にサイズが変わって不便なことになります。ですから、対角項に d_j をもつ正方行列 $m \times m$ として表現しています。

サイズを確認したところで、実際に行列計算をしてみましょう。

$$\begin{aligned}
 R &= \frac{1}{n} Z' Z \\
 &= \frac{1}{n} (FA' + UD)' (FA' + UD) \\
 &= \frac{1}{n} \{ (FA')' + (UD)' \} (FA' + UD) \\
 &= \frac{1}{n} (AF' + D'U') (FA' + UD) \\
 &= \frac{1}{n} AF'FA' + \frac{1}{n} AF'UD + \frac{1}{n} D'U'FA' + \frac{1}{n} D'U'UD
 \end{aligned}
 \tag{8.14}$$

と、このように展開できました。記号を見ているとわかりにくいので、サイズ感を確認しましょう。最終的には、次のようになっています。

$$R = \frac{1}{n} \underset{m \times m}{A} \underset{n \times m}{F'} \underset{m \times n}{F} \underset{n \times m}{A'} + \frac{1}{n} \underset{n \times m}{A} \underset{m \times pp}{F'} \underset{pp \times nn}{U} \underset{nn \times mm}{D} + \frac{1}{n} \underset{n \times m}{D'} \underset{m \times mm}{U'} \underset{mm \times nn}{F} \underset{nn \times pp}{A'} + \frac{1}{n} \underset{n \times m}{D'} \underset{m \times mm}{U'} \underset{mm \times nn}{U} \underset{nn \times mm}{D}$$

ここで、要素ごとに計算していた時のことを思い出してください。第二項 $\frac{1}{n} AF'UD'$ と第三項 $\frac{1}{n} D'U'FA'$ の中にある、 $F'U$ と $U'F$ のところは、共通因子得点と独自因子得点の積ですし、いずれも標準化されていますから、 $\frac{1}{n}$ と合わせて考えると、これは相関係数を表していることになります。また、共通因子と独自因子は相関しませんので、これはイコール 0 となり、この2つの項が消えてしまうのでした。

また、第一項の $\frac{1}{n} F'F$ は、共通因子同士の相関を表しています。 $F'F = C$ とすると、これは因子得点間相関 C を表すことになります。これが直交であると仮定する、つまり他の因子と相関しないと考え、 $C = I$ 、つまり単位行列です。単位行列は計算に影響を与えませんから、 $AF'FA' = ACA' = AIA' = AA'$ となり、この式は簡単に次のように変形できます。

$$R = AA' + D^2 \tag{8.15}$$

先ほどの代数的展開を、そのまま行列で表現しただけですが、この方がシンプルに表現できていますね。この表現は、因子分析の第一定理と第二定理の両方を含んで一度に表せているのです。

8.3.1 因子分析とスペクトル分解

さて、相関行列が因子負荷行列と誤差の行列に分解できることが示されましたが、相関行列は正方対称行列ですので固有値と固有ベクトルを使ってスペクトル分解できます。

$$R = \Lambda X X'$$

この式は、間に入る Λ が対角行列であり、 X が各固有ベクトルを列に持つ行列であることを考えると、次のように書くこともできます。

$$R = \lambda_1 x_1 x_1' + \lambda_2 x_2 x_2' + \cdots + \lambda_n x_n x_n'$$

念のため、 2×2 の簡単な例で確認してみましょう。

$$\begin{aligned}
 & \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\
 &= \begin{pmatrix} \alpha a & \beta c \\ \alpha b & \beta d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\
 &= \begin{pmatrix} \alpha aa + \beta cc & \alpha ab + \beta cd \\ \alpha ab + \beta cd & \alpha bb + \beta dd \end{pmatrix} \\
 &= \alpha \begin{pmatrix} aa & ab \\ ab & bb \end{pmatrix} + \beta \begin{pmatrix} cc & cd \\ cd & dd \end{pmatrix} \\
 &= \alpha \begin{pmatrix} a \\ b \end{pmatrix} (a \ b) + \beta \begin{pmatrix} c \\ d \end{pmatrix} (c \ d)
 \end{aligned}$$

さて $\lambda \mathbf{x}\mathbf{x}' = \mathbf{a}\mathbf{a}'$ となるようにベクトルの大きさを整えてみましょう。これは λ を分解してベクトルの中に溶け込ませるようなものですから、 $\mathbf{a} = \sqrt{\lambda} \mathbf{x}$ とすればよいでしょう。行列のままで表現するなら、固有値の平方根を対角にもつ行列 $\Lambda^{\frac{1}{2}}$ を使って次のように表現できます。

$$\mathbf{R} = \mathbf{X}\Lambda\mathbf{X}' = \mathbf{X}\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}\mathbf{X}' = \mathbf{A}\mathbf{A}'$$

つまり $\mathbf{A} = \mathbf{X}\Lambda^{\frac{1}{2}}$ と考えることができます。これは n 次元ベクトルが n 個並んでいることになりませんが、ある基準で区切って m 個のベクトルだけを共通因子、残りはまとめて誤差因子と考えて次のように書き直してみましょう。

$$\mathbf{R} = \mathbf{a}_1\mathbf{a}_1' + \mathbf{a}_2\mathbf{a}_2' + \cdots + \mathbf{a}_m\mathbf{a}_m' + \mathbf{d}\mathbf{d}' = \mathbf{A}\mathbf{A}' + \mathbf{D} \quad (8.16)$$

ここでは共通因子の数が m 個だとわかっている体で分解していますが、基本的にはサイズ n の行列からは n 個の固有値がずら一と並んでいたのです。それをどこかで「共通しているのはここまで」と判断し、残りは誤差であるとしてまとめて $\mathbf{d}\mathbf{d}'$ にしてみました。このように数学的にはここからが共通因子、ここからが独自因子といった区別をすることなく、最後のひとかけらまで固有値分解を行なっているのですが、その次元の重要度をもって共通因子と誤差因子に（研究者が恣意的に）分割しているのが因子分析のやっていることなのです。

実際に因子分析を（代数的に）解くときは、 $\mathbf{R} = \mathbf{A}\mathbf{A}' + \mathbf{D}$ から、 $\mathbf{R}_{\dagger} = \mathbf{R} - \mathbf{D} = \mathbf{A}\mathbf{A}'$ として、相関行列 $\mathbf{R}$ から、誤差因子の負荷量の二乗和が対角に並んだ $\mathbf{D}$ を引いた行列 $\mathbf{R}_{\dagger}$ をスペクトル分解/対角化し、

$$\mathbf{R}_{\dagger} = \mathbf{X}\Lambda^{\frac{1}{2}}\Lambda'^{\frac{1}{2}}\mathbf{X}' = \mathbf{A}\mathbf{A}'$$

として共通因子負荷量を求めます。

この $\mathbf{R}_{\dagger}$ は、

$$\mathbf{R}_{\dagger} = \begin{pmatrix} 1 - d_1^2 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 - d_2^2 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & 1 - d_m^2 \end{pmatrix} = \begin{pmatrix} h_1^2 & r_{12} & \cdots & r_{1m} \\ r_{21} & h_2^2 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & h_m^2 \end{pmatrix}$$

のように対角項に共通性が含まれた行列であり、純粋な相関行列ではありません。

さて事前に h_j^2 の大きさがわかっているのですが、最初はわからないので、これをどういう数字にするかが実際に計算する上での悩み事になります。これが**共通性の推定問題**で、さまざまな方策が考えられて

きました。一般的には重相関係数の二乗 (SMC)^{*5} をラフな推定値として初期解を求め、そのときの固有値の減衰状況などから因子数を決めて改めて計算し、推定値に変動が起きなくなるまで反復を繰り返す、という方法が取られます^{*6}

また相関係数を固有値分解すると、その固有値はすべて足し合わせると行列のトレースから n になることがわかります。元のデータから計算される相関行列は、1つの項目が一単位分の情報を持っていると考えますが、固有値分解はそれを次元の重要性順に並べ替えているともいえるわけです。固有値は項目いくつかの重要度があるかということを表す指標だと考えることができます。どこから共通因子でどこからが誤差か、ということを考えてときに、たとえば固有値が 1.0 よりも小さくなるようであれば、項目 1 つ分の情報もないのだからということでも誤差因子だと判断することがあります。

ところで因子分析モデルの A 、因子負荷行列ですが、これは共通因子の固有ベクトルをセットで扱ったものです。つまり、 $A = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \end{pmatrix}$ と縦ベクトルを並べたものになっています。エレメントワイズで表現すると次のようになります。

$$A = \left(\begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1m} \end{pmatrix}, \begin{pmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2m} \end{pmatrix}, \cdots, \begin{pmatrix} a_{p1} \\ a_{p2} \\ \vdots \\ a_{pm} \end{pmatrix} \right)$$

ここで AA' の間に単位行列 I を挟んでも、別に結果は変わりません。つまり $AA' = AIA$ です。

$I = TT' = T'T$ になるような m 次の行列を使うと、 $AA' = ATT'A'$ の関係は保たれたままです。この T はこれまた行列の座標を変えてしまう変換行列であり、とくに回転行列 (rotation matrix) ともいうのでした。さて、これを挟むことができるということは因子負荷量の値はなんでもありだということになってしまいます。因子分析はこのようにどんな回転でもできる、回転に関する不定性があるのです。あらゆる因子負荷量の組み合わせがあり得る、というのは困りものなので、なんらかの形で因子負荷行列 A に制約をかける必要があります。言い方を変えれば、色々な制約の中ではあっても好きな値を取ることができますから、ユーザにとって便利な基準を考えてやれば良いでしょう。因子分析では一般に、因子軸の回転を行います、それはこうした理由からです。

図 6.5 や図 6.6 でみたように、回転することで軸ごとの座標にメリハリがつかます。因子分析の場合、「どの項目がどの軸に乗っているか」がはっきりしている単純構造 (Simple Structure) が好まれますから、わかりやすく回転させるのは便利なことなのです。

ちなみに TT' に $\text{diag}(T\Phi T') = I_m$ という制約のある行列 Φ を挟んでやっても、元の計算モデルに影響はありません。この Φ は因子間相関 (factor correlations) とよばれ、相関を持った因子軸の回転、すなわち斜交回転 (oblique rotation) も色々なものが考えられています^{*7}。

直交回転と斜交回転の違いを考えると、強調される座標について注意が必要です。直交の場合、座標の取り方に悩むことはありませんが、斜交の場合は「軸に垂線を下ろす」という考え方ももちろんありますが (正射影)、「他の軸に平行な線を引く」 (平行射影) という考え方もできるからです。

図 8.2 にあるように、正射影の座標 (s_{i1}, s_{i2}) を因子構造 (Factor structure) とよび、平行射影の座標

^{*5} 項目 j に対して、それ以外の項目を使って回帰分析を行い、その重相関係数の二乗 $R_{y_j}^2$ を共通性の推定値とみなす方法です。重相関係数の二乗は、目的変数の分散が、説明変数の線型結合でどの程度説明されたかを表す指標です (この辺りは (小杉, 2018) を参照してください)。共通性とはある項目が他の項目とどの程度共通して空間を張るか、という程度の指標だと考えられますから、重相関係数の二乗がその推定値として合理的だと考えられるわけです。

^{*6} 反復主因子法と呼ばれる推定法です。推定法はこの他にも、多変量正規分布を仮定する最尤法や、残差を最小化するように求める最小二乗法などがあります。

^{*7} これに対して $TT' = I$ のような因子間相関がない (単位行列) な回転を直交回転 (orthogonal rotation) といいます。

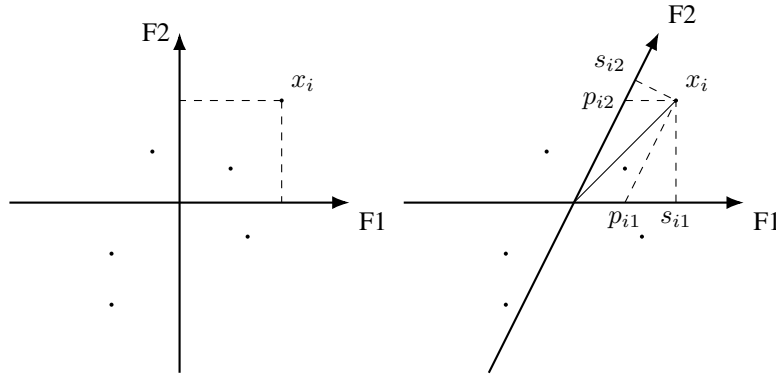


図 8.2 直交の場合は正射影と平行射影が同じになる (左図) が, 斜交の場合はそうではない (右図)

(p_{i1}, p_{i2}) を **因子パターン (Factor Pattern)** と呼んで区別します。

8.4 主成分分析と因子分析

8.4.1 主成分の分散と固有値

セクション 6.2 (Pp.93) では分散共分散行列の軸を回転させる, という話がありました。固有値と固有ベクトルについての知識を踏まえて, 改めてこの回転について見てみましょう。

データ行列から各変数の平均を引いた, 平均偏差行列 V を作ったとします。 n 人 m 変数からなるデータセットにおいて, 分散共分散行列 S は,

$$\begin{aligned} S &= \frac{1}{n} V' V \\ &= \frac{1}{n} \begin{pmatrix} x_{11} - \bar{x}_1 & x_{21} - \bar{x}_1 & \cdots & x_{n1} - \bar{x}_1 \\ x_{12} - \bar{x}_2 & x_{22} - \bar{x}_2 & \cdots & x_{n2} - \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{1m} - \bar{x}_m & x_{2m} - \bar{x}_m & \cdots & x_{nm} - \bar{x}_m \end{pmatrix} \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1m} - \bar{x}_m \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2m} - \bar{x}_m \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{nm} - \bar{x}_m \end{pmatrix} \end{aligned}$$

として計算できます。

さてこの S の固有値を $\lambda_1, \lambda_2, \dots, \lambda_n$ とし, また対応する固有ベクトルを $u_1, u_2, \dots, u_n$ とすると,

$$U' S U = \Lambda$$

とできるのでした。つまり,

$$U' S U = U' \frac{1}{n} V' V U = \Lambda$$

となります。ここで $\frac{1}{n}$ はスカラーですからベクトルの中にかけてしまうことができます。特に, $W = \frac{1}{\sqrt{n}} V$ としておくと $U' W' W U$ と表せます。

U は各列が互いに直交していますから, UW は平均偏差を $\sqrt{n}$ で割った, いわばスケールダウンした座標を回転させて新しい座標に移したことです。

$$\begin{aligned}
U'W'WU &= \\
U' \frac{1}{\sqrt{n}} \begin{pmatrix} x_{11} - \bar{x}_1 & x_{21} - \bar{x}_1 & \cdots & x_{n1} - \bar{x}_1 \\ x_{12} - \bar{x}_2 & x_{22} - \bar{x}_2 & \cdots & x_{n2} - \bar{x}_2 \end{pmatrix} \frac{1}{\sqrt{n}} \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 \\ \vdots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 \end{pmatrix} U \\
&= \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \Lambda
\end{aligned}$$

二変数のイメージを図 8.3 に示しました。各次元の座標の二乗和が、各変数の分散に該当しますし、新しい座標の分散がそれぞれの固有値に対応しています。回転させてもデータ点が全体的に占める面積 (楕円の領域) に変化はありません。行列から得られる固有値を大きい順に並べたとき、最大固有値に対応する固有ベクトルは、分散を最大にする軸であると言えます。

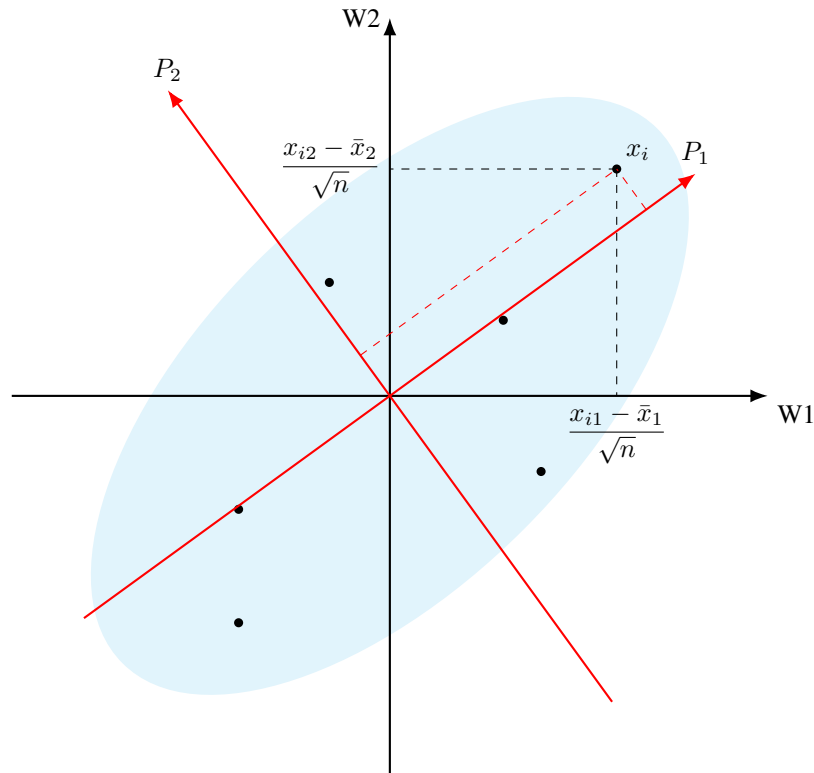


図 8.3 スケールダウンした平均偏差行列をつかって表したデータ空間

主成分分析を別の表現で説明するなら、複数の変数 $x_1, x_2, \dots, x_m$ があつたときに、合成変数 P_i を

$$p_i = w_1 x_{i1} + w_2 x_{i2} + \cdots + w_m x_{im}$$

の分散を最大にするように w_j を決めることともいわれます。これはむしろ、上の行列計算を逆に読み解くようなものですね。ここで x_{ij} は平均からの偏差であるとすれば、

$$\mathbf{p} = w_1 \begin{pmatrix} x_{11} - \bar{x}_1 \\ x_{21} - \bar{x}_1 \\ \vdots \\ x_{n1} - \bar{x}_1 \end{pmatrix} + w_2 \begin{pmatrix} x_{12} - \bar{x}_2 \\ x_{22} - \bar{x}_2 \\ \vdots \\ x_{n2} - \bar{x}_2 \end{pmatrix} + \cdots + w_m \begin{pmatrix} x_{1m} - \bar{x}_m \\ x_{2m} - \bar{x}_m \\ \vdots \\ x_{nm} - \bar{x}_m \end{pmatrix}$$

となります。これは $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \end{pmatrix}$ という重みベクトルを平均偏差ベクトルにかけて、

$$\mathbf{p} = \mathbf{V}\mathbf{w}$$

としていることであり、

$$\frac{1}{n}\mathbf{p}'\mathbf{p} = \frac{1}{n}\mathbf{w}'\mathbf{V}'\mathbf{V}\mathbf{w} = \mathbf{w}'\mathbf{S}\mathbf{w}$$

とすることができます。一方、固有値・固有ベクトルを使った分散共分散行列のスペクトル分解を考えると、

$$\mathbf{S} = \lambda_1 \mathbf{x}_1 \mathbf{x}_1' + \lambda_2 \mathbf{x}_2 \mathbf{x}_2' + \cdots + \lambda_m \mathbf{x}_m \mathbf{x}_m'$$

であり、 $\mathbf{x}_j, \mathbf{x}_k$ はそれぞれノルム 1 で互いに直交する ($\mathbf{x}_j' \mathbf{x}_k = 0$) ことから、

$$\mathbf{x}_j' \mathbf{S} \mathbf{x}_j = \lambda_j$$

とすることができます。つまり主成分を作る重み $\mathbf{w}$ は固有ベクトルを選べばよく、合成変数 $\mathbf{p}$ の分散はその固有ベクトルに対応する固有値であるといえます。

8.4.2 因子分析と主成分分析の違い

主成分分析と因子分析は、似ているような似ていないようなところがあります。両者の違いについて、次のように言われることがあります。

- 主成分分析は一成分だけ求めるものであり、複数の共通成分といった考え方はしない。
- 因子分析は軸の回転をするが、主成分分析は軸の回転をしない。
- 主成分分析は分散共分散行列を対象に行われるものであり、因子分析は相関行列を対象に分析するものである。

しかしこれらは、数学的には不必要な制約、解釈をしているといえます。

主成分分析は「分散を最大にする合成変数を作る」という目的で行われますから、一番重要な成分だけ作って「総合的に判断する」という使い方がされてきました。しかしここまでの議論で明のように、分散説明率を効率よく再配分しているだけですから、複数の成分を考えることは別に問題ではありません。心理学でつかわれる因子分析は、その誕生の経緯から多因子論が一般的ではありますが、主成分分析において第二、第三の次元を抽出することを否定するものではありません。

軸の回転についても同様で、因子分析は心理学的な理論的背景から斜交回転のような考え方が出てきているとも言えます。確かに直交する次元を抽出したのに、わざわざ斜交で解釈するのは、側からみるとなやあってんの、と思われるところもあるかもしれません。しかし複数の主成分を抽出して、解釈しやすくするために斜交回転することは、技術的に可能ですし数学的に問題があるわけではないのです。あくまでも使う状況によって、必要とされる表現方法が異なるというだけです。

主成分分析は分散の説明という意味がありますから分散共分散行列をスペクトル分解する、というのは確かにそうです。しかし分散共分散行列と相関行列の違いは、単位がある尺度かどうかということに過ぎません。心理学でよく使われる因子分析が相関行列から始められるのは、心理学の測度は単位を持たない、相対的な関係で表現するものだからです。相関行列に対して主成分分析を行っても問題があるわけではなく、統計パッケージによっては主成分分解として相関行列をもとにした主成分分析を選ぶこともできます。

それよりも大きな違いがある点は、主成分分析は $\mathbf{S} = \mathbf{L}\mathbf{L}'$ と分解するのに対し、因子分析は $\mathbf{R} = \mathbf{A}\mathbf{A}' + \mathbf{D}^2$ とするところです。ここで $\mathbf{D}^2$ は誤差成分からなる対角行列です。ですから、因子分析がスペクトル分解するのは $\mathbf{R}_\dagger = \mathbf{R} - \mathbf{D}^2 = \mathbf{A}\mathbf{A}'$ なのです。

どちらから計算を始めるか, というのは数学の問題ではなく, 応用領域固有の背景に依存する問題です。人文社会科学の領域においては, 項目が持つ情報に誤差成分が含まれていると考える方が妥当であることが多いでしょう。そのような時は因子分析を使うべきであり, 逆に経済指標など誤差なく計算され, また単位に意味があるデータセットを対象にするのであれば, 主成分分析でも十分だということになります。

ちなみに因子分析の場合は, 相関行列から誤差の情報を取り除いた上で次元解析をしますので, 項目の数だけ因子が出てこないこともあります。ランク落ちが甚だしい場合はデータセット上に問題があることも考えられるので, 実践上は注意が必要です。

第9章

共分散構造

最後に**構造方程式モデリング (Structural Equation Modeling, SEM)** について説明していきます^{*1}。この手法はこれまで学んできた因子分析や回帰分析を統合したもので、分散共分散行列に方程式を作り込んでいくものです。我々が調査で得たデータから得られる情報は、分散共分散行列がすべてですから、その変数間関係を細かく作り込んでいき、行列の隅々まで考え尽くすという意味で、究極の手法といえるでしょう。実際、SEM は因子分析や回帰分析の他にも、多くのモデルをその下位モデルとして包含します。言い換えれば、SEM が登場する前に考えられてきたさまざまなモデルが、SEM の枠組みで理解・表現できるようになりました。であれば、この究極の技さえ知っておけば、非常に広範囲に拡張できるわけですから、こんなに便利なことはありません。

構造方程式モデリングについても、データが作る行列の世界をイメージすると理解しやすくなります。

9.1 パスダイアグラムと方程式

構造方程式モデリングでは、変数間関係を方程式で表して分析を進めますが、論文などで一般的に目にするのは変数間関係を図示したもの、すなわち**パス図**あるいは**パスダイアグラム (Path Diagram)** とよばれる表現でしょう。パス図は直感的に非常にわかりやすいのですが、そこで表現されているものと方程式で表現されているものは同じものの別表現であることを意識しましょう。

9.1.1 準備 1; パス図の書き方

分析する際の変数には**観測変数 (Observed Variables)** と**潜在変数 (Latent Variables)** があります。観測変数はその名の通り、観測された変数であり、データとして数値化されたものになります。これに対して潜在変数とは、モデルで仮定された変数のことです。たとえば因子分析やテスト理論では、因子や学力といったものを想定します。それが潜在変数です。潜在変数はデータとして数値化されているのではなく、抽象的な概念ですが、図にする場合はそれも表現しなければなりません。そこで観測変数を矩形 (長方形) で、潜在変数を楕円で表現することにします。

また、変数同士の関係を表現する必要がありますが、関係には**因果関係**と**相関関係**があります。調査データから因果関係を見出すのは原理的に不可能ですが、モデル上は一方が他方の原因になっている、と考えることがあります。回帰分析はその典型例で、説明変数が原因となって、被説明変数のあたいが結果的に変わる、という関係ですね。こうした関係を一方向の矢印で表現します。他方、相関関係は因果関係よりも緩やか

^{*1} この手法は別名**共分散構造分析**という名前でも呼ばれています。書籍や論文を検索する時は、こちらの名称もキーワードにしてみてください。

で、何らかの関係がある、ということの意味しているに過ぎません。因果の向きが定まっているのではなく、どちらの向きにも影響しうる、ということで双方向矢印でもってこれを表現します。注意してほしいのは、ここでいう因果関係とは一方が原因、他方が結果という一方向的な影響を表してはいますが、本当に現象として因果的であるかどうかは一時点のデータセットから確定できるものではないという点です。回帰分析でいうところの、変数を説明・被説明の関係とみなすというのと同じ意味での因果関係に過ぎないことに注意してください。



図 9.1 SEM による変数と変数間関係の描画

基本的な表記ルールはこれぐらいです。「誤差変数は潜在的だが楕円にしない」とか、「誤差変数は煩雑になるので描画しない」といったマイナーな作法がいろいろ見られますが、本質的な違いはありません。

9.1.2 準備 2; 確率変数とそのベクトル表記

さて、本書でのここまでの話は、特に推測統計学的な説明を含んでいませんでした。しかし心理学では一般に標本と母集団を区別し、母集団での一般化した議論を展開していきますし、構造方程式モデリングでも基本的には母数の話をしたいところですので、少し準備をしておきます。具体的には、これまでの変数＝データセットを確率変数と見做し、標本を使って母集団のそれらを推測するもの、と考えるようにしたいので、確率変数の表記ルールを導入します。

これまでの平均、分散は次のように定義してきました。

$$\bar{x} = \frac{1}{n} \sum_i^n x_i$$

$$s_x^2 = \frac{1}{n} \sum_i^n (x_i - \bar{x})^2$$

これらはあくまでも標本の統計量であって、母数の推定値とするためには分散を

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_i^n (x_i - \bar{x})^2$$

で計算します。

また、推定したい母集団の状態は、基本的に**確率分布**していると考えますので、確率変数についての考え方も少し取り入れる必要があります。確率変数の平均は**期待値**と呼ばれ、確率変数 X の期待値は

$$E[X] = \sum p_k x_k$$

で定義します。ここで p_k は変数 X が x_k の値になる確率であり、確率と実現値の積和で表されるものです。確率変数の分散は、次のようになります。

$$V[X] = \sum p_k (x_k - E[X])^2 = E[(X - E[X])^2]$$

行列計算も同様に考えることができます。

さて、構造方程式モデリングではこのように、変数を確率変数と捉えます。ベクトルは基本的に列ベクトルで考えることにしますので、確率変数が縦に並んでいるものとして考えると、本書のここまでの考え方をコロリと転置させた方が良さそうです。すなわち、ここまではデータセットを観測度数 n のベクトルを列に含む $n \times m$ 行列 $\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix}$ として考えていましたが、観測されるデータ x_{ij} は確率変数 x_j の実現値ですから、データ行列は確率変数ベクトルを列に並べた組、

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

であると考えことにします。

こうすると、たとえば平均ベクトル

$$\bar{x} = \frac{1}{n} \mathbf{X} \mathbf{1}$$

については、

$$E[\mathbf{x}] = (E[x_1], E[x_2], \dots, E[x_m])' = (\mu_1, \mu_2, \dots, \mu_m)' = \boldsymbol{\mu}'$$

のように表現しますし、母集団における分散共分散行列も、

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] = E[\mathbf{v}\mathbf{v}']$$

のように表現できます。母集団における相関行列は、

$$\boldsymbol{\Sigma}_r = E[\mathbf{z}\mathbf{z}']$$

などします。

以下の説明の利便性のために少し表記が変わっただけで、本質的な違いではありませんのでご容赦いただければと思います。

これで準備が整いました。それでは構造方程式モデリングの内容に進んでいきましょう。

9.2 構造方程式

さて、構造方程式モデリングはその名の通り、構造を方程式で書いて表すことです。この時、方程式は大きく2種類に分けることができ、変数間の関係を表現するものを**構造方程式**、(潜在)変数の組成について表現するものを**測定方程式**といいます。ここではまず構造方程式の方から見ていきましょう。まずは図 9.2 にある例からです。

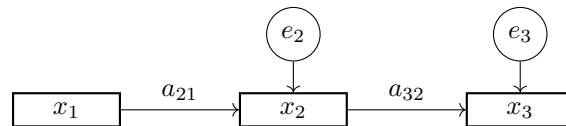


図 9.2 パス解析

この分析モデルは、 x_1 が x_2 の説明変数になり、 x_2 がまた x_3 の説明変数になるという関係を表しています。影響力の大きさをパス係数 a_{jk} で表しています。ここで j が説明される変数の番号、 k が説明する変数の番号とするのが一般的な表記です (これには少し慣れが必要かもしれません)。このモデルは、変数間の影

響関係がある通り道を形成することから、**パス解析**と呼ばれています。 x_1 と x_2 , x_2 と x_3 だけピックアップしてみれば、回帰分析の関係になっていますから、この図に対応する方程式は

$$\begin{aligned}x_2 &= a_{21}x_1 + e_2 \\x_3 &= a_{32}x_2 + e_3\end{aligned}$$

と表すことができます。するとデータセットは

$$\mathbf{x} = \mathbf{Ax} + \mathbf{e}$$

のように表せます。これが構造方程式です。

今回のモデルの構造方程式を書き下すと次のようになります。

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

やや技巧的ですが、ベクトル表記にするために x_1 が係数行列にも加わっていつつ、最後の残差ベクトル $\mathbf{e}$ の中にも含まれているところに注目してください。変数 x_1 にはどこからも影響力が入ってきていない、モデルの中では生まれてこない変数です。誤差も同じく変数間関係の外にあるもので、こうした変数を**外生変数**と呼びます。残差ベクトルは外生変数ベクトルでもあるのです。それに対して、 $\mathbf{Ax}$ で表されるのは影響し合う関係で、 x_1 はここではお呼びでないため、 $\mathbf{A}$ の 1 行目は 0 が入っています。書いておきながら、影響力がないことを表現するというのがポイントです。なお x_2, x_3 はモデルの中での影響関係があるので、外生変数に対して**内生変数**と呼ばれます。

さて行列の方程式を変形すると、

$$\mathbf{Ix} = \mathbf{Ax} + \mathbf{e}$$

より

$$(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{e}$$

とすることができます。ここで $(\mathbf{I} - \mathbf{A})$ に逆行列が存在するならば、

$$\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{e}$$

と書き直すことができます。

ここで $(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{T}$ とおき、また $\mathbf{x}$ が平均偏差ベクトルであると考えれば、分散共分散行列は $\Sigma = E(\mathbf{xx}')$ で表せますから、

$$\Sigma = E(\mathbf{xx}') = E(\mathbf{T}\mathbf{e}(\mathbf{T}\mathbf{e})') = E(\mathbf{T}\mathbf{e}\mathbf{e}'\mathbf{T}') = \mathbf{T}E(\mathbf{e}\mathbf{e}')\mathbf{T}' = \mathbf{T}\Sigma_e\mathbf{T}'$$

となります。ここで Σ_e としているのは残差ベクトルの分散共分散行列で、対角項には分散が、非対角項には共分散が入っている行列ということになります。誤差は相関がゼロであると仮定できますので、この行列の要素は次のようになります。

$$\Sigma_e = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_{e2}^2 & 0 \\ 0 & 0 & \sigma_{e3}^2 \end{pmatrix}$$

続いて図 9.3 の例を見てみましょう

これも内生変数に対する方程式を書くことができ、

$$\begin{aligned}x_3 &= a_{31}x_1 + a_{32}x_2 + e_3 \\x_4 &= a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + e_4\end{aligned}$$

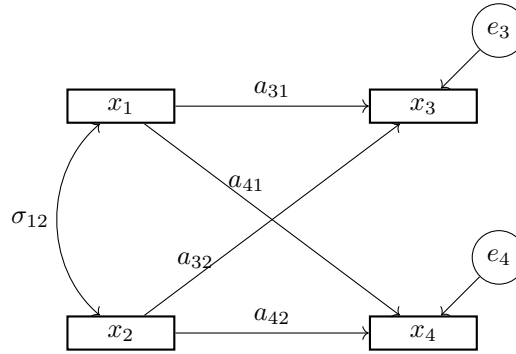


図 9.3 逐次モデル

これも行列で書くと

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

です。書き下して見ると

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & 0 & 0 \\ a_{41} & a_{42} & a_{43} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} x_1 \\ x_2 \\ e_3 \\ e_4 \end{pmatrix}$$

となります。同じロジックで $\Sigma = \mathbf{T}\Sigma_e\mathbf{T}'$ となりますが、このときの左辺には

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & 0 & 0 \\ \sigma_{12} & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_{e3} & 0 \\ 0 & 0 & 0 & \sigma_{e4} \end{pmatrix}$$

と、一部共分散が含まれていることに注意してください。

ともあれ一般に次の式、

$$\Sigma = (\mathbf{I} - \mathbf{A})^{-1}\Sigma_e(\mathbf{I} - \mathbf{A})^{-1'}$$

が成り立つということは、左辺で表されている観測変数による分散共分散行列と、モデルの係数を含んだ行列がイコールで結びついているということ、つまり観測変数の分散共分散という「変数から得られる情報」が、モデルの関数で表現されている、すなわち連立方程式になっているわけです。

9.3 測定方程式

測定方程式とは、潜在変数をどのようにして構成するかを書いている方程式のことです。

図 9.4 には単純な 1 因子モデルを描いてみました。これは何らかの構成概念 (因子) を測定するのに 3 つの項目を使っていることとなりますが、これを表す方程式は次のようになります。

$$\begin{aligned} x_1 &= \lambda_{11}f_1 + e_1 \\ x_2 &= \lambda_{21}f_1 + e_2 \\ x_3 &= \lambda_{31}f_1 + e_3 \end{aligned}$$

ここで f は潜在変数であり、因子分析の文脈で言えば因子得点にあたります。 λ_{jk} はパス係数、因子分析の文脈で言えば因子負荷量にあたります。これを行列の形で書けば

$$\mathbf{x} = \mathbf{\Lambda}\mathbf{f} + \mathbf{e}$$

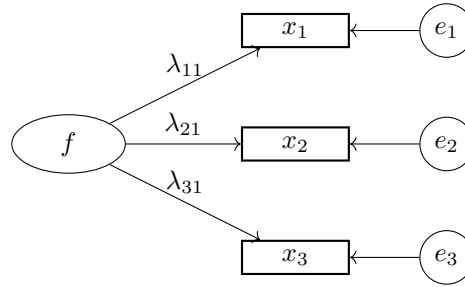


図 9.4 1 因子モデル

であり、既にみてきたように

$$\Sigma = E(xx') = \Lambda\Lambda' + \Psi$$

とすることができます。ここで Ψ は対角に誤差分散が入った対角行列です。右辺を書き下すと次のようになります。

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^2 + \sigma_1^2 & \lambda_{11}\lambda_{21} & \lambda_{11}\lambda_{31} \\ & \lambda_{21}^2 + \sigma_2^2 & \lambda_{21}\lambda_{31} \\ & & \lambda_{31}^2 + \sigma_3^2 \end{pmatrix}$$

これもまた、左辺の (データによる) 分散共分散行列を、モデルの方程式の形で書いたわけですから、これも方程式を解く形になるわけです。

9.4 複雑なモデルの場合

では次に、図 9.5 のような複雑なモデルでも方程式で表せることを示しましょう。

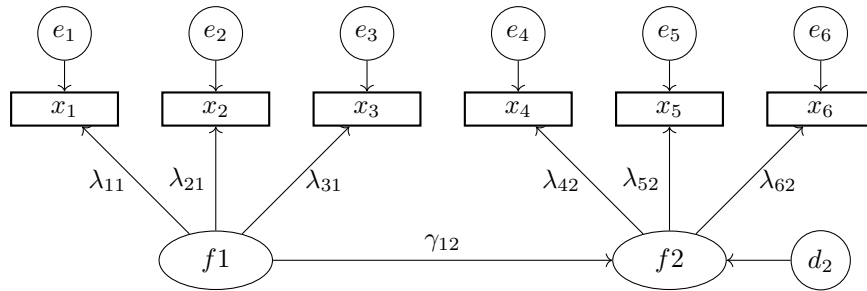


図 9.5 構造方程式と測定方程式。記号の意味は後述

ここで図 9.5 の係数について説明します。ギリシア文字 λ や γ がパス係数を表しています。パス係数の添字は、被説明変数の変数番号 j と、説明変数の変数番号 k をつかって λ_{jk} のようにかき、これで $k \rightarrow j$ の影響力の大きさを表していると思ってください。測定方程式の係数を λ で、構造方程式の係数を γ で表現しました。

それを踏まえて、モデルの方程式を書くと、次のようになります。

$$\begin{aligned} f2 &= \gamma_{21}f1 + d2 \\ x1 &= \lambda_{11}f1 + e1 \\ x2 &= \lambda_{21}f1 + e2 \\ x3 &= \lambda_{31}f1 + e3 \\ x4 &= \lambda_{42}f2 + e4 \\ x5 &= \lambda_{52}f2 + e5 \\ x6 &= \lambda_{62}f2 + e6 \end{aligned}$$

ややこしいだけで、書けるのは書けましたね。そして当然、これを行列の形に書いてみたくなってきたはず
です。

まず次のようなベクトル t および u を用意します。 t はモデルの中に含まれる全ての確率変数を持つもの
で、**構造変数ベクトル**といいます。 u は既に出てきましたが、**外生変数ベクトル**ですね。

$$t = \begin{pmatrix} f_1 \\ f_2 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}, u = \begin{pmatrix} f_1 \\ d_2 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{pmatrix}$$

これと係数行列 A から、 $t = At + u$ を考えましょう。具体的に書いてみると、次のようになっています。

$$\begin{pmatrix} f_1 \\ f_2 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{42} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{52} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{62} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} + \begin{pmatrix} f_1 \\ d_2 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{pmatrix}$$

左辺のベクトルに f_1 が入っているところがポイントで、これは方程式的には $f_1 = f_1$ というだけの式なの
ですが、これを入れたおかげで全体を整合的に表現できましたね。

先ほどと同じように、行列の式を変形して

$$It = At + u$$

$$(I - A)t = u$$

ここで $T = (I - A)^{-1}$ とすると、

$$t = Tu$$

です。

ここで、構造変数ベクトルから観測変数だけ抜き出すことを考えます。**選択方程式**とよばれる次の操作、

$$x = Ft$$

によって、左辺を観測変数だけにしてしまいます。この操作は次のようなことをしています。

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

変数を抜き出すのも行列の操作でやるわけですね。面白くないですか、これ。

ともかくこれをつかって、先ほどの $t = Tu$ との関係から、

$$x = FTu$$

であることがわかります。左辺は観測変数だけになりましたから、分散共分散構造を構成することを考えましょう。

$$\begin{aligned}\Sigma &= E(xx') = E(FTu(FTu)') \\ &= E(FTuu'T'F') \\ &= E(FTuu'T'F') \\ &= FTE(uu')T'F' \\ &= FT\Sigma_uT'F'\end{aligned}$$

ここで式の中にある Σ_u は誤差などの未知数で作った分散共分散行列になっています。誤差は他の変数との共分散がないと考えられるので、これは結局対角行列です。

$$\Sigma_u = \begin{pmatrix} V(f_1) & & & & & & \mathbf{0} \\ & V(d_2) & & & & & \\ & & V(e_1) & & & & \\ & & & V(e_2) & & & \\ & & & & V(e_3) & & \\ & & & & & V(e_4) & \\ & & & & & & V(e_5) \\ \mathbf{0} & & & & & & & V(e_6) \end{pmatrix}$$

さあややこしい話はここまでにして、何を考えていたのかをもう一度考えてみましょう。

この式の左辺は Σ , すなわちデータからできる分散共分散行列であり, 具体的に数字の計算のできるデータセットです。この式の右辺は $\mathbf{F} \mathbf{T} \Sigma_u \mathbf{T}' \mathbf{F}'$ であり, 係数と潜在変数の分散からなる方程式です。左辺にデータ, 右辺にモデルで表した分散共分散行列があり, これがイコールで結びついているのですから, この連立方程式を解けば, 係数を求めることができます。やったね!

例題 23

図 9.6 の構造変数ベクトル, 外生変数ベクトル, 係数行列を考えてみましょう。

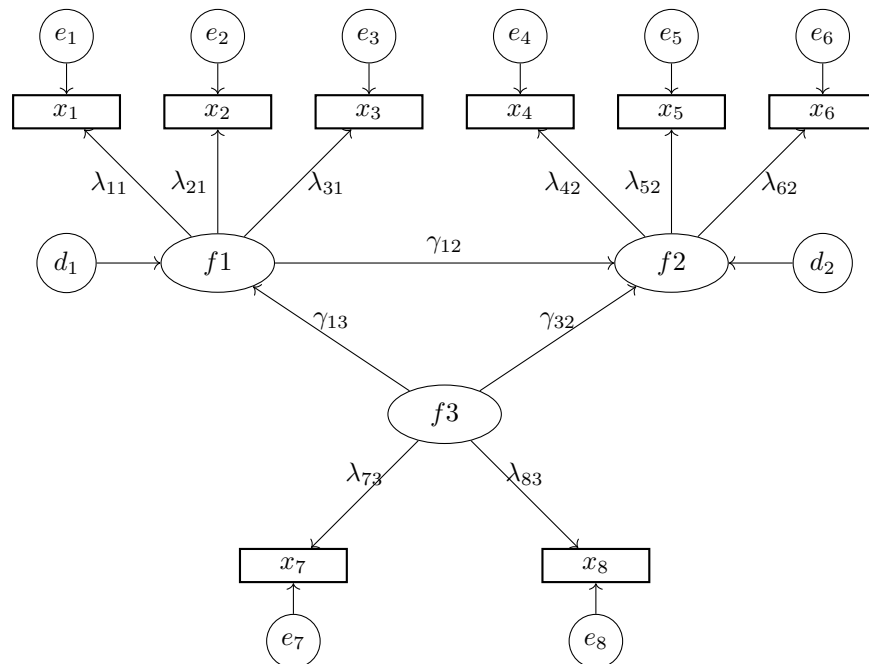


図 9.6 練習問題;ベクトルと行列でかけるかな

9.5 構造方程式モデルによる未知数の推定

9.5.1 識別の問題

さて、ここまでデータとモデルの方程式を書いてきましたが、果たしてこの方程式は解けるのでしょうか。セクション 9.3 でやってみた測定方程式モデルを考えてみましょう。

最終的には、 $\Sigma = E(\mathbf{x}\mathbf{x}') = \mathbf{\Lambda}\mathbf{\Lambda}' + \Psi$ のかたちから、モデル式は次のように書けるのです。

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^2 + \sigma_1^2 & \lambda_{11}\lambda_{21} & \lambda_{11}\lambda_{31} \\ \lambda_{21}^2 + \sigma_2^2 & \lambda_{21}\lambda_{31} & \lambda_{31}^2 + \sigma_3^2 \end{pmatrix}$$

一方、この式の左辺はデータから計算された分散共分散行列、右辺はモデルから計算された分散共分散行列です。3 つの変数からなる分散共分散行列ですから、 3×3 の行列ができ、対称行列ですので重複を除くと、

$$\begin{aligned} s_1^2 &= \lambda_{11}^2 + \sigma_1^2 \\ s_2^2 &= \lambda_{21}^2 + \sigma_2^2 \\ s_3^2 &= \lambda_{31}^2 + \sigma_3^2 \\ s_{12} &= \lambda_{11}\lambda_{21} \\ s_{13} &= \lambda_{11}\lambda_{31} \\ s_{23} &= \lambda_{21}\lambda_{31} \end{aligned}$$

という関係になります。データからは 6 つの数字が提供され、右辺の方程式には $\lambda_{11}, \lambda_{21}, \lambda_{31}, \sigma_1, \sigma_2, \sigma_3$ の 6 つの未知数がありますから、この連立方程式は解けます。とくに、未知数と既知数の数が一致しているので、**丁度識別 (just identification)**、あるいは単に**識別可能 (identifiable)** といいます。方程式が解けるということは、解が一意に定まるということ、それ以外の答えがないということになります。SEM の応用シーンである、人文社会科学の領域で、答えがひとつしかないということに違和感があるかもしれません。

とは言え、少し話を進めます。次に、観測変数が 2 つで潜在変数が 1 つの場合を考えてみましょう。これはデータと方程式の関係を考えると、

$$\begin{aligned} s_1^2 &= \lambda_{11}^2 + \sigma_1^2 \\ s_2^2 &= \lambda_{21}^2 + \sigma_2^2 \\ s_{12} &= \lambda_{11}\lambda_{21} \end{aligned}$$

となり、右辺に 4 つの未知数 ($\lambda_{11}, \lambda_{21}, \sigma_1, \sigma_2$) となりますので、解けない方程式です。そのような場合は**識別不可能** (解が求められない) といわれます。

例題 24

観測変数が 5 つ、潜在変数が 2 つの場合は識別できるでしょうか? 判定してみてください。

さて、一般に観測変数が n 個あれば、そこから得られる情報 (既知数) は n 個の分散と $n(n-1)/2$ 個の共分散あります。変数が多くなってくると、モデルで想定するパラメータ (未知数) の数が増えるスピードよりも、データから提供される既知数の増えるスピードの方が飛躍的に多くなります。連立方程式の観点から言えば「解なし」状態なのです。

未知数よりも既知数が多い連立方程式の例として、たとえば次のようなものがあったとします。

$$\begin{cases} x + y &= 5 \\ 2x + y &= 4 \\ x + 2y &= -1 \end{cases}$$

この場合、未知数は x, y の 2 つです。これの係数行列 $\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \end{pmatrix}$ の階数は 2 であるのに対し、拡大係数行列 $\begin{pmatrix} 1 & 1 & 5 \\ 2 & 1 & 4 \\ 1 & 2 & -1 \end{pmatrix}$ の階数が 3 ですから、セクション 3.3.2, Pp.50 でみたように、この方程式は解がありません。しかし、上の 2 本の式、すなわち $x + y = 5$ と $2x + y = 4$ だけであれば、 $x = -1, y = 6$ ですし、下の 2 本の式、 $2x + y = 4$ と $x + 2y = -1$ だけであれば、 $x = 3, y = -2$ になります。残る 2 本を使えば $x = 11, y = -6$ になります。左辺が SEM のモデルとしての式、右辺がデータから得られた値と考えれば、モデルとデータが矛盾している状態であり、どこかが間違えている、あるいは答えをどれにしようか迷える解の自由度がある状況だといえます。一般的な SEM のモデルのほとんどは、識別不能でも丁度識別でもない、この過剰識別状態にあるといえます。

連立方程式は一意に解が求められなければ意味がないのでは、と思われるかもしれませんが、SEM が多変量データ解析の文脈で使われるモデルであることを考えると、むしろ解の自由度があること、過剰識別であることの方がよいのです (谷 & 平島, 2018)。たとえば回帰分析をする時のことを考えてみてください。回帰分析は散布図に回帰直線を引くことが目的ですが、直線は空間上の 2 点が定まれば一意に引くことができます。しかし 2 件だけのデータで回帰係数を求め、「これ以外の答えはあり得ない」というのはあまりにも不自然でしょう。データの揺らぎ、誤差を想定した上で、なるべくもっともらしい線を引こう、と考えるほうが自然です。多数のデータ点があるというのは、多数の相互に矛盾するデータがあるということかもしれませんが、それらを総合的に判断して、全体的に当てはまる係数を求めようとします。「総合的に判断」とか「全体的に当てはまる」というのは、言葉を変えるなら、「誤差が最小になるように」とか、「誤差が正規分布からきているという仮定に最も適合するように」と表現できます。これが最小二乗法や最尤法であり、合理的な判断基準を外挿的に加えて答えを求めるのです。そうすることで、相互に矛盾する情報の情報源や、モデルの適切さを評価できるようにもなります。

構造方程式モデリングは、変数間関係を方程式で表し、データの分散共分散行列と、モデルの分散共分散行列を照らし合わせつつ、統計的に最も良い答え (モデルとデータのずれが最も少ないとか、尤度が最も高いとか) を考えていく手法です。我々がデータから手に入れられる情報の上限が分散・共分散であるなら、その素性を隅々まで書き込んで方程式を解いていく。そういう意味で構造方程式モデリングは「究極の」分析手法なのです。

第 10 章

まとめにかえて

10.1 さまざまな多変量解析

本書の冒頭で語ったことですが、心理統計を学ぶのになぜ線形代数が必要なのでしょう。

線形代数は数字のリストと方程式、そしてそれらの作る空間の理解を与えてくれます。方程式は計算問題として、空間の問題は抽象的な話をイメージするための方便、ということでもいいのですが、これを繋げてくれる理論的根拠があることで、**視座が増える**ことが一番の利点かな、と思います。

心理学を学ぶものにとって、心理統計はツールであり使えれば良いものです。料理人が腕を磨くのは調理の腕であって、包丁を研いだり台所を片付けたりする腕を磨くことの意義がピンとこないというようなところでしょうか。しかもそれが数字の羅列でコーティングされているのですから、余計に近寄りたくない、と思ってしまうのも当然だと思います。

しかしものを見る観点が増えると、理解が深まるということもあるでしょう。理解するということの 1 つの側面は、そのことについてさまざまな角度から自らの言葉で語ることができるようになる、ということかと思えます。データを空間的な広がりとして理解し、例えば

- 分散分析はその空間を直和に分解していくこと
- 回帰分析はある変数ベクトルに他の変数ベクトルを射影すること
- 因子分析はデータ空間の基底を見つけること
- 構造方程式モデリングはデータ空間とモデル空間を合わせること

というような理解ができれば、分析モデルのことがちょっと好きになるし、「間違った使い方」をするようなことも減るのではないのでしょうか。

そしてこれらのイメージができると、本書で扱わなかったその他の分析方法についても、イメージを広げていくことができます。

因子分析のところで、データ空間の規定を見つけるための固有値分解という話をしました。相関行列から始まって、共通性の推定といったちょっとしたテクニックが入っていましたが、基本的には空間座標を作るためのものだったわけです。実はその他の分析方法も、

- 主成分分析は分散共分散行列の固有値分解
- 因子分析は相関行列の固有値分解
- 多次元尺度構成法は距離行列の固有値分解

という特徴があります。最後の**多次元尺度構成法**は本書では扱いませんでしたが、計算的なメカニズムは固

有値分解で、元になるデータが距離行列という違いだけである、ともいえます^{*1}。

要するに、データの特徴を表す正方行列さえ作られれば、そこから軸を取り出す＝意味を見出すことができる、というわけです。さてそうすると、ここまではずっと $m \times m$ サイズの正方行列、すなわち変数間関係の行列を分解してきました。ここで元のデータ行列を転置して、 $n \times n$ サイズの正方行列を構成すれば、ケース間の共通次元を見出すこともできるのではないか、と思いつきますよね。これは因子分析の **Q 技法** と呼ばれる手法になりますが、これを使って「個人に共通する因子」を考えることも可能なわけです。

最近では時系列的なデータ、そのほか色々な角度から集められたデータが手に入りますが、なんらかの側面で正方行列を作ってしまうと、あとは分析し放題！それがただの数字の羅列だと思ってしまうと、数学的には色々な計算が可能です。さらにまた、ただの数字の羅列にわたしたちは「人の心」や解釈できる「意味」を見出したわけですから、数字の集め方や解釈の仕方に敬意を払って、丁寧に分析していこうという気持ちにもなれるのではないのでしょうか。

ついでに、「正方行列じゃないとだめか、長方形（矩形）の場合はどうなるのか」と言いたくなる人もいるかもしれません。それでもいいんです！固有値分解は正方行列から始めるしかありませんが、矩形行列の場合は**特異値分解 (Singular value decomposition)** を使うことで行列の分解が可能です。たとえば行・列それぞれに別の意味をもたせる、行空間と列空間にデータを射影するという分析方法は**数量化 III 類**として知られています^{*2}。このような方法をとると、変数やケースの考え方も、もっと多角的に見ることができるかと思えます。

ところで、(重) 回帰分析はある変数空間に他の変数を射影することだ、と言いました。この手法はいわば、多変数 $\rightarrow$ 1 変数の関数関係を考えることでした。しかし人間に関わる変数は、インプットもさまざまあれば、その出力・表現系・アウトプットも多岐にわたります。この多変数 $\rightarrow$ 多変数の関数関係を考えるのに、線形でない関数関係も許容したモデルが今流行の**機械学習**というやつです。ここでも背後にはデータの空間を射影するという数学的世界があります。線形代数と違って非線形な世界を考えることになりますが、より複雑なデータを扱うときにも線形代数の知識がきっと、お役に立つことと信じています。

10.2 数学書案内

私は数学者ではなく、心理学者であろうと自認していますので、数学のテキストを書いたり教壇に立ったりするのは正直怖いなあとと思っています。数学という世界は整合的で、広く広く、どこまでも広がっていきますから、私の偏った理解や狭い見識で語っていると、ミスリーディングなことをしてしまうのではないか、という不安に苛まれています。

本書でも、たとえば逆行列の計算は 2×2 でやめてしまいました。行列を小行列に分割したりすることもスキップしています。というのも、やはり心理学者の使う範囲に限定し、みなさんが必要としているところに絞ってお話したほうが良いだろうと思ったからです。それしかできないというのもありますし、言い訳してんじゃねえよ、といわれたらごめんなさいとしか言えないです。

このテキストは私がわかる範囲で、私が教えられる範囲で、私が学んできた範囲での知識を書いたものであり、私オリジナルな定理も定義も含まれるものではありません。そういう意味では、膨大なコピペのかたまりでもあります。これから先、心理学の新しい理論、モデル、数学を作らんとするみなさんは、是非以下に紹介する優れたテキストを読んで、さらに幅広く学んでいただければと思います。

^{*1} これはもちろん計量多次元尺度法に限った話で、多次元尺度構成法は心理測定、尺度構成の観点から他にも面白い話がいっぱいあるのですが。

^{*2} ここでは数量化 III 類の名前を挙げましたが、数学的には**双対尺度法**や**対応分析**も同じく特異値分解であり、数量化 III 類と数学的には同等なものです。

■**数学を学ぶ** 数学の知識は使わないと忘れてしまうので、折に触れて学び直し、見直すことが重要です。中野 (2017) は美文、積分、対数、指数、その他さまざまな基本的数学がコンパクトにまとまっているのでお勧めです。線形代数に限って言えば、竹内 (2010) はその名の通り高校生にもわかるような口調で書いてあるのでわかりやすかったと思います。そしてより広く、数学的な線形代数を学ぶという意味では村上 et al. (2016) がいいでしょう。版を重ねていることから、多くの人にとって良いものであることは間違いありません。

さらに本腰を入れてやってやろう、という人は是非永田 (2005) にチャレンジしてみてください。少し難易度は高くなりますが、この一冊を学び通せば「自分は数学ができるぞ」といってもいいんじゃないでしょうか。

■**統計を学ぶ** 残念ながら岡太 (2008) は、ものすごくいい本なのですが絶版になってしまったようです。データを要素ごとに、計算ステップを書きながら論じてくれているのは珍しく、この本はとても良い入門書になったのですが。図書館などで探してみてください。

心理統計というみでは、南風原 (2002) が一番いいでしょう。サブタイトルにある「統合的理解のために」というのはまさに本書も目指したところで、数学的な視点を持つことができればいろいろつながって見える、その喜びを感じることできる一冊です。

多変量解析、とくに因子分析については芝 (1979) が私のテキストでした。行列の基本から書いてくれているので、とてもわかりやすいです。より進んだ内容としては、柳井 et al. (1990) がいいでしょう。

最後に紹介した、矩形行列の分析については、その計算方法から丁寧に解説がある西里 (2010) が良いと思います。これを開くと、もしかするとみたことのない景色が広がっているかもしれません。

付録 A

行列操作の R コード

A.1 行列の基本変形

行列の計算は、一度は手計算をして計算方法を「身体化」した方が良いと思いますが、単純な計算でも繰り返しが多くて人間がやるとどうしてもミスが生じます。そこで R など計算機を使つての検算をすることになると思います。本書執筆時に、行列の基本変形 (セクション 3.2, Pp.39) に言及しましたが、その操作をするコードを以下に示します。

code : A.1 ある行に 0 ではない数字を掛ける

```
1 pic <- function(size, target, constant){
2   vec = rep(1, size)
3   vec[target] = constant
4   return(diag(vec))
5 }
```

code : A.2 ある行に他の行を何倍かしたものを加える

```
1 pij <- function(size, from, to){
2   vec = rep(1, size)
3   vec[from] = 0
4   vec[to] = 0
5   mat = diag(vec)
6   mat[from, to] = 1
7   return(mat)
8 }
```

code : A.3 2つの行を入れ替える

```
1 pij <- function(size, from, to){
2   vec = rep(1, size)
3   vec[from] = 0
4   vec[to] = 0
5   mat = diag(vec)
6   mat[from, to] = 1
7   return(mat)
8 }
```

これらのコードはいずれも、引数として行列のサイズを入れます。関数の中では R の `diag` 関数で単位行列を作ってから、必要な要素だけ入れ替えて返すようにしています。例題にあった、拡大係数行列

$$A = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & -1 & 3 \\ 3 & 1 & 4 & 1 \end{pmatrix}$$

を変形する操作ですが、最終的には

$$P_{32}(-2)P_{31}(1)P_3(1/2)P_{23}(-7)P_{21}(2)P_2(1/14)P_{13}(-3)P_{12}(-5)A = I$$

という計算をすることになったのでした。これを今回のコードを使ってやると、以下のようになります。

code : A.4 拡大係数行列の変形

```
1 A <- matrix(c(1,5,3,-2,4,1,-5,3,-3,3,1,6),ncol=4)
2 size <- 3
3 P1 <- pijc(size,2,1,-5)
4 P2 <- pijc(size,3,1,-3)
5 P3 <- pic(size,2,1/14)
6 P4 <- pijc(size,1,2,2)
7 P5 <- pijc(size,3,2,-7)
8 P6 <- pic(size,3,-1/2)
9 P7 <- pijc(size,1,3,1)
10 P8 <- pijc(size,2,3,-2)
11 P8 %%% P7 %%% P6 %%% P5 %%% P4 %%% P3 %%% P2 %%% P1 %%% A
```

R の行列の積は `%%` という記号になることに注意してください。

また、全ての操作を結合した行列 P が逆行列に一致することを確認するために、R の逆行列関数 `solve` を使って算出したものと見比べてみてください (R の出力 A.1)^{*1}。

code : A.5 逆行列の算出

```
1 I <- diag(rep(1,size))
2 P8 %%% P7 %%% P6 %%% P5 %%% P4 %%% P3 %%% P2 %%% P1 %%% I
3 solve(matrix(c(1,5,3,-2,4,1,-5,3,-3),ncol=3))
```

R の出力 A.1: 逆行列の検算結果

```
> P8 %%% P7 %%% P6 %%% P5 %%% P4 %%% P3 %%% P2 %%% P1 %%% I
      [,1]      [,2] [,3]
[1,]  0.5357143  0.3928571 -0.5
[2,] -0.8571429 -0.4285714  1.0
[3,]  0.2500000  0.2500000 -0.5
> solve(matrix(c(1,5,3,-2,4,1,-5,3,-3),ncol=3))
      [,1]      [,2] [,3]
[1,]  0.5357143  0.3928571 -0.5
[2,] -0.8571429 -0.4285714  1.0
[3,]  0.2500000  0.2500000 -0.5
```

^{*1} `solve` は「解く」という意味です。方程式を解く、すなわち逆行列を求めると言うことです！

A.2 固有値の近似演算

固有値と固有ベクトルの計算は基本的に機械に任せていけばよく、R では `eigen` という関数で固有値と固有ベクトルを返してくれます。ここではちょっとマニアックに、どうやって計算しているのかについて考えてみたいと思います。

固有値や固有ベクトルの計算は固有方程式を解くことにあり、変数が多くなると近似的な計算をすることになり、計算機科学の分野ではさまざまな近似計算法が提案されています。ここではそのうちの 2 つをご紹介します。

A.2.1 パワー法

まずはパワー法から。ここでのパワー power とは力という意味ではなく、累乗の方の意味です。

ある行列 A の固有値を $\lambda_1, \lambda_2, \dots, \lambda_n$ とします。かつ、これらが絶対値の大きい順に並んでいるとしましょう。すなわち、

$$\lambda_1 > \lambda_2 > \dots > \lambda_n$$

です。またこの各固有値に対応する固有ベクトルをそれぞれ $v_1, v_2, \dots, v_n$ と表します。任意のベクトル v は $v_1, v_2, \dots, v_n$ の線形結合で表すことができるという性質を利用して、

$$v = c_1 v_1 + c_2 v_2 + \dots + c_n v_n$$

と書くことができます。このとき、 c_j はあるスカラーです。さて、この v を行列 A にかけてどうなるかというと、

$$Av = A(c_1 v_1) + A(c_2 v_2) + \dots + A(c_n v_n)$$

$$= c_1 Av_1 + c_2 Av_2 + \dots + c_n Av_n$$

v はいずれも固有ベクトルだから

$$= c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \dots + c_n \lambda_n v_n$$

ここで λ_1 を括り出すと、

$$= \lambda_1 (c_1 v_1 + c_2 (\lambda_2/\lambda_1) v_2 + \dots + c_n (\lambda_n/\lambda_1) v_n)$$

となります。このようにしながら、 A を累乗していく、すなわち $AAA \dots Av$ としていくと、

$$A^p v = \lambda_1^p c_1 v_1 + c_2 (\lambda_2/\lambda_1)^p v_2 + \dots + c_n (\lambda_n/\lambda_1)^p v_n$$

となっていきます。ここで p 回の以上を繰り返しているわけですが、 $\lambda_1 > \lambda_2$ のように最初の固有値 λ_1 は最大のものですから、 $c_2 (\lambda_2/\lambda_1)^p v_2$ は $(\lambda_2/\lambda_1)^p$ がどんどん 0 に近づいていきます。以下の $c_n (\lambda_n/\lambda_1)^p v_n$ についても同様で、繰り返し続けると

$$A^p v \approx \lambda_1^p c_1 v_1$$

となることがわかります。これを利用して、固有値を求める式は

$$\lambda_1 = \lim_{p \rightarrow \infty} \frac{(A^{p+1}v)_r}{(A^p v)_r}$$

とすることができます。式中の r は分子・分母のベクトルの、第 r 番目の要素を指しています。

これがパワー法による固有値の求め方です。この方法では最大固有値しか求められませんが、第2固有値は第1固有値の要素を元の行列から抜き去った行列、すなわち $A - v_1 v_1'$ を作り、第1固有値の次元情報が抜けた行列の中の最大固有値、というようにして求めていくことができます。

数値例で見てみましょう。

4×4 の行列 A を次のような行列とします。

$$A = \begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix}$$

これに、まずあるベクトル $v = \begin{pmatrix} 1.0 \\ 0.0 \\ 0.0 \\ 0.0 \end{pmatrix}$ をかけると、

$$Av = \begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix} \begin{pmatrix} 1.0 \\ 0.0 \\ 0.0 \\ 0.0 \end{pmatrix} = \begin{pmatrix} 1.0 \\ 0.3 \\ 0.4 \\ 0.9 \end{pmatrix}$$

この新しくできたベクトルを $v^\dagger$ として、もう一度 A にかけて

$$Av^\dagger = \begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix} \begin{pmatrix} 1.0 \\ 0.3 \\ 0.4 \\ 0.9 \end{pmatrix} = \begin{pmatrix} 2.06 \\ 1.16 \\ 1.58 \\ 2.20 \end{pmatrix}$$

となります。この新しくできたベクトル $v^\dagger$ の要素は大きいものでは2.0を超えているので、最大値2.2を取り出して少しスケールを整えてみましょう。

$$w = 2.2 \begin{pmatrix} 0.936 \\ 0.527 \\ 0.718 \\ 1.000 \end{pmatrix}$$

この大きさの整えられたベクトル w を次の計算に用いることにします。

$$Aw = \begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix} \begin{pmatrix} 0.936 \\ 0.527 \\ 0.718 \\ 1.00 \end{pmatrix} = \begin{pmatrix} 2.282 \\ 1.567 \\ 2.056 \\ 2.556 \end{pmatrix}$$

ここでも同様に、新しくできたベクトルは最大要素で割ってサイズを整え、 $w^\dagger = 2.556 \begin{pmatrix} 0.893 \\ 0.613 \\ 0.804 \\ 1.000 \end{pmatrix}$ として次の

計算に用いることにします。

以下のプロセスは同様なので、数値の変化だけ追いかけていきましょう。

$$Aw^\dagger = \begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix} \begin{pmatrix} 0.893 \\ 0.613 \\ 0.804 \\ 1.00 \end{pmatrix} = \begin{pmatrix} 2.298 \\ 1.683 \\ 2.168 \\ 2.612 \end{pmatrix} = 2.612 \begin{pmatrix} 0.880 \\ 0.644 \\ 0.830 \\ 1.00 \end{pmatrix}$$

$$\begin{aligned}
&\rightarrow \begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix} \begin{pmatrix} 0.880 \\ 0.644 \\ 0.830 \\ 1.00 \end{pmatrix} = \begin{pmatrix} 2.305 \\ 1.724 \\ 2.204 \\ 2.631 \end{pmatrix} = 2.631 \begin{pmatrix} 0.876 \\ 0.655 \\ 0.838 \\ 1.00 \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix} \begin{pmatrix} 0.876 \\ 0.655 \\ 0.838 \\ 1.00 \end{pmatrix} = \begin{pmatrix} 2.308 \\ 1.737 \\ 2.216 \\ 2.637 \end{pmatrix} = 2.637 \begin{pmatrix} 0.875 \\ 0.659 \\ 0.840 \\ 1.00 \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix} \begin{pmatrix} 0.875 \\ 0.659 \\ 0.840 \\ 1.00 \end{pmatrix} = \begin{pmatrix} 2.309 \\ 1.741 \\ 2.220 \\ 2.639 \end{pmatrix} = 2.639 \begin{pmatrix} 0.875 \\ 0.660 \\ 0.841 \\ 1.00 \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix} \begin{pmatrix} 0.875 \\ 0.660 \\ 0.841 \\ 1.00 \end{pmatrix} = \begin{pmatrix} 2.309 \\ 1.743 \\ 2.221 \\ 2.640 \end{pmatrix} = 2.640 \begin{pmatrix} 0.875 \\ 0.660 \\ 0.841 \\ 1.00 \end{pmatrix}
\end{aligned}$$

このように、取り出される数値が徐々に一定の値 (今回は $\rightarrow 2.640$) に収束していくし、ベクトルの方も一定の値に収まっていくことになります。この変化量がとても小さくなると「十分に収束した」とされ、取り出された値=固有値、ベクトル=固有ベクトル、とすることができます。

この計算をする R コードを添付します。

code : A.6 パワー法による最大固有値の算出関数

```

1 power <- function(A, max.itr = 100, eps = 1e-20) {
2   u <- rep(1, length = nrow(A))
3   v <- rep(0, length = nrow(A))
4   e.val <- 0
5   for (i in 1:max.itr) {
6     v <- A %*% u
7     if (abs(max(v) - e.val) < eps) {
8       break
9     }
10    e.val <- max(v)
11    e.vec <- u <- v / e.val
12  }
13  return(list(Eval = e.val, Evec=e.vec))
14 }

```

A.2.2 QR 法

行列を $A = QR$ の形に分解することを QR 分解といいます。ここで Q は直交行列, R は上三角行列です。上三角行列というのは、非ゼロ要素が体格の右上半分だけにある行列、すなわち

$$\begin{pmatrix} * & * & * & \cdots & * & * \\ 0 & * & * & \cdots & * & * \\ 0 & 0 & * & \cdots & * & * \\ 0 & 0 & 0 & \cdots & * & * \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & * & * \\ 0 & 0 & 0 & \cdots & 0 & * \end{pmatrix}$$

という形の行列です。ここで*には任意の数値です。

この分解を使って、

1. A を直交行列 Q と上三角行列 R の積に分解する。すなわち、 $A^{(t)} = QR$
2. その結果を逆順に掛ける。 $A^{(t+1)} = RQ$

という2つの操作を反復して、固有値を求めることになります。これらの計算は、固有値を変化させずに行列の形を変形するアルゴリズムであり、このような反復を繰り返すと非対角要素の絶対値は次第に小さくなり、 t が十分に大きくなれば $A^{(t)}$ は対角行列に収束し、その対角項が固有値となるという技です。

A を Q と R との積に分解する方法はいくつかありますが、ひとつはセクション 5.1.1, Pp.70 で解説したグラム・シュミットの正規直交化法が使えます。

この方法では、行列 A の第一列 a_1 , 第二列 a_2 , $\dots$, 第 n 列 a_n を元に、行列 Q の第一列 q_1 , 第二列 q_2 , $\dots$, 第 n 列 q_n を作っていきます。このとき Q は

- q_i と q_j が互いに直交する。すなわち、 $(q_i, q_j) = 0$ 。ただし $i \neq j$
- 各 q_i の長さは1である。すなわち、 $(q_i, q_i) = 1$

の条件を満たすように作っていきます。

そこで、

$$\begin{aligned} b_1 &= a_1 \\ b_2 &= (a_2 \text{ から } a_1 \text{ に平行な成分を抜き取ったもの}) \\ b_3 &= (a_3 \text{ から } a_1, a_2 \text{ に平行な成分を抜き取ったもの}) \\ &\vdots \\ b_n &= (a_n \text{ から } a_1, \dots, a_{n-1} \text{ に平行な成分を抜き取ったもの}) \end{aligned} \quad (\text{A.1})$$

という要領で、 $b_1, b_2, \dots, b_n$ を作り、各 b_i をその長さ $\|b_i\|$ で割って q_i にすることになります。

数値例を見てみましょう。行列 A を QR の形に分解してみます。

$$\begin{pmatrix} 1.0 & 0.3 & 0.4 & 0.9 \\ 0.3 & 1.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 1.0 & 0.7 \\ 0.9 & 0.4 & 0.7 & 1.0 \end{pmatrix} = \begin{pmatrix} 0.697 & -0.286 & -0.338 & -0.565 \\ 0.209 & 0.903 & -0.374 & 0.025 \\ 0.279 & 0.299 & 0.855 & -0.319 \\ 0.627 & -0.116 & 0.120 & 0.761 \end{pmatrix} \begin{pmatrix} 1.435 & 0.808 & 1.101 & 1.533 \\ 0.000 & 0.920 & 0.555 & 0.197 \\ 0.000 & 0.000 & 0.617 & 0.265 \\ 0.000 & 0.000 & 0.000 & 0.039 \end{pmatrix}$$

そしてこれから、 $A^{(1)} = RQ$ を作ります。

$$\begin{pmatrix} 2.437 & 0.470 & 0.338 & 0.025 \\ 0.470 & 0.974 & 0.153 & -0.005 \\ 0.338 & 0.153 & 0.559 & 0.005 \\ 0.025 & -0.005 & 0.005 & 0.030 \end{pmatrix}$$

この操作を繰り返して、 $A = QR \rightarrow RQ = A^{(1)} \rightarrow A^{(1)} = QR \rightarrow \dots$ を続けてみたいと思います。

$$\begin{aligned} A^{(2)} &= \begin{pmatrix} 2.620 & 0.180 & 0.067 & 0.00 \\ 0.180 & 0.859 & 0.037 & 0.00 \\ 0.067 & 0.037 & 0.492 & 0.00 \\ 0.000 & 0.000 & 0.000 & 0.03 \end{pmatrix} \\ A^{(3)} &= \begin{pmatrix} 2.638 & 0.059 & 0.012 & 0.00 \\ 0.059 & 0.844 & 0.018 & 0.00 \\ 0.012 & 0.018 & 0.488 & 0.00 \\ 0.000 & 0.000 & 0.000 & 0.03 \end{pmatrix} \end{aligned}$$

$$A^{(4)} = \begin{pmatrix} 2.640 & 0.019 & 0.002 & 0.00 \\ 0.019 & 0.843 & 0.010 & 0.00 \\ 0.002 & 0.010 & 0.487 & 0.00 \\ 0.000 & 0.000 & 0.000 & 0.03 \end{pmatrix}$$

$$A^{(5)} = \begin{pmatrix} 2.640 & 0.006 & 0.000 & 0.00 \\ 0.006 & 0.843 & 0.006 & 0.00 \\ 0.000 & 0.006 & 0.487 & 0.00 \\ 0.000 & 0.000 & 0.000 & 0.03 \end{pmatrix}$$

どんどん対角化されていくのが見えてきましたね。これを 10 回もすると、

$$A^{(10)} = \begin{pmatrix} 2.64 & 0.000 & 0.000 & 0.00 \\ 0.00 & 0.843 & 0.000 & 0.00 \\ 0.00 & 0.000 & 0.487 & 0.00 \\ 0.00 & 0.000 & 0.000 & 0.03 \end{pmatrix}$$

となり、対角項に 4 つの固有値が並んでいるのが見えると思います。

最後に正方形行列 A を与えると、QR 分解する R コードを添付します。

code : A.7 QR 分解する R コード

```
1 QRdecompose <- function(A){
2   Q = matrix(0,ncol=ncol(A),nrow=nrow(A))
3   R = matrix(0,ncol=ncol(A),nrow=nrow(A))
4   for(i in 1:N){
5     b <- A[,i]
6     for(j in 1:i){
7       if(j<i){
8         R[j,i] <- t(A[,i])%*%Q[,j]
9         b <- b - R[j,i]*Q[,j]
10      }
11    }
12    R[i,i] <- sqrt(sum(b^2))
13    Q[,i] <- b/R[i,i]
14  }
15  return(list(Q=Q,R=R))
16 }
```


付録 B

ギリシア文字一覧

ギリシア文字ってカッコいいけど、読み方わからない・・・という人のための一覧。ついでに $\text{T}_\text{E}\text{X}$ 表記も。

表 B.1 ギリシア文字とその読み方

読み方	大文字	小文字	英語表記	$\text{T}_\text{E}\text{X}$ 表記
アルファ	A	α	alpha	<code>\alpha</code>
ベータ	B	β	beta	<code>\beta</code>
ガンマ	Γ	γ	gamma	<code>\gamma</code>
デルタ	Δ	δ	delta	<code>\delta</code>
エプシロン	E	ε	epsilon	<code>\varepsilon</code>
ゼータ	Z	ζ	zeta	<code>\zeta</code>
イータ	H	η	eta	<code>\eta</code>
シータ	Θ	θ	theta	<code>\theta</code>
イオタ	I	ι	iota	<code>\iota</code>
カッパ	K	κ	kappa	<code>\kappa</code>
ラムダ	Λ	λ	lambda	<code>\lambda</code>
ミュー	M	μ	mu	<code>\mu</code>
ニュー	N	ν	nu	<code>\nu</code>
クサイ	Ξ	ξ	xi	<code>\xi</code>
オミクロン	O	$\omicron$	omicron	<code>\mathrm{o}</code>
パイ	Π	π	pi	<code>\pi</code>
ロー	R	ρ	rho	<code>\rho</code>
シグマ	Σ	σ	sigma	<code>\sigma</code>
タウ	T	τ	tau	<code>\tau</code>
ウプシロン	U	υ	upsilon	<code>\upsilon</code>
ファイ	Φ	ϕ	phi	<code>\phi</code>
カイ	X	χ	chi	<code>\chi</code>
プサイ	Ψ	ψ	psi	<code>\psi</code>
オメガ	Ω	ω	omega	<code>\omega</code>

引用文献

- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59, 1–23. <https://doi.org/10.18637/jss.v059.i10>
- 中野, 友. (2017). 大学新入生のためのリメディアル数学 (第2版). 森北出版.
- 千野, 直., 岡田, 謙., & 佐部利, 真. (2012). 非対称 mds の理論と応用 (単行本). 現代数学社.
- 南風原, 朝. (2002). 心理統計学の基礎 - 統合的理解のために. 有斐閣.
- 小杉, 考. (2018). 言葉と数式で理解する多変量解析入門. 北大路書房.
- 岡太, 彬. (2008). データ分析のための線形代数. 共立出版.
- 村上, 正., 佐藤, 恒., 野澤, 宗., & 稲葉, 尚. (2016). 教養の線形代数 (単行本). 培風館.
- 柳井, 晴., 前川, 真., 繁杵, 算., & 市川, 雅. (1990). 因子分析—その理論と方法 (統計ライブラリー). 朝倉書店.
- 永田, 靖. (2005). 統計学のための数学入門 30 講 (単行本). 朝倉書店.
- 竹内, 淳. (2010). 高校数学でわかる線形代数—行列の基礎から固有値まで (ブルーバックス) (新書). 講談社.
- 竹山, 美. (2022). 数学書の読みかた. 森北出版.
- 絹田, 村. (2019). 数字であそぼ。 (2) (フラワーコミックス α). 小学館.
- 芝, 祐. (1979). 因子分析法 (単行本). 東京大学出版会.
- 西里, 静. (2010). 行動科学のためのデータ解析—情報把握に適した方法の利用. 培風館.
- 谷, 伊. & 平島, 太. 伊藤, 大. (Ed.) (2018). 心理学・社会科学研究のための構造方程式モデリング: *mplus* による実践 基礎編. ナカニシヤ出版.
- 長沼, 伸. (2011). 物理数学の直観的方法〈普及版〉 (Kindle 版). 講談社.

索引

Q

Q 技法 132

あ

一次結合 56
 一次従属 56
 一次独立 56
 位置ベクトル 55
 一般線形モデル 34
 因果関係 121
 因子間相関 116
 因子構造 116
 因子軸の回転 116
 因子得点 109, 110
 因子パターン 117
 因子負荷量 109, 110
 因子分析 109
 上三角行列 14

か

核 75
 回帰分析 33
 階数 48, 74
 外生変数 124
 外生変数ベクトル 127
 階段行列 48
 回転行列 94, 116
 解の自由度 130
 ガウスの消去法 44
 拡大係数行列 41
 確率分布 122
 過剰識別 130
 数ベクトル 55
 観測変数 121
 機械学習 132
 幾何ベクトル 55
 期待値 122
 基底 56
 基本行列 42
 基本ベクトル 22, 69
 逆行列 44, 100
 列基本変形 42
 行基本変形 41
 共通因子 110
 共通性 112
 共通性の推定 115
 共分散 26
 共分散構造分析 121
 行ベクトル 11
 行列 9, 12
 行列式 45
 偶然誤差 108
 グラム・シュミットの正規直交化法 70
 係数行列 41
 系統誤差 108
 構造変数ベクトル 127
 構造方程式 123
 構造方程式モデリング 121

固有値 99
 固有ベクトル 99
 固有方程式 100

さ

最小二乗法 86, 116, 130
 最尤法 116, 130
 識別可能 129
 識別不可能 129
 次元 70
 下三角行列 14
 射影 61
 射影行列 79, 81
 斜交回転 116
 写像 74
 重回帰分析 33
 自由度 89
 小行列 21
 信頼性 108
 数量化 III 類 132
 スカラー 12
 スペクトル分解 103
 正規直交基底 66
 正規分布 108
 正射影 78
 整然データ 9
 正則行列 45
 正方行列 12
 積空間 72
 線型結合 56
 潜在変数 121
 選択方程式 127
 像 75
 相関関係 121
 相関行列 13, 32
 相関係数 26
 双対尺度法 132
 測定方程式 123

た

対応分析 132
 対角 13
 対角化 104
 対角行列 13
 対称行列 13
 多次元尺度構成法 131
 多重共線性 87
 単位行列 13
 単純構造 116
 単純構造の原則 113
 丁度識別 129
 直和 72
 直交 66
 直交回転 116
 直交行列 94
 直交射影 81
 直交補空間 72
 デザイン行列 36

転置	20
部分ベクトル空間	72
特異値分解	132
独自性	113
トレース	21, 92

な

内生変数	124
内積	60
ノルム	59

は

掃き出し法	44
バス解析	124
バス図	121
バスダイアグラム	121
標準誤差	89
標準偏差	26
分割	21
分散	25
分散共分散行列	13, 31
平均平方	89
平均偏差	25
平方和	89

ら

列ベクトル	11
-------	----

わ

和空間	72
-----	----