統計学入門

安田正實

2006 年 4 月 16 日

目次

図目次

表目次

第 1 章

統計学の体系と歴史

統計学（とうけいがく）とは、統計に関する学問のことである。統計学は、経験的に得られたバラツキのあるデータから、応用数学の手法を用いて数値上の性質や規則性あるいは不規則性を見いだす。統計的手法は、実験計画、データの要約や解釈を行う上での根拠を提供する学問であり、幅広い分野で応用されている。英語で統計または統計学を statistics と言うが、語源はラテン語で「状態」を意味する statisticum であり、この言葉がイタリア語で「国家」を意味するようになり、国家の人力、財力等といった国勢データを比較検討する学問を意味するようになった。現在では、経済学、自然科学、社会科学、医学 (疫学、EBM)、薬学など広い分野で必須の学問となっていることは論をまたない。

1.1 統計的手法

記述統計　記述統計とは、収集したデータの要約統計量（平均、分散など）を計算して分布を明らかにする事により、データの示す傾向や性質を知ること。

推測統計　データからその元となっている諸性質を確率論的に推測する分野。推測統計学の項に詳述。

データ構造　データ（あるいは変数、測定）の尺度はふつう次のような種類（水準）に分類される。尺度水準によって、統計に用いるべき縮約統計量や統計検定法が異なる。**質的データ（カテゴリデータ）**リストや項目別にデータが得られる。名義尺度：単なる番号、項目で順序の意味がない。電話番号、県別リストなど。順序尺度：順序が意味をもつ番号あるいは項目リスト。階級や階層など、たとえば秀優良可はこの順序に応じて並べられる。**量的データ（数値データ）**個数、回数を数えたりするときには離散データとして得られるが、測定の精度に応じて、連続的に小数点以下までも求められるような実数をデータとするものは連続データ。。間隔尺度：順序に加え間隔にも意味がある（単位がある）が、ゼロには絶対的な意味はない。摂氏・華氏温度、知能指数など。比率尺度：ゼロを基準とする絶対的尺度で、間隔だけでなく比率にも意味がある。絶対温度、金額など。

実験計画, 品質管理データ収集の規模や対象、割付方法をコントロールし、より公正で評価可能なデータが収集できるよう検討すること。統計の世界には Garbage in, garbage out という格言がある。これは「ゴミのようなデータを使っていくら解析しても出てくる結果はゴミばかりだ」という意味であり、データ収集の前にその方法を十分に検討する必要があることを強調したものである。

1.2 歴史

統計学の源流は国家または社会全体における人口あるいは経済に関する調査（東西を問わず古代から行われている）にある。学問としては、17 世紀にはイギリスでペティの『政治算術』などが著述され、その後の社会統計学につながる流れが始まった。またライプニッツやエドモンド・ハレーによる死亡統計の研究も行われた。これらの影響のもと 18 世紀にはドイツのジュースミルヒが『神の秩序』(1741 年）で人口動態にみられる規則性を明らかにしたが、これには文字通り「神の秩序」を数学的に記述する意図があった。

ドイツでは 17 世紀からヨーロッパ各国の国状の比較研究が盛んになったが、1749 年にアッヘンヴァルがこれにドイツ語で Statistik（「国家学」の意味）の名をつけている。19 世紀初頭になるとこれに関して政治算術的なデータの収集と分析が重視されて、Statistik の語は特に「統計学」の意味に用いられ、さらにイギリスやフランスなどでも用いられるようになった。この頃アメリカ、イギリス、フランスなどで国勢調査も行われるようになる。

一方パスカル、フェルマーに始まった確率論の研究がフランスを中心にして進み、19 世紀初頭にはラプラスによって一応の完成を見ていた。またオイラーによる誤差や正規分布についての研究も統計学発展の基礎となった。ラプラスも確率論の社会的な応用を考えたが、この考えを本格的に広めたのが「近代統計学の父」と呼ばれるアドルフ・ケトレーであった。彼は『人間について』(1835 年)、『社会物理学』(1869 年）などを著し、自由意志によってばらばらに動くように見える人間の行動も社会全体で平均すれば法則に従っている（「平均人」を中心に正規分布に従う）と考えた。ケトレーの仕事を契機として、19 世紀半ば以降、社会統計学がドイツを中心に、特に経済学と密接な関係を持って発展する。代表的な人物にはアドルフ・ワグナー、エルンスト・エンゲル（エンゲル係数で有名)、ゲオルク・フォン・マイヤーがいる。またフローレンス・ナイチンゲールも、社会医学に統計学を応用した最初期の人物として知られる。

同じく 19 世紀半ばにダーウィンの進化論が発表され、彼の従弟に当たるゴルトンは数量的側面から進化の研究に着手した。これは当時 Biometrics*(生物測定学) と呼ばれ、多数の生物（ヒトも含めて）を対象として扱う統計学的側面を含んでいる。ゴルトンは回帰の発見で有名であるが、当初生物学的と思われたこの現象は一般の統計学的対象の解析でも重要であることが明らかとなる。ゴルトンの後継者となった数学者カール・ピアソンはこのような生物統計学をさらに数学的に発展させ（数理統計学)、19 世紀終わりから 20 世紀にかけ記述統計学を大成する。(*注：現在の言い方では生物統計学 Biostatistics に当たり、この単語は現在では生体認証という別の意味で使われている）

20 世紀に入ると、ゴセット、続いて R.A. フィッシャーが農学の実験計画法研究をきっかけとして数々の統計学的仮説検定法を編み出し、記述統計学から推計統計学の時代に移る。ここでは母集団から抽出された標本を基に、確率論を利用して逆に母集団を推定するという考え方がとられる。続いてネイマン、エゴン・ピアソンらによって現代の推計統計学の理論体系が構築され、これは社会科学、医学、工学などの様々な分野へ応用されることとなった。

1.3　統計調査

統計（とうけい、英:statistic）とは、現象を調査することによって数量で把握すること、または、調査によって得られた数量データのことである。

1.3.1　統計の変遷

統計は、国を統治するための基礎資料であり、建造物建設のための調査や兵役や徴税のための調査といったように、人口や土地等については古くから統計が取られている。また、近代国家が成立した頃から政策の企画・立案のために利用されるようになり、調査範囲も広がった。 ナポレオンは「統計は事物の予算である。そして予算なくしては公共の福祉も無い」と語り、1800 年にはフランス、1828 年にはオーストリアで国の調査機関が設立された。さらに、パソコンの普及、分析手法の発達によって大学や企業なども統計を利用するようになり、「国のためのデータ」から「国民のためのデータ」へとその性質は変わってきている。

1.3.2　統計の種類

国や地方公共団体などによる統計

　国が行う統計調査には、法的根拠や強制力などがあり、一般的に調査結果への信頼性が高い。また、調査項目の連続性が考慮されることが多いことから分析に使いやすく、定期性、速報性にも優れている。ただし、社会の変化への対応は鈍い。調査項目の改正などには時間がかかり、特に新たな分野に対しては調査の遅れや、調査しても対象の補足が満足にできないなど、不十分なものにもなりやすい。

分類

　手続き面からみた分類と、作成手段面からみた分類とに分けて解説する。

《手続き面》

指定統計

　国や地方公共団体 (都道府県、市町村) が作成する統計のうち、統計の中でも特に重要なものであるため総務大臣が「指定統計」に指定したものをいう。統計法により申告義務が課せられており、57 調査が指定統計となっている (2004 年 4 月現在)。(例：国勢調査、家計調査等)

承認統計

　国が作成する統計のうち、個人や民間事業所を対象にし、指定統計以外の物をいう。報告者負担の観点から、実施に当たっては総務大臣の承認が必要となっており、121 調査が承認されている (承認期間が 2004 年 6 月 1 日以降有効なもの)。(例：消費動向調査等)

届出統計

　指定統計及び承認統計以外の国や地方公共団体が作成する統計。実施に当たっては事前に総務大臣への届出が必要となっている。また、国が行う届出調査は、地方公共団体を対象とする物が中心となっている。(例：住民基本台帳人口移動報告)

《作成面》

調査統計 (一次統計)

　統計を作成することを目的として行われる調査から得られる統計のこと。国勢調査や工業統計のように、集団の構造を把握する事が目的の調査と、家計調査や毎月勤労統計調査のように、消費や賃金といった限られた項目に焦点を絞って、項目の時間的な動きを把握する事が目的の調査とにわかれる。その性質上、前者の構造把握用の統計は対象範囲が広く、時間とお金がかかることから大規模 (特に全数調査のことを、センサスと呼ぶことがある)、隔年実施となる。一方、後者の特定項目の動きを把握する調査 (動態統計と呼ばれることが多い) は、速報性を重視して小規模な標本調査で時間もお金もあまりかけずに、月、四半期ごとに調査を行う。

業務統計 (一次統計)

　登録や届出、業務記録など、行政機関が行政上、業務上の必要から集めた記録などをもとに作成する統計のこと。輸出入の通関書類から作成される貿易統計などがある。中には出生・死亡・婚姻の届出をもとにした人口動態統計のように、統計調査の形を取って統計法の適用を受けるものもある。

加工統計 (二次統計)

　一次統計を利用、加工した統計のこと。国民経済計算や鉱工業指数がこれにあたる。統計学を駆使し、直接調査することが困難な事象を把握するために作成される。

　一次統計と二次統計の違いは、一次統計は調査対象を直接調べる統計のことで、二次統計は一次統計を加工した統計のことである。

統計調査の流れ

　一般的な調査の流れは、次の通りである。

1. **調査の企画、設計**　行政等のニーズに応じ、ニーズを所管する省庁が統計を企画、設計し、総務省 (大臣) に承認申請、届出を行い、許可をもらう。
2. **調査の説明**　都道府県 (大規模な調査は都道府県が調査作業を行う場合が多い。国勢調査は市区町村まで関わる) に調査方法や、調査の主旨などを説明する。
3. **調査の実施**　省庁又は都道府県が、調査対象の個人や事業所に調査票を配布、記入後回収する。
4. **調査の集計**　回収した調査票を検査 (桁ズレ、差し引きがおかしいなどがあった場合には調査対象に照会して修正する) 、集計を行う。
5. **調査の発表**　集計した調査結果を分析し、公表する。

○地方統計機構

国が行う大規模な統計については、地方統計機構整備要項 (1947 年 7 月 11 日　閣議決定) によって地方公共団体を活用するとなっており、都道府県に対して統計主管課を設置、統計専任職員を配置。また統計専任職員の配置などによってかかる経費は国が統計調査事務地方公共団体委託費として交付。

参考：地方財政法 (地方公共団体が負担する義務を負わない経費) 第 10 条の 4　専ら国の利害に関係のある事務を行うために要する次に掲げるような経費については、地方公共団体は、その経費を負担する義務を負わない。　(略)　二　国が専らその用に供することを目的として行う統計及び調査に要する経費

○事業者団体・企業などによる統計

業界団体や企業が調査、推計している統計。国と違い、あくまで団体に所属している企業の自主回答や、企業の取引対象への聞き取り等が主要なものとなるため強制力を伴わず、統計によっては回答率が 10 %台というものもある。そのため国の調査に比べて信頼性で劣ったり、毎年公表がほとんどのため速報性に欠ける。また、調査項目が必ずしも一定ではなく、分析に使用しづらい場合があるが、これは統計の主目的が業界の現状の把握のため、時節に即した調査項目が選択されるからであり、悪いことではない。そのため、社会の変化に対する反応は早く、国が調査を行わないような業界、項目についてもいち早く数字で把握することは出来る。また業界公認の値ということである程度の信頼性は担保される。企業が行う調査についても、電通の広告に関する発表資料のように、公式の値として広く利用されるものも中にはある。

1.4　調査のもつ課題

社会の変化への対応：社会・経済情勢の変化のスピードが速くなってきているため、国の統計は変化への対応が遅く、業界団体の統計では詳細な分析が行えないという様に、現状の正確な把握ができなくなってきている。 特に第三次産業でこの傾向が強い。

調査対象の意識の変化への対応：個人については従来からプライバシー保護といった意見はあったが、個人情報保護法の施行以降、個人情報保護の考えは急速に広まり、個人を対象とする調査は行いづらくなってきている。2005 年に行われた国勢調査についても、調査拒否が頻発した。企業については自社の情報を出したくないという防衛意識は以前からあったが、規制緩和による国の行政指導力の低下、業界団体への未加入企業の増加、外資系企業の増加から、統計調査への協力拒否が増えてきている。

調査項目の見直し：国の統計の場合、調査主体が各省庁に分かれるため、似たような項目を複数の調査から聞かれるなどの問題点が指摘されており、「記入者負担の軽減からこれらを一本化すべし」などの意見がある (例：日本経済団体連合会 「ペーパーワーク負担の実態と改善方策に関する調査報告」)

推計精度、速報性向上への対応：統計が行政だけでなく、広く社会で使用されるようになるにつれて、統計の推計精度や、速報性が問題になり始めた。最もやり玉に挙がるのは四半期 GDP だが、これは注目度の裏返しともいえる。基本的に統計は予算・速報性と精度とはトレードオフの関係にあるものの、双方の両立が求められている。

予算制約への対応：国が行う統計においては財政状況の厳しさから統計も予算削減の対象になっており、毎年調査を隔年調査に切り替えたり、本調査と簡易調査とに切り替える等が行われている。また、予算上の制約

は新規統計の設計も困難にしており、スクラップアンドビルド (新しい統計を行う場合には、現在の統計を削る) が基本となってきている。

官から民への業務移管：小泉内閣の打ち出した聖域なき構造改革において、政府が調査している統計は官から民への業務移管の検討対象になっており、市場化テストの候補に挙がっている。

1.5　指定統計一覧

以下は省庁別の指定統計リスト。指定番号、指定年月日など、詳しくは総務省統計局の一覧を参照。

* 総務省：国勢調査、事業所・企業統計、住宅・土地統計、労働力調査、小売物価統計、家計調査、個人企業経済調査、科学技術研究調査、地方公務員給与実態調査、就業構造基本調査、全国消費実態調査、全国物価統計、社会生活基本統計、サービス業基本統計

* 財務省：法人企業統計

* 国税庁：民間給与実態統計

* 文部科学省：学校基本調査、学校保険調査、学校教員統計、社会教育調査

* 厚生労働省：人口動態調査、毎月勤労統計調査、薬事工業生産動態統計調査、医療施設統計、患者統計、賃金構造基本統計、国民生活基礎統計

* 農林水産省：農林業センサス、牛乳乳製品統計、作物統計、海面漁業生産統計、漁業センサス、製材統計、農業経営統計

* 経済産業省：工業統計調査、経済産業省生産動態統計、商業統計、埋蔵鉱量統計、ガス事業生産動態統計、特定機械設備統計調査、石油製品需給動態統計、商業動態統計調査、特定サービス産業実態統計、特定業種石油等消費統計、企業活動基本調査、商工業実態基本調査

* 国土交通省：湾港調査、船舶船員調査、造船造機統計、建築着工統計、鉄道車両等生産動態統計調査、建設工事統計、船員労働統計、自動車輸送統計、内航船舶輸送統計、法人土地基本調査

”http://ja.wikipedia.org/wiki/” より作成

第 2 章

数学の基礎事項

2.1 数の種類

- 自然数: $1, 2, 3, \cdots$,(0 は入れない)、 数え上げにもちいる。数えることは自然数との 1 対 1 対応をすることです。偶数: $0, \pm 2, \pm 4 \cdots$ (2 で割り切れる数) 、奇数: $\pm 1, \pm 3, \pm 5 \cdots$ (2 で割り切れない数)
- 整数とは自然数の集まりにゼロと自然数に符号をつけた負の数とを合わせたもの。正の整数は $1, 2, 3, \cdots$, で自然数と同じ集まり。非負の整数:$0, -1, -2, -3, \cdots$,
- 分数の形で表せる数を有理数といいます。分数は具体的に割り算をして、小数として表すことが多い。小数どうしの計算では有効桁数が統計学で重要になります。電卓での計算、とくに積や 2 乗の計算には桁上がり、桁落ちには注意します。また指数表示も覚えておきましょう。
- 無理数とは分数の形で表せない数をいいます。たとえば $\sqrt{2}, \dfrac{\sqrt{5}+1}{2}$, 2 次方程式、3 次方程式などの解となるものです。さらにこれ以外にも $\pi = 3.1415926538\cdots$, $e = 2.71828\cdots$ もありますが、これらはとくに超越数とよばれ、多項式による方程式では表すことができません。有理数は自然数と整数を含むものですし、さらに無理数を合せた集まりを実数とします。数直線には、この実数全体があるものとします。実数以外の数は数直線の上には存在していないものと仮定をします。これを実数の連続性公理とします。
- 複素数とは新たに虚数単位を導入して実数の世界を拡張したものです。負の数の平方根を純虚数とよび、実数と純虚数との和を複素数と定めます。

問 2.1.1

数のもつ性質にはどんなことがあるか、いくつか挙げてみてください。また演算にはどういうことがありますか？

問 2.1.2

電卓を用いて計算するときに表示される数を具体的にきちんと書いてみてください。たとえば $2.3E2 = 2.3 * 10^2 = 230, 1.5E-2 = 1.5 * 10^{-2} = 1.5/100 = 0.015$ です。

(1) $3.14E3$ (2) $5.0E10$ (3) $6.723E-3$

問 2.1.3

未知数の含む式について、方程式の解と数の関係はどんなことが知られていますか？

問 2.1.4

円周率のパイ (π) や自然対数の底に使われるイー (e) はだいたい $\pi = 3.1415926538\cdots$, $e = 2.71828\cdots$ となりますが、どうやって計算するのか考えてみてください。

2.2　数列

自然数と 1 対 1 の対応

番号	1	2	3	4	$\cdots$	k	$\cdots$
数	a_1	a_2	a_3	a_4	$\cdots$	a_k	$\cdots$
数	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$\cdots$	$x^{(k)}$	$\cdots$

添え数（番号）が 2 つついているもの (i,j) に対応して $a(i,j), a_{(i,j)}, a_{ij}$ (カッコや誤解がなければカンマを省略することが多い) については 2 重和 $\sum_i \sum_j a(i,j)$ をもちいたり、2 つを同時にまとめ、ひとつの $\sum_{i,j} a(i,j)$ で表現します。有限個の和であれば、順序を入れ替えても同じですから、すべて同じ値となります。ただし $\sum_i \sum_j a(i,j)$ と $\sum_j \sum_i a(i,j)$ の表現では前者については、先に j で足し算をしてから i で加え、この順が逆ものが後者です。有限個ではない場合には、一般には順序の交換も一般には成り立ちませんから、複雑になりますが、交換可能となる場合もあります。

2.3　ベクトルと行列

実数を長方形にならべてカッコでくくったもの。横を行とよび、縦を列とよぶ。数を対応させるために、添え数（インデックス）が組み (i,j) であり、その対応する実数 $a_{i,j}$ を成分あるいは要素とよぶ。、行と列

$$(i,j) \Rightarrow a_{ij}$$

縦 (列）ベクトルと横 (行）ベクトル: 3 行 1 列の行列と 1 行 3 列の行列とみなせる

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \qquad (a,b,c)$$

正方行列 (3 行 3 列)

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

無限行列

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots \\ a_{21} & a_{22} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

行列の演算: (1) 行列の和 $A+B$, (2) 行列のスカラー倍 cA (c スカラー)。ここでスカラーとはベクトルに対する概念で、実数として考えてよい。また行列の積の計算（内積）; AB, BA などをよく復習しておいてください。

2.4　和の記号

足し算をする一般項が規則的に表現できるときには、足し算を表す記号（シグマとよび） $\sum$ で書き表します。とくに範囲を指定しなくても自明であれば、和をとるインデックスの範囲を省略することもあります。またインデックスそのものを省略してしまうこともよくあります。

インデックス変数 $k(=1,2,3,\cdots,N)$ に対応する第 k 項が a_k ならば

$$\sum_{k=1}^{N} a_k = a_1 + a_2 + a_3 + a_4 + \cdots + a_k + \cdots + a_N$$

と表します。また

$$\sum_j x^{(j)} = x^{(1)} + x^{(2)} + x^{(3)} + x^{(4)} + \cdots + x^{(k)} + \cdots$$

$$\begin{aligned}\sum_{i,j} a_{ij} = \quad & a_{11} + a_{12} + a_{13} + a_{14} + \cdots \\ & +a_{21} + a_{22} + a_{23} + a_{24} + \cdots \\ & + \cdots\end{aligned}$$

$$\sum_{i=1}^{3}\sum_{j=1}^{2} a_{ij} = \{a_{11} + a_{12}\} + \{a_{21} + a_{22}\} + \{a_{31} + a_{32}\}$$

という意味を表しています。範囲があらかじめ分かっている場合には省略をすることが多いです。インデックス変数には自然数がよく用いられます。

和を求める計算式

$$\textstyle\sum_{k=1}^{n} k = 1 + 2 + \cdots + n = \dfrac{n(n+1)}{2}$$

$$\textstyle\sum_{k=1}^{n} c = c + c + \cdots + c = nc \quad \text{（ n 個の定数和）}$$

$$\textstyle\sum_{k=1}^{n} (-1)^{(k-1)}/k = 1 - 1/2 + 1/3 + \cdots + (\pm 1)^n/n$$

$$\textstyle\sum_{k=1}^{n} (2k-1) = 1 + 3 + 5 + \cdots + (2n-1) = n^2$$

$$\textstyle\sum_{k=1}^{n} k^2 = 1^2 + 2^2 + \cdots + n^2 = \dfrac{n(n+1)(2n+1)}{6} \quad \text{（2 乗和の公式）}$$

$$\textstyle\sum_{k=1}^{n} k(k+1) = 1 \times 2 + 2 \times 3 + \cdots + n \times (n+1) = \dfrac{n(n+1)(n+2)}{3}$$

問 2.4.1

（積が 3 つ）各項が 3 個の積のとき、$\sum_{k=1}^{n} k(k+1)(k+2)$ がどういう式にまとめられるか考えなさい。そして実際いくつかについて確かめてみなさい。

問 2.4.2

（3 乗和の計算）$\sum_{k=1}^{n} k^2$ を、上の問いで得られた結果ももちいて、求めてみなさい。

問 2.4.3

（三角型の和）$\sum_{k=1}^{n}\sum_{j=1}^{k} a_{kj}$ をそれぞれの項で書き並べてみましょう。この和のとり方の順序がはじめ、列をとってから、行で加える順ですが、この順序を逆にしてはじめは行でとり、あとから列を加えることにすると、シグマを用いるとどういう式になりますか？

問 2.4.4

つぎの数列の和をシグマにより表現しなさい。

1. $a_1 + a_3 + a_5 + \cdots + a_{31}$
2. $\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots$
3. $\frac{1}{1} + (\frac{2}{1} + \frac{2}{2}) + (\frac{3}{1} + \frac{3}{2} + \frac{3}{3}) + (\frac{4}{1} + \frac{4}{2} + \frac{4}{3} + \frac{4}{4}) + \cdots$

2.5　極限、級数

数列の和に関して、項の数をどんどん大きくしていった状況を考えてみます。有限個で終わるのではなく、無限に続いて多数の項をもつ数列を級数といいます。もしこの和を順次計算していったとき、項数とともに増加あるいは減少していく傾向をもち、ある一定の値に近づくならば、この極限値を級数の和といいます。つまり、部分和の極限値として

$$\lim_{N\to\infty}\sum_{n=1}^{N} a_n = \lim_{N\to\infty} S_N = \sum_{n=1}^{\infty} a_n = s$$

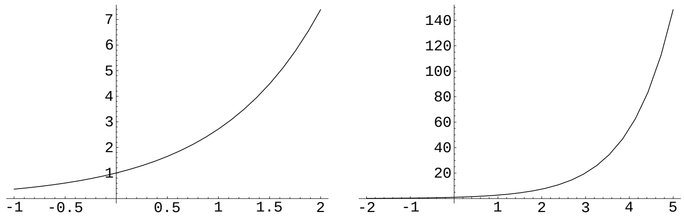

図 2.1　指数関数 (その 1) $y = \exp(x)$

と表します。しかし一般項がゼロに近づいていても、和をもつ（有限和の極限値が存在する）とは限りません。たとえば $a_n = 1/n$ などはその例です。これよりももっと早くゼロになるような項でなければ、和をもちません。またここで収束する (convergent)、つまり $S_n \to s(n \to \infty)$ であることは絶対値で表される距離 $|S_n - s|$ が n の増加としたがい、ゼロに近づくことです。記号では $\lim_{n\to\infty} S_n = s$ と表します。

2.6　対数、指数関数

確率や統計では、コイン投げ、サイコロ振りを多数回繰り返した場合をよく考えます。どんどんと繰り返し数を多くしていくので、数列の極限を調べることがとても重要な基本事項となります。このときに表れる関数が、対数関数や指数関数です。概要をつぎで述べてみましょう。

指数関数:「べき」とは乗積のことで、正の数 $a > 0$ を m 乗するとは、$a^m = aa \cdots a$ で明らかに $a^m a^n = a^{m+n}$ などが成り立ち m を指数とよび、この性質を指数法則といいます。m は自然数だけでなく負とでもよくて、一般に整数であっても法則が成り立ち、さらに分数、実数についても拡張して定義されます。

たとえば利率の計算では、元金が a 円あり、複利の利率が各期間ごとに r_n であれば、n 期日後には $a(1+r_n)^n$ となります。もし利率の値 $r_n = r/n$ であり、n を大きくすると（利率の計算期間を細かくしていくこと）、$a \lim_n (1 + r_n)^n$ を考えることです。基本的な項は $r = 1$ とおいた $\lim_n (1 + \dfrac{1}{n})^n$ で、これは少しづつ増加し、数値 3 を超えることはない。この極限値を $e = 2.71828 \cdots$ と表します。一般に $\lim_n (1 + r_n)^n = e^r = \exp r$ とも表します。この r を変化させた値がいわゆる $y = \exp x, x > 0$ が指数関数とよばれるものです。先の述べた a^m の形が e^x となっています。

対数関数: 上で述べた指数関数は、べき乗を細かくすることで表れた関数でしたが、対数関数は、指数関数の逆関数として定義されます。大きな数を計算するときはたいへん重要な役割をします。計算機が十分に発達していない時代には対数表とよばれるものを用いて、大きな数の計算をしていました。

与えられた指数関数に対して、その逆の対応を表す

$$y = e^x = \exp x \quad \leftrightarrow \quad x = \log y$$

ものです。独立変数 x から従属変数 y の形で書けば、$y = \log x$ と表します。これは電卓などでは $y = Ln(x)$ とキー入力します。実用上の計算には e の代わりに 10 を用いて、$y = Log(x)$ とキー入力します。ここでは省略していますが対数の底（てい）が e（ネイピア数ともよばれる）の場合、つまり自然対数の場合と一方は底が 10 の場合での常用対数のものです。実用上の計算には、常用対数が便利ですが、理論の展開には自然対数のほうが表現が簡単になるときが多いです。どちらを使っても本質は変わりません。

2.7　集合

ある定まった数学的な対象の集まりあるいは集団を**集合** (set) とよび，この集合を構成する一つ一つの対象を集合の要素 (element) あるいは、元という．一般に，集合は大文字をつかい A, B などと表し、この要素は小文字 x, y などをもちいる。ある要素 x が集合 A に属する (belong) かどうかによって定められている。もし属するならば $x \in A$ と表す．また，逆に，A の要素でないときは $x \ni A$ と表す．

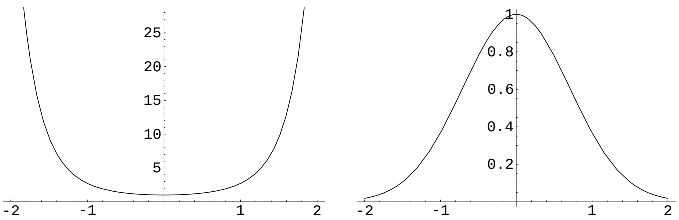

図 2.2　指数関数 (その２) $y = \exp(x^2)$ と $y = \exp(-x^2)$

要素を並べ、括弧でくくることで $\{a, b, c, d, e\}$ や $\{x_1, x_2, x_3, \cdots\}$ と表したり, あるいは {要素—条件} の形をもちいて、$\{x_i | i = 1, 2, 3, \cdots\}$ や $\{x | a < x \leqq b\}$（つまり実数の半開区間 $(a, b]$）などの表し方ももちいる。

とくに集合の中で全くこれに属する要素をもたない場合には**空集合** (empty set) とよび、$\emptyset$ で表す。2 つの集合において $\forall x \in A \Rightarrow x \in B$ が成り立つとき、$A \subset B$ で表す。$A \subseteq B$ も同じ意味でこのとき、A は B の**部分集合**という。$A = B$ とは、それぞれの部分集合であるときをいう。とくに真の部分集合とは $A \subset B$ かつ $A \neq B$ のときであり、$A \subsetneq B$ で表す。

元（要素）の個数が有限個しか含まれないときを有限集合、そうでない場合、自然数の全体とか、実数の全体は無限集合という。

和集合と積集合 数と数のあいだに足し算や引き算があるように、集合と集合のあいだにも演算を考える。

$$\text{和集合：}\quad A \cup B = \{x \mid x \in A \,\mathrm{or}\, x \in B\}$$

$$\text{積（共通）集合：}\quad A \cap B = AB = \{x \mid x \in A \,\mathrm{and}\, x \in B\}$$

$\cap$ を省略することもある。数の掛け算でも多くは省略している。

積集合が空集合のとき、$A \cap B = AB = \emptyset$ には**互いに素**、あるいは共通部分がない、交わりをもたないなどという。**差集合**とは $A \setminus B = \{x \mid x \in A \,\mathrm{and}\, x \notin B\}$ いま対象とる全体の集合を Ω(オメガ) とするとき, この全体集合 Ω から A を引いた差集合、$\Omega \setminus A$ を A の**補集合** (complement) といい、全体集合を明示せずに A^c や $\overline{A}$ という記号をもちいる。

交換法則と結合法則：演算の順序をかえることでも同じ場合にこの法則が成り立つという。$A \cup B = B \cup A$ などや $(A \cup B) \cup C = A \cup (B \cup C)$ など。したがって単にカッコをもちいず $A \cup B \cup C$ と表しても問題ないということでこの表現が意味をもつことになる。よって一般に多数のばあいについては

$$\text{和集合：}\quad \textstyle\bigcup_{i=1}^{n} A_i = \bigcup_i A_i = A_1 \cup A_2 \cup \cdots \cup A_n,$$

$$\text{積集合：}\quad \textstyle\bigcap_{i=1}^{n} A_i = \bigcap_i A_i = A_1 \cap A_2 \cap \cdots \cap A_n.$$

と表す。分配法則：３つの集合に対して、和と積の順序について交換するとどうなるかという関係式をいう。これをみれば明らかなように通常の数の演算 $+, \times$ が集合の和、積に対応している。

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

ドモルガン (de Morgan) の法則: 和と差、あるいは積と差のあいだの分配法則をあらわすもの、

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

一般に n 個の集合 $A_i, i = 1, 2, \cdots, n$ についても同様に

$$\overline{\bigcup_i A_i} = \bigcap_i \overline{A_i}, \qquad \overline{\bigcap_i A_i} = \bigcup_i \overline{A_i}.$$

２つの集合の直積: それぞれの集合からとった作られる順序のついた全体からなる集合を**直積**といい、$A \times B$ で表す。$A \times B = \{(a, b) \mid a \in A, b \in B\}$ 掛け算の記号をつかっているが、その内容には注意をする。もし

$A = \{1,2,3,4,5,6\}$ であれば、$A \times A$ は 36 個の要素をもつ $(i,j); i,j = 1,2,\cdots,6$ からなる集合でとなる。これはサイコロを 2 個投げたときに出るすべての目の組の可能性を列挙したものとなる。

集合族、べき集合：集合の集合、つまり元（要素）がまた集合となっているものを集合族 (a family of sets) という。とくに**べき集合** (power set) とは記号で集合 A の部分集合全体 2^A で表す。たとえば $A = \{2,4,5\}$ の 3 個の有限集合ならば $2^A = \{\emptyset, \{2\}, \{4\}, \{5\}, \{2,4\}, \{2,5\}, \{4,5\}, \{2,4,5\}\}$ の $2^3 = 8$ 個の集合となる。

上極限集合、下極限集合：無限個の集合族について積や和をとったの概念である。数に対する集合でも同じような概念が、集合の無限列についても定義されている。これは確率の定義をするときには重要な働きをする。

2.8 写像

写像 (map,mapping) の定義は 2 つの集合 A, B において一つの要素 $x \in A$ から他の一つ $y \in B$ への対応があるとき、これを f と表し、$f : A \longrightarrow B$ つまり $y = f(x), x \in A, y \in B$ とおく。

像 (image) と逆像 (inverse image)：写像 f で、$b = f(a)$ のとき、b を a の像といい、a を b の原像という。集合についてある部分集合 A_1 で、$A_1 \subset A$ のとき、$f(A_1) = \{f(a) \mid a \in A_1\}$ を A_1 の像といい、ある部分集合 B_1 で、$B_1 \subset B$ のとき、$f^{-1}(B_1) = \{a \mid f(a) \in B_1\}$ を B_1 の逆像という。

定理 写像 $f : A \longrightarrow B$ でつぎの集合間の関係式が成り立つ (ただし $A_i \subset A, B_i \subset B, i = 1,2$).
(1) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$ (2) $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$ (部分集合となる)
(3) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$ (4) $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$
(5) $A_1 \subset f^{-1}(f(A_1))$ (部分集合となる) (6) $f(f^{-1}(B_1)) \subset B_1$ (部分集合となる)

なぜなら (2) については

$$
\begin{aligned}
& \{b \in B \mid b = f(a), \exists a \in A_1 \cap A_2\} \\
\subset \quad & \{b \in B \mid (b = f(a_1), \exists a_1 \in A_1) \,\text{and}\, (b = f(a_2), \exists a_2 \in A_2)\}
\end{aligned}
$$

であるから。

問 2.8.1
たとえば実数上の 2 次関数 $f(x) = x^2$ において上の例になる集合をあげてみよ。答えのひとつにはつぎがあり、他の例も各自考えてみよ。$A_1 = [-3,1], A_2 = [-1,2], B_1 = [-1,1]$

写像としてとくに実数の集合を考えるには関数とよぶことがふつうであり、関数の定められている範囲やとり得る値の範囲を定義域 (domain)、 値域 (range) とよぶ。また 2 つの関数の合成関数 $f \circ g(x) = f(g(x))$ とか逆関数 f^{-1}、対数関数の逆関数として指数関数、指数関数の逆関数として対数関数、$f(x) = x^2\,(x \geq 0)$ の逆関数として $f^{-1}(x) = \sqrt{x}\,(x \geq 0)$ も定義される。逆三角関数は三角関数の逆関数である。

2.9 平方根の開平

ルートの計算値を手計算で求めましょう。電卓がないと計算できないことはありません。つぎのように求められます。知っておいたほうがよいです。

たとえば、$\sqrt{2} = 1.4142213562$ と電卓では出てきますが、あるいは少し欲張ると、1.414213562373095048 8ぐらいはパソコンでもでてきます。しかし機械に頼るではなく、理由を考えることもとても大切なことなのです。

かんたんなものから、始めます。$256 = 16^2$ ですから、これを考えてみます。手順をしてこれの理由を述べましょう。

(1) 1 の位から 2 桁づつに仕切りをいれます。2 と 56 に分かれます。(2) 左端からはじめて順次に進めます。平方して 2 を越えない最大の自然数を求めます。これは 1 です。 2 であれば、4 となり、2 を越えてしまいます。この最大の数を縦に書きます。

$$\begin{array}{r l}
 & \ \ 1 \\
1 & \sqrt{2\ \ 56} \\
1 & \ \ 1 \\
\hline
2 & \ \ 1\ \ 56
\end{array}$$

(3) つぎは $\begin{array}{r} 2\ x \\ \times \qquad x \\ \hline y \end{array}$ の掛け算で、答え y が 156 を越えない数 x を考えます。これは $x = 6$ と求められるでしょう。$26 \times 6 = 156$ となりますから、これを右側に書き、元の値から引き算をして、答えは 0 となり、これで終了です。

$$\begin{array}{r l}
 & \ \ 1\ \ 6 \\
1 & \sqrt{2\ \ 56} \\
1 & \ \ 1 \\
\hline
26 & \ \ 1\ \ 56 \\
6 & \ \ 1\ \ 56 \\
\hline
 & \qquad\ \ 0
\end{array}$$

結局、$\sqrt{256} = 16$ が得られることになりました。(4) もし引き算の答えが 0 でなかったら、繰り返します。2 桁づつ下ろしてきてます。(4) 仮に先ほどの引き算がゼロでなければ、$26 + 6 = 32$ ですから、$\begin{array}{r} 32\ x \\ \times \qquad x \\ \hline y \end{array}$ の掛け算で、答え y が下ろしてきた数を越えないように x を定めていきます。

さて問題はどうしてこのような手順で開平（平方根の計算）ができたのでしょうか？ヒントは $(10a + b)^2 = 100a^2 + 20ab + b^2$ で $a = 1$, $b = 6$ がいまの場合に対応しているのです。問いとして、上の $\sqrt{2}$ を計算してみてください。

また計算機のアルゴリズムを考えてみることもプログラムのよい練習になります。

第 3 章

場合の数、数え上げの方法

3.1　場合の数、数え上げ

階乗（かいじょう、ファクトリアル）の記号とは $n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1(n \geqq 1), 0! = 1$ とする。「r 種類の箱に n 個のボールの中から選んで入れること」が場合の数を計算するための基本のことがらです。
順列「区別のある箱」に「区別のあるボール」を入れる
重複順列「重複を許してひとつの箱に何個入れてもよい場合」という条件で取り出された場合の数。

$$n^r = \overbrace{n \times n \times \cdots n}^{r} \tag{3.1}$$

順列(permutation)「重複は許さずにひとつの箱には高々ひとつしか入れない場合」という条件で取り出された場合の数。

$$\overbrace{n(n-1)(n-2)\cdots(n-(r-1))}^{r} = (n)_r = \frac{n!}{(n-r)!} \tag{3.2}$$

組合せ「区別のある箱」に「区別のないボール」を入れる
重複組合せ「重複を許してひとつの箱に何個入れてもよい場合」という条件。

$$\frac{n(n+1)(n+2)\cdots(n+(r-1))}{r!} = \binom{n+r-1}{r} = \binom{n+r-1}{n-1} \tag{3.3}$$

組合せ(combination)「重複は許さずにひとつの箱には高々ひとつしか入れない場合」という条件。

$$\frac{n(n-1)(n-2)\cdots(n-(r-1))}{r!} = \frac{(n)_r}{r!} = \binom{n}{r} = \binom{n}{n-r} \tag{3.4}$$

問 3.1.1
箱とボールがそれぞれ $n=5, r=2$ のとき、重複順列、順列、重複組合せ、組合せについてすべての場合を書き並べてみなさい。そして答えが一致していることを確かめよ。たとえば 5 個のボールを a,b,c,d,e としてそれぞれ書き上げてみなさい。

3.2　パスカル三角形

数式 $(a+b)^3$ を展開してみれば、

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

となることは知られている。ここでそれぞれの係数をみると、$1,3,3,1$ となっている。$(a+b)^3 = (a+b)\times(a+b)\times(a+b)$ であるから、たとえば $3a^2b$ という項は 第 1 項から a, 第 2 項から a, 第 3 項から b としたときには a^2b が得られ、これを他の取り出しかた、第 1 項から a, 第 2 項から b, 第 3 項から a ととしても同じ

a^2b を得る。もうひとつの取り出し、第 1 項から b, 第 2 項から a, 第 3 項から a として、これらの 3 通りがすべて a^2b になるから、結局 a^2b の係数は 3 で、$3a^2b$ となることがわかる。つまり次のような塊りとしてくくることで理解できるでしょう。

$$\begin{array}{llllll} & & +aab & +abb & \\ (a+b)^3 = & aaa & +aba & +bab & +bbb \\ & & +baa & +bba & \end{array}$$

いくつかの項を計算してみます。最初の 0 次（ゼロ乗）が 0 となるのは計算ではなく、このように定めると考えていてください。順次に計算していけることがつぎの定理にある命題に結びついてくるはずです。

$$\begin{array}{llllllll} (a+b)^0 \cdots\cdots & 1 \\ (a+b)^1 \cdots\cdots & 1 & 1 \\ (a+b)^2 \cdots\cdots & 1 & 2 & 1 \\ (a+b)^3 \cdots\cdots & 1 & 3 & 3 & 1 \\ (a+b)^4 \cdots\cdots & 1 & 4 & 6 & 4 & 1 \\ (a+b)^5 \cdots\cdots & 1 & 5 & 10 & 10 & 5 & 1 \\ (a+b)^6 \cdots\cdots & 1 & 6 & 15 & 20 & 15 & 6 & 1 \end{array}$$

三角形に並べられてこの表をパスカル三角形とよびます。パスカルは確率論ではよく出てきますが、日本の有名な江戸時代の数学者、関孝和もこの関係式を導いています。２項係数をもちいれば、

$$\begin{array}{llllllll} (a+b)^0 \cdots\cdots & \binom{0}{k} \\ (a+b)^1 \cdots\cdots & \binom{1}{0} & \binom{1}{1} \\ (a+b)^2 \cdots\cdots & \binom{2}{0} & \binom{2}{1} & \binom{2}{2} \\ (a+b)^3 \cdots\cdots & \binom{3}{0} & \binom{3}{1} & \binom{3}{2} & \binom{3}{3} \\ (a+b)^4 \cdots\cdots & \binom{4}{0} & \binom{4}{1} & \binom{4}{2} & \binom{4}{3} & \binom{4}{4} \\ (a+b)^5 \cdots\cdots & \binom{5}{0} & \binom{5}{1} & \binom{5}{2} & \binom{5}{3} & \binom{5}{4} & \binom{5}{5} \\ (a+b)^6 \cdots\cdots & \binom{6}{0} & \binom{6}{1} & \binom{6}{2} & \binom{6}{3} & \binom{6}{4} & \binom{6}{5} & \binom{6}{6} \end{array}$$

となります。

3.3　２項係数

ここで２項係数とはある自然数 n に対して $k = 0, 1, 2, \cdots, n$ について

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

と表したものです。したがって一般式では

$$(a+b)^n = \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b^1 + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{k}a^{n-k}b^k + \cdots + \binom{n}{n}a^0 b^n$$

と表現できます。なぜ２項という理由は、$(a+b)$ の２つの項を展開しているからで、$(a+b+c)$ とか $(a+b+c+d)$ などの３項、４項もあって当然です。もし興味があれば、調べてみることもよいでしょう。

定理　つぎの関係式が成り立つ；

1. $\binom{n}{r} = \binom{n-1}{n-r} + \binom{n-1}{r}$ $\cdots\cdots\cdots$ （加法性）
2. $\binom{n}{r} = \binom{n}{n-r}$ $\cdots\cdots\cdots$ （対称性）

問 3.3.1
加法性 $\binom{n}{r} = \binom{n-1}{n-r} + \binom{n-1}{r}$ が成り立つことを確かめ、さらにこの式は、どういうことを意味しているか、パスカル三角形の表で考えなさい。

問 3.3.2
つぎの式は定義に代入して計算すれば得られますが、パスカル三角形の表から考えてみなさい。さらに積和の計算式, $\sum_{k=1}^{n} k(k+1)$ などとの関係を考えてみなさい。

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-2}{r-1} + \cdots + \binom{r-1}{r-1}$$

問 3.3.3
定義の式を代入して計算してみて確かめなさい。

1. $\binom{m}{0}\binom{n-m}{r} + \binom{m}{1}\binom{n-m}{r-1} + \cdots + \binom{m}{r}\binom{n-m}{0} = \binom{n}{r}$
2. $1 - \binom{n}{1} + \binom{n}{2} + \cdots + (-1)^n \binom{n}{n} = 0$
3. $\binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + n\binom{n}{n} = n2^{n-1}$

問 3.3.4
トランプ 52 枚から 5 枚を選ぶポーカーで、場合の数を考えなさい。

(1) すべての場合は何通りか？答え $\binom{52}{5}$ 約 259 万通り
(2) ローヤルフラッシュ（同じ種類の絵札、(10,J,Q,K,A)）
(3) フォーカード（4 枚が同じ数、$x \neq y, (x,x,x,x,y)$)、答え $13 \cdot 12 \cdot 4$
(4) ストレート（種類を無視して、続きの番号）答え $9 \cdot 4^5$
(5) スリーカード（3 枚が同じ数, (x,x,x,y,z)）答え $13\binom{12}{2}4 \cdot 4^2$
(6) ツーペア（2 組の同じ数, (x,x,y,y,z)）答え $\binom{13}{2}11 \cdot 6 \cdot 6 \cdot 4$
(7) ワンペア（2 枚のみ同じ数で残りはバラバラ, (x,x,y,z,w)）答え $13\binom{12}{3}6 \cdot 4^3$

(補足) とくに見かけは大したことがないような $\binom{52}{5}$ の値が驚くほどの約 259 万通り (正確には 2598960) にも大きな数になりますが、この概算の計算を試みてみます。スターリングの公式

$$n! \sim n^n e^{-n}\sqrt{2\pi n}$$

では

$52!$
$= 80658175170943878571660636856403766975289505440883277824000000000000$
$\sim 8.05290203838866 * 10^{67}$

このようなたいへんな計算もパソコンのレベルでも計算できます。

問 3.3.5
いままで 2 項係数の表し方は ${}_nC_k$, C_k^n とも表したかもしれませんが、同じものです。しかしこれを拡張した意味を上の式ではこめているのです。たとえば、自然数 n のかわりに 実数 α としても意味をもちます。

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!} \quad k = 0,1,2,3,\cdots$$

これをもちいて、$k = 0,1,2,3,\cdots$ に対して $\binom{-1}{k}, \binom{-2}{k}$ などを計算してみなさい。どんなことが気づきますか？ パスカル三角形と関係があるのです。