

自然科学の歩き方
先進工学部

武藤恭之

工学院大学
基礎・教養教育部門

目次

第1章	データの表現方法	2
1.1	自然科学の考え方	2
1.2	データとその可視化	3
1.3	数学における「関数」の概念とグラフ	5
1.4	グラフの読み方	6
1.5	演習課題	7
第2章	測定データとモデル化	8
2.1	現象を表すモデルとは	8
2.2	データの推定	8
2.3	モデルに求められる条件	10
2.4	二乗誤差	12
2.5	演習課題	14
第3章	モデルパラメータの推定	15
3.1	二乗誤差による最適なモデルパラメータの推定	15
3.2	演習課題	16
第4章	最小二乗法の定式化	17
4.1	微分を用いた最小二乗法の定式化	17
4.2	演習課題	19
第5章	科学的な文章の書き方	20
第6章	レポート・論文の構成	21

第1章

データの表現方法

1.1 自然科学の考え方

自然科学の研究は、自然現象をよく観察することから始まる。その観察の方法は様々で、様々な動植物をスケッチする・星の明るさや色を調べる・何かをたたいて音を聞く、など、挙げていけば枚挙に暇は無い。

自然科学の大きな目標は、「自然現象を人間が理解する」ことである。では、何をもって「理解した」と言えるのか。これは、非常に難しい問題である。一つの方法としては、自然を様々に分類してみる、ということが挙げられる。例えば、植物を系統的に分類したり、動物の行動をまとめあげてみたり、星のカタログを作ってみたり、といったことが行われている。このことに依って、雑多に見える自然をある程度大きな分類のもとにまとめ上げていくことができる。

自然を分類することができると、その裏に潜む法則性を見出していくことが出来る。例えば、星・太陽・月などの動きを調べることで、まず、どんな天体も東から昇って西から沈む、ということが分かる。ここから、かつては、地球の周りを天体が回っている、という「予想」を立てられた。しかし、季節ごとの惑星の位置を詳しく調べると、地球の周囲の単純な円の上を回転していると思うとうまく説明できず、「周転円」という、別の円を考えることで惑星の動きを理解しようとする試みがなされた。ところが、この考え方は非常に複雑であり、実は、「太陽の周囲を地球が回転している」と考えた方が、より簡単に天体の運動を表現できるということが、後に分かってきた。これがさらに万有引力の概念につながり、今日の物理学の基礎の一つである、重力の理論につながっていく。このようにして、人間は、「なぜ星の運動が起こっているのか」ということを理解しようとしてきた。

このように、多くのデータの蓄積の上に、人間の理解が成り立っているということが分

かる。つまり、自然をよく観察するという事は、自然科学の研究を進める上で非常に重要なことである。さらに、観察をした結果を数値的に表現することが出来ると、そこに数学という一般的な道具を持ち込めることになる。数学は、「数」を扱うための、高度に抽象化された手法であり、「誰がやっても、同じ方法で行えば同じ結果が出る」という、強い特徴を持っているものである。そこで、自然科学の色々な分野では

- 自然現象を観察・測定し、数値として表現する
- その数値の間に関連性を見出し、自然界の法則を理解する

ということが行われる。

もちろん、自然現象は様々で、中には数値として表現できない（正確には、おそらく、人間がまだ数値として表現する方法を知らない）現象も多くある。そのような現象では、現象を言葉として表現した上で、そこから何らかの関連性を見つけていくという方法が取られるが、一方で、いかに数値化して定量的に物事を表すか、という方向性も重要になる。

この講義では、「現象を数値化する方法」はすでに確立しているものとして、それらの数値からいかに法則性を見出していくか、という部分の方法論を主に扱う。実験や観察から得られるデータは大量にあり、単にその数値を眺めていても、何も理解することは出来ない。それらの数値をいかに整理して、人間の頭でわかる形にするか、というところが、自然を理解する上で重要なことになる。

1.2 データとその可視化

実験や測定は、何のために行うのだろうか。自然を理解するためだが、何をもって自然を理解したと言えるのか。一つの答えとしては、「二つの量の間に関係があること」を見いだせば、それは自然の理解に一步近づいたことになる。つまり、実験や測定を行う一つの目的は「ある二つの量の間に関係を見出すこと」といえるだろう。

「二つの量の間に関係を見出す」には、どうすればよいだろうか。例えば、A という量と B という量があったとする。この時、例えば、「A を増やすと、B も増える」というのは、A と B の間の「関係」であるが、この関係を自然現象の中から見出していくことが、実験や測定の目的である。そのためには、「A を少し増やしたら、B はどうなるか」、そして、「A をさらに増やしたら B はどうなるか」といった具合にして、少しずつ片方の量を変えていき、もう片方の量がどのように変化をするかを見ていけばよいだろう。結果として、「A を増やすと、B も増える」ということが分かるかもしれない。

実験などの測定結果は、例えば、表にしてあらわすことができる。表 1.1 は、ある実験

V [V]	1.50	3.00	4.50	6.00	7.50	9.00
I [A]	5.64×10^{-2}	1.12×10^{-1}	1.86×10^{-1}	2.22×10^{-1}	3.25×10^{-1}	3.32×10^{-1}

表 1.1 ある電気抵抗を流れる電流と電圧の例

で見出された、電気抵抗を流れる電圧と電流の関係とする。実験としては、ある電気抵抗に電源をつなげ、電圧を少しずつ変えていき（表の例では、1.5 V から、1.5 V 刻みで 9 V まで、電圧という言葉が分かりにくければ、抵抗につなぐ電池を直列に増やしていったと考えても良い）、その時に電気抵抗を流れる電流の値がどうなったかを調べる、というものであると考えれば良い。

この結果を見るだけでも、「電圧を増やすと電流も増えていく」ということが分かるであろう。

しかし、「増やす」とか「減らす」というのは、少し曖昧さが残る表現である。せっかく測定データが数字として表現されているのだから、そのメリットをもっと活かしていきたい。そこで、さらに理解を一步進めるために、「どのくらい」電圧を増やしたら「どのくらい」電流が増えるのか、ということ調べていこう。

この時に役に立つのが、グラフによるデータの表現である。グラフとは、測定する量を縦軸・横軸の数直線の上に取り、測定した値を平面上の点として表すものである。このことによって、数値で表現されていた測定結果を、視覚的に表現することができ、（人間にとって）より理解がしやすい形に測定結果を表現することができる。

グラフを描く上で、最も基本的で重要な点を以下に列挙しておこう：

- 縦軸・横軸が何を表しているのかを、単位を付けて示す
- 縦軸・横軸の目盛り、およびゼロの位置をしっかりと示す
- 軸の線はまっすぐ描く
- 測定データの点は、分かりやすく示す
- グラフ全体に測定データがプロットされるよう、縦軸・横軸の目盛りの範囲と間隔を適切に設定する

いずれも、基本的なことであり、大学に入る前から習っていた事だろうが、実践で身に付けるにはなかなか時間のかかることである。

1.3 数学における「関数」の概念とグラフ

高校までの学習の中で、「グラフ」はどこに出てきたらうか。もちろん、いわゆる「理科」の中では様々なグラフが出てくるが、一方、「数学」の中でも「関数のグラフ」というものが出てきたらう。理科の科目では、二つの量の間の関係を示すもの（例えば、速度と位置など）としてグラフが出てきたらう。一方で、数学では「関数」というものの性質を表現するものとして、グラフが出てきた。

数学の中では「関数」という形で抽象化されているものの、数学の「グラフ」と理科の「グラフ」に、なんら違いはない。このことを理解するためには、数学で出てくる「関数」とは何か、ということ、もう一度理科の言葉で言い換えてみる必要がある。

関数とは、「ある数を別の数に変換するルール」である。例えば、「1 という数字を 2 に置き替え、3 という数字を 6 に置き替え、5 という数字を 10 に置き替える」というルールがあるとすれば、これは $\{1, 3, 5\}$ という数の集合に対して定義された関数であるといえることができる。数学では、様々な計算の方法があるので、関数を式を用いて表現することもできる。しかし、式を用いて表現するか、上記のように「ある数にはこの数に対応させる」というルールで表現するかは本質的では無い。どちらにしても、「一つの数を別の数に変える」という部分が本質である。

一方で、自然科学の実験でどのようなことをやっているのか、ということを考えてみよう。実験では、「ある量と別の量の間を探る」ということをやる。つまり、例えば、ある電気抵抗を持ってきて、「電圧がこの値ならば、電流はこの値になる」ということを調べ、電圧を色々な値に変化させながら、その関係を調べる。今の場合、電流も電圧もある数字で表現する方法が確立している。ということは、実験を行うとは、（この例でいえば）ある一つの電気抵抗という材料を用いて、「電圧を表す数字を、電流を表す数字に変換するルール」が何かを調べていると言って良いだろう。この「ルール」は、自然が持っている性質であり、人間は、実験に依って、そのルールが何か、ということを知ることができる。もう少し、数学の言葉と理科の言葉を対応させるとすれば、「電圧の値を変えながら、電流の値を調べる」とは「電圧の値を独立変数、電流の値を従属変数とするような関数の形を調べる」ということに他ならない。ここに、数学の授業で出てくる「グラフ」と、理科の授業で出てくる「グラフ」の間につながりが見えてくる。

1.4 グラフの読み方

自然科学では、様々なデータをグラフで表して、その振る舞いを調べることで、二つの量の間関係をつかんでいく、ということをしてきた。

逆に、グラフが与えられた時に、そのグラフが何を表現しているのか、ということをつかえるということも重要である。数学的には、あくまでも抽象化された数字の間関係として「関数」が定義されているので、そのグラフも、関数そのものの性質を表現するものとしての意味がある。しかし、理科で出てくるような自然現象を相手にする場合、測定値は必ず現実の現象としての意味を持っている。そこで、グラフも、現実の現象を表す手段としての意味を持っていくということに注意しよう。

例えば、縦軸が物体の位置、横軸を時刻というグラフを描いたとき、その傾きが大きければ、「1秒あたりの位置の変化が大きい」ということを表している。これはつまり、「その物体が速く動いている」ということを表している。

また、別の例を挙げれば、横軸に物体の大きさ、縦軸に物体の重さ（質量）を描いたグラフを描いた時、大きさが大きくなればその分物体は重くなるから、そのグラフは右上がりになるだろう、ということが容易に想像できる。

しかし、もしも、そのグラフが、大きさが大きくなっても質量が変わらない、というようなことがあったら、どのようなことを想像するだろうか。このようなことは実際に起こることである。物体は、たくさんの構成粒子から成っているが、構成粒子一つ一つの重さは変わらない。その数が変わらずに、全体の大きさが変わらないということは、構成粒子の間距離が離れ、物体が大きくなるにつれて、スカスカになっていっているということを表している。これは例えば、気体の温度を上げて、全体を膨張させるような時に起こることである。

このように、グラフの振る舞いを見て、現実の現象を想像できるかどうか、というところが、自然科学を学んでいく上での一つの重要なポイントとなる。この感覚を身に付けるためには、様々な現象を見て、測定・実験し、グラフを描くという経験を積んでいくことが重要になる。これから、たくさんの実験の授業などがあるが、その意義は、こういった自然現象に対する感覚を身に付けていくことにあると言えるだろう。

1.5 演習課題

状況

ある日、あなたは無性に電気抵抗を測定したくなった。

気がつくと、目の前には、一つの電気抵抗が転がっていた。これ幸いにと電気抵抗を測定しようと思ったが、これだけではどうしようもない。途方に暮れる。このまま、電気抵抗を測れずに人生が終わるのか…。

ところが、偶然にも、端子電圧が 1.5 V の電池をいくつか恵んでくれる人が現れた。

この人の恩に報いるためにも、是非とも電気抵抗を測定して報告書にまとめなければならない。

すぐにどこの家庭にでもある銅線を用意して大学の実験室に行ってみると、なんと、電流計と電圧計が無造作に置かれている。

これは運命だ。回路を組んで測定しよう…。

課題 1

以下のデータは、ある電気抵抗に電圧 V [V] をかけた時に流れる電流 I [A] の値を測定したものである。このデータを、横軸に V 、縦軸に I を取ったグラフに表せ。

V [V]	1.50	3.00	4.50	6.00	7.50	9.00
I [A]	5.64×10^{-2}	1.12×10^{-1}	1.86×10^{-1}	2.22×10^{-1}	3.25×10^{-1}	3.32×10^{-1}

課題 2

「状況」の文章から必要な情報を抜き出し、この実験の「目的」が何かを簡潔な文章で表せ。

第2章

測定データとモデル化

2.1 現象を表すモデルとは

自然科学の大きな目標は、様々な自然現象を理解していくことである。人間が理解できることというのは、そこまで多くは無いので、実験をして膨大なデータを得られたとしても、それを全て見ていくということは難しい。「木を見て森を見ない」という言葉があるが、自然の全体が大きな「森」だとすれば、それぞれの実験データのの一つ一つというのは、その森を構成する「木」である。実験データは、「ある値に対して、別の値がどうなるか」という個々の事情を教えてくれるが、単にその一つを見ていても、「結局それらの量の間関係は何か」ということは教えてくれない。

そこで、自然の理解のためには、それぞれの「木」の個性はいったん置いて、「森」がどのようなになっているか、ということ把握する必要が出てくる。そのために役立つのがグラフで、様々なデータを一つの点として、それらの点の全体像をまとめて見ることが出来る。その上で、その森の全体を「大雑把に」把握しよう、というのが、「モデル化」と呼ばれる考え方である。

例えば、前章の課題で描いたグラフを見てみよう。電圧を大きくすると、その分電流も大きくなっているということは、グラフを見るとすぐにわかるだろう。では、「どのように」大きくなっていくのか？この部分を把握することで、電圧と電流の関係に、より深く迫っていくことが出来る。

2.2 データの推定

データのモデル化をするうえでまず気を付けるべきなのは：

- 測定されていないデータを、人間は知ることはできない

ということである。つまり、前章のデータで言えば、電圧が 1.5 V の時や 3 V の時の電流の値は測定で知ることが出来ているが、その間の 2 V の時の電流の値はわからない、ということである。このようなデータは、再度測定するか、あるいは推定をするしかない。

データを再度測定することは、基本的にはいつでも可能である。しかし、再度測定することに依って、電圧が 2 V の時の電流の値を知ることが出来ても、今度は、2.5 V の時の値を知ることが出来ない、となって、堂々巡りに陥る。実数は無限にあるのだから、適当に言ったある値の電圧に対し、必ず電流を求められるような測定データの表を作るとは、原理的に不可能である。

そこで、「データの推定」ということが必要になってくる。つまり、例えば、2 V の時のデータが無くても、それを他のデータから推定出来るだろう、という考え方である。例えば：

- 電圧が 1.5 V の時の電流値が 5.64×10^{-2} A だった
- 電圧が 3 V の時の電流値が 1.12×10^{-1} A だった
- したがって、電圧が 2 V の時の電流値は、これら二つの値の間になるだろう

というような考え方である。このことは、なんとなく直観的にはわかることだろう。

この、「直観的にわかる」ことの根拠は何だろうか。この部分で、グラフの重要性が見えてくる。まず気付くのは、グラフに表現された点を見ると、全体として、右上がり（電圧が大きくなると電流が大きくなる）の傾向にある、ということが見て取れる。さらに、その上がり方が、なんとなく、直線状になっており、あるところから急に電流が大きくなる、といったような傾向が見えない、ということがある。これらの二つから、「おそらく」電圧が 2 V の時も、1.5 V の時・3 V の時の二つの場合の間に、電流の値が来るだろう、ということが「想像」できるのだ。本当のことは、常に測定するまでわからない、というスタンスで居ることが大事だが、このように「結果の見当をつける」ことはさらに重要なことで、結果の見当が付くからこそ、追加の実験をやる時の計画を立てやすくなるのである。

さて、データの推定を行うのためには、どのようなことをすれば良いだろうか。ここで、数学の力を借りることになる。数式で表現できるような「関数」は、全ての実数（あるいは、複素数）に対して計算できるようなものであることが多い。そこで、測定されたデータの振る舞いを最も良く表現するような数学の式を見つけてくれば、その式を元に、測定の無いようなところでも、測定値がどのようになるか、ということ、数字として予測することが可能になる。そこで、このような式をどうやって求めていくか、ということ

を次に考えていこう。

2.3 モデルに求められる条件

ある程度のデータがあれば、モデルを作り、データを推定することが可能になる。しかし、モデル作成の上で注意することがある。それは：

- 測定データを信用し過ぎないこと

である。自然現象を知る上で、実験・観察・測定といったことが非常に重要である、と言いつつ、このようなことを主張するのは不自然かもしれない。しかし、測定や実験は、人間が行うものであり、「完璧な測定」をすることは不可能である。つまり、測定や実験の値には、常に誤差が付きまとっている。誤差の原因としては、人間の使っている測定機器の限界かもしれないし、実験上の不注意かもしれないし、あるいは、人間の想定していない全く別の自然現象かもしれない。これらを挙げだすときりが無いが、測定データには、必ず誤差がついて回るものであり、逆に、誤差の評価ができないデータは全く意味が無い、と言っても過言ではない。

しかし、この講義の中では、それぞれの測定データの誤差について、詳細に立ち入ることとはしない。誤差やデータの扱い方については、実験の授業などでより詳しく学んでほしいことである。ここでは、データがある程度の誤差を持っていることを前提としたうえで、「モデル化」をどのように考えるか、ということを書いていこう。

まず、データに誤差があることを考えると、それぞれの測定データを完全に合わせるような関数を求めても意味が無い。実際、例えば、6つのデータ点があれば、5次関数を使うことで、それらのデータ点を完全に再現するようなモデルを作成することはできる。しかし、そのようなモデルは、誤差を考慮した場合に対応しきれない、という難点がある。つまり、誤差の範囲で測定値が変わった時に、モデルの形もまた大きく変わる、という可能性が高い。そうすると、本当に、測定されていない点での推定が出来るのかどうか、ということも分からなくなってくる。

そこで、モデルを作る際に、必要な条件としては：

- ある程度単純な式で、他のデータの予測ができること

が求められる。ここで、単純というのは、「モデルパラメータの数が少ない」と言い換えられる。

モデルパラメータとは何か、例えば、前章で導入した、電流と電圧の関係を表現するモ

デルを例にとって、考えてみよう。今、データ点が全部で6つあるので、電流 I は、電圧 V の5次関数として表現すれば、これらの測定値は（原理的に）全て説明できるようなモデルを作成することができる

$$I = c_0 + c_1V + c_2V^2 + c_3V^3 + c_4V^4 + c_5V^5 \quad (2.1)$$

しかし、これは、 I を V の関数として表すために必要な定数が $c_0 \sim c_5$ の6つもある。これらの、モデルの関数を決めるために必要な、不定の値のことを「モデルパラメータ」といい、パラメータは、測定値に基づいて決められるものである。今の場合、データの数に対してパラメータの数が多すぎる。 $c_0 \sim c_5$ のパラメータを決めるためには、少なくとも6つ以上の測定が必要であるから、測定データを全て使ってようやくモデルを決められる、ということになっている。

一方、グラフを見てすぐにわかることとしては、 V と I は比例していそうだ、ということであろう。そこで思い切って

$$I = aV \quad (2.2)$$

というモデルを立てることもできる。こうすると、モデルのパラメータは a の1つだけになる。もちろん、こんなことをすると、全ての測定値を与えられるようなモデルを作成することは不可能である。しかし一方で、パラメータの数が少ないので、例えば、電圧 3 V での電流の測定値があれば、それをもとに、電圧 1.5 V での電流の測定値が何になるかを推定することができる。電圧 1.5 V に対しては、電流の測定値があるので、モデルから推定した値と、実際の測定値を比較することができる。この二つがそこまでずれていなければ、もともと実験にも誤差があることを考えれば、このモデルで「測定していない電圧での電流の値」を推定することに対して、正当化の根拠を付けることが出来る。

もちろん、単純な関数では、その分、予測できることもそこまで多くは無いだろうし、本当はもっと複雑な自然法則を見落とす可能性もある。しかしそれは、現状のモデルで説明できない測定データが出てきたときに、少しモデルを複雑にして対応することができる。このようなモデルの複雑化が、時に、新しい自然の法則の発見へとつながっていく。もともと複雑すぎるモデルを立てていたのであれば、それに手を加えてもっと複雑にすると、ややこしくなりすぎ、結局、何が分かったのか、という肝心の部分が分からなくなってしまう。

モデルには、「全ての測定データをそれなりに説明できる複雑さ」と「モデルの予測がはっきりとわかる単純さ」の、両方が必要である。この兼ね合いの感覚は、実践の中で培っていく必要があるが、現段階では、「良いモデルとは単純なモデルである」ということを強調しておこう。

2.4 二乗誤差

では、モデルの善し悪しは、どのように判断するのか？繰り返せば、モデルの善し悪しは主に：

- 測定データをそれなりに再現できる
- 式として単純である

という点にあるが、特に、最初の点が重要である。ここで、「それなりに」という、少し曖昧な言葉が出てきているが、測定データの再現性を表現するための量は何か、ということを考えてみよう。

再び、前章の電圧と電流の測定データについて考える。式を少し簡単に書くために、 V_i ($i = 1, 2, 3, 4, 5, 6$) を電圧の測定値、その時の電流の測定値を I_i ($i = 1, 2, 3, 4, 5, 6$) としよう。例えば、 $V_1 = 1.5$ [V], $I_1 = 5.64 \times 10^{-2}$ [A] である。電流と電圧の関係について

$$I = aV \quad (2.3)$$

というモデルを立てたものとする。ここで、 a がモデルパラメータになり、ひとまず、人間が「それらしい値」として、測定値から決めたものだとしよう。このモデルが予測する、 V_i における電流の測定値を、 \tilde{I}_i とすると

$$\tilde{I}_i = aV_i \quad (2.4)$$

となるから、実際の測定値とのずれは

$$\tilde{I}_i - I_i \quad (2.5)$$

である。測定値が全部で6つあるので、これら全てでずれを計算すると、「ずれが大きければ再現性が悪い」ということになる。

しかし、6つの量をまとめて扱うのは少し大変なので、この「再現性の悪さ」を、1つの数値で表現できないだろうか。最も単純なのは、この6つの「ずれ」の値を全て足し合わせることだが、これは良くない。なぜなら、「ずれ」の値は正にも負にも成り得る量だからである。正の量と負の量を足した時、それぞれのデータのずれは大きいのに、見かけとしてずれが小さくなってしまいうように見えることがありうる。

そこで、二乗誤差と呼ばれる量（あるいは、それを少し変形した量；平均二乗誤差など）が使われることが多い。これは、それぞれのデータ点での「ずれ」の量の二乗を取って、

足し合わせたものである。二乗誤差の値を E とした時、数式としては：

$$E = (\tilde{I}_1 - I_1)^2 + (\tilde{I}_2 - I_2)^2 + \dots = \sum_{i=1}^6 (\tilde{I}_i - I_i)^2 \quad (2.6)$$

と表される。ずれの二乗ならば、必ず正になる量なので、見掛け上ずれが小さくなるような心配は無い。この E の値が大きければ再現性が悪く、小さければ再現性の良いモデル、ということになる。

ただし、はじめに注意したように、パラメータの数を多くすれば、必ず測定値を全て説明できるようなことになるので、「再現性が良すぎるモデル」にも注意が必要である。そのような時は、そもそものモデル化のやり方に間違いが無かったか、あるいは測定データそのものがおかしくないか、ということ念入りに調べなければならない。

2.5 演習課題

測定データ

V [V]	1.50	3.00	4.50	6.00	7.50	9.00
I [A]	5.64×10^{-2}	1.12×10^{-1}	1.86×10^{-1}	2.22×10^{-1}	3.25×10^{-1}	3.32×10^{-1}

課題

- (1) 前回の課題で作成したグラフをもとに、電圧 V と電流 I の関係を最もよく表すと思われる直線を、点線で かきこめ。ただし、グラフは原点を通るものとしてよい。
- (2) かきこんだ点線を $V = aI$ という式で表す。グラフから数値を読み取り、傾き a を計算せよ。
- (3) この直線と実際の測定データの間のずれを、二乗誤差を用いて評価せよ。ただし、二乗誤差 E は、以下の式で表される。

ある電圧 V に対し、電流の測定値を I_{mes} とし、電流の推定値 (かきこんだ直線が表す電流の値) を I_{est} とする時

$$E = \sum (I_{\text{mes}} - I_{\text{est}})^2$$

ここに、総和は、全ての電圧の値について取るものとする。

第3章

モデルパラメータの推定

3.1 二乗誤差による最適なモデルパラメータの推定

前回、モデル化の考え方について学び、モデルの再現性を定量的に表現する量として、二乗誤差を導入した。

モデルを作成する際に大事なのは、モデルパラメータをどのように決めるか、という点である。そのための指針として、二乗誤差を使う。つまり、「データの再現性を最も良くする」という条件は、「二乗誤差が最小になるようにモデルパラメータを決める」ということになる。

再び、これまでも使っている電流 I と電圧 V のデータをもとに考えよう。まず、単純なモデルとして

$$I = aV \quad (3.1)$$

という形のモデルを考える。この時、二乗誤差 E は

$$E = \sum_{i=1}^6 (\tilde{I}_i - I_i)^2 = \sum_{i=1}^6 (aV_i - I_i)^2 \quad (3.2)$$

となる。ここで、右辺は、 V_i も I_i も測定量であるので、すでに知られている量であり、決まっていない量は a のみである。そこで、 a の値を色々と試して、 E の値が最小になるような a の値を推定すれば良い。ここでは、演習課題としてその計算を実際に行ってみて、 a の値に依って、二乗誤差がどのように振る舞うかを調べてみよう。

3.2 演習課題

測定データ

V [V]	1.50	3.00	4.50	6.00	7.50	9.00
I [A]	5.64×10^{-2}	1.12×10^{-1}	1.86×10^{-1}	2.22×10^{-1}	3.25×10^{-1}	3.32×10^{-1}

課題

電流 I と電圧 V の間に、 $I = aV$ という関係があるものと推定し、今回の測定値を最もよく表す a の値を探することを考える。

- (1) a の値を 0.02 から 0.01 刻みで、0.06 まで変化させる。それぞれの a の値について、二乗誤差 E を計算し、以下の表に書き込め。

a の値	0.02	0.03	0.04	0.05	0.06
E の値					

- (2) 今回調べた a の値の中で、最も良く測定データを表すものは何か。その a を用いた $I = aV$ の直線を、グラフ中に 実線 でかきこめ。

第4章

最小二乗法の定式化

4.1 微分を用いた最小二乗法の定式化

この章では、最小二乗法の数学的な定式化を行う。ただし、一般的にやると難しいので、これまでに一貫して用いている、電流と電圧のデータを使い、モデルとして

$$I = aV \quad (4.1)$$

という形の式を使うものとしよう。

最小二乗法とは、モデルパラメータの推定の際によく用いられる手法で、「二乗誤差を最小にせよ」という条件で、モデルパラメータを決めることである。この考え方自体は、すでに3章で扱った。3章では、そのために a として色々な値を試して二乗誤差の値を調べたが、「最小化」という作業は、微分を用いることで簡単に行うことができる。ここでは、このことを見ていこう。

今回の測定データとモデルに対して、二乗誤差 E は

$$E = \sum_{i=1}^6 (aV_i - I_i)^2 \quad (4.2)$$

と表される。ここで、 V_i および I_i が測定値であり、 a が推定したいパラメータを表すことから、この式は、「モデルパラメータ a を独立変数とする関数」と見ることができる。つまり、独立変数を明示して書けば

$$E(a) = \sum_{i=1}^6 (aV_i - I_i)^2 \quad (4.3)$$

となる。そして、問題としては、これを a の関数と見た時に、 $E(a)$ を最小にするような a の値は何か、ということになる。

この式を良く見ると、 $E(a)$ は a の二次式になっている。つまり、 $E(a)$ は単純な二次関数であり、しかも、 a^2 の部分が

$$E(a) = (V_1^2 + V_2^2 + V_3^2 + V_4^2 + V_5^2 + V_6^2)a^2 + \dots \quad (4.4)$$

と書けることから、二次の項の係数が正なので、必ず最小値を持つということが分かる。このように、今回の場合は、 $E(a)$ がただ一つの最小値を持つということが保証されているが、これは、常にそうなるとは限らない。今回のモデルは、たまたま、パラメータの二次関数になるようにモデルが設定されている、ということである。

このまま、平方完成を用いて $E(a)$ を最小にする a を求めても良いが、それだと $E(a)$ が二次式だという性質を使っているので、ある程度一般性を保つために、微分を用いた計算で進めよう。最小となる a は、 $E(a)$ の微分がゼロとなる点として求められる

$$\frac{dE}{da} = \sum_{i=1}^6 2V_i(aV_i - I_i) = 0 \quad (4.5)$$

こうなると、すでに a の一次方程式となっており

$$a = \frac{\sum_{i=1}^6 V_i I_i}{\sum_{i=1}^6 V_i^2} \quad (4.6)$$

と、 $E(a)$ を最小にする a の値が求められるということになる。

4.2 演習課題

測定データ

V [V]	1.50	3.00	4.50	6.00	7.50	9.00
I [A]	5.64×10^{-2}	1.12×10^{-1}	1.86×10^{-1}	2.22×10^{-1}	3.25×10^{-1}	3.32×10^{-1}

課題

測定データから、電圧 V と電流 I の間に、 $I = aV$ という関係があるものとして、測定値を最もよく表す a の値を推定したい。

そのために、以下の式で表される二乗誤差 E を最小にするような a を求めることを考える。

ある電圧 V に対し、電流の測定値を I_{mes} とし、電流の推定値 (かきこんだ直線が表す電流の値) を I_{est} とする時

$$E = \sum (I_{\text{mes}} - I_{\text{est}})^2$$

ここに、総和は、全ての電圧の値について取るものとする。

以下の問いに答えよ。

- (1) 3章で作成した、 a と E の表を、グラフに示せ。
- (2) 定義式をもとに二乗誤差 E を a の関数として式で表し、そのグラフの概形を、前問で作成したグラフ上に実線で示せ。
- (3) E を最小にする a の値を、式に基づいて計算せよ。

第5章

科学的な文章の書き方

準備中…

第6章

レポート・論文の構成

準備中…