

# 計算機を用いた実験データの解析

## 1 目的

測定した量には必ず誤差が伴う。(誤差と雑音(ノイズ)とは区別すること。)誤差には統計的な誤差や系統的な誤差などいろいろな種類がある。系統的誤差は個々の測定に依存するものであり個々の測定の十分な理解が必要となる。一方、統計的な誤差は最も基本的なもので、熱等によるランダムな雑音(ノイズ)のために測定値がばらつく偶然誤差や、現象が確率的に起こるために生じる統計誤差などがあり、これらは数学的にどのように取扱うかはよくわかっている。

本実験では電気回路等で生じる雑音を含んだ電圧データを取得し、それを計算機を用いて解析することを通して誤差の性質や誤差伝播法則など誤差を含んだデータの扱い方を習得する。(確率的な要因による統計誤差や Poisson 分布については「放射線」の実験で詳しく学ぶ。)

## 2 原理

### 2.1 母集団の平均と分散

取得できるデータの数は有限であり、ある集団(母集団)に属するデータの一部を抽出したものに過ぎない。実際には母集団の性質は未知なのであるが、充分にたくさんの個数を抽出したデータはほぼ母集団の性質を表しているとしてよいだろう。取得したデータの頻度分布を調べると、分布の性質がわかる。分布の平均と標準偏差は最も確からしい値と誤差の大きさを反映していると考えられる。

$i$  番目のデータを  $d_i$ 、データの個数を  $N$  と書いたとき、平均  $m$  と標準偏差  $\sigma$

$$m = \frac{\sum_{i=1}^N d_i}{N} \quad (1)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (d_i - m)^2}{N}} \quad (2)$$

である。 $\sigma$  の計算では  $m$  ではなく真の値を用いるのが正しいが、真の値が未知であるので  $m$  で代用している。そのことを考慮すると母集団の標準偏差は、データの不偏標準偏差

$$\sigma_s = \sqrt{\frac{\sum_{i=1}^N (d_i - m)^2}{N - 1}} \quad (3)$$

で推定するのがよい。

### 2.2 $\chi^2$ 法

物理実験では求めたい物理量が直接には測定できず、一連の測定結果からデータを解析して求めるという場合も多い。基礎物理実験(理科実験)のなかでは「電流の熱作用」、「ヤング率の測定」、「プランク定数の測定」、「放射線」などがそれに相当する。これらの実験の解析では最小二乗法を用いて実験データを説明する最適な物理量が定量的に評価される。また、測定されたデータにはそれぞれ実験誤差があるので、求める物理量も誤差をもつ。その関係は誤差の伝播法則で結ばれている。

電気回路等の雑音には時間的に徐々に変化する（ドリフトする）ようなものがあり、特に電源の投入直後によく起こる。そこでこの実験では測定した電圧データがそのような時間変動を持つかどうかを調べる。変化が時間の一次関数であると仮定し、最小二乗法の一つである  $\chi^2$ （カイ二乗）法を適用してその程度を評価する。

最小二乗法と誤差伝播については「レポートの作成」の後半に詳しく書かれている。事前によく読んでおき、実験開始前に理解しておくこと。

ある量の測定を  $N$  回行い、各測定値  $m_i (i = 1, 2, \dots, N)$  は誤差  $\sigma_i$  をもつとき、最小二乗法と誤差伝播の考え方をを用いると最良の推定値  $m$  とその誤差  $\sigma_m$  を求めることができる。特に  $\sigma_i$  がすべてほぼ同じ大きさ  $\sigma$  であるしてよいときには、 $m = \sum_i m_i / N, \sigma_m = \sigma / \sqrt{N}$  となる。

## 3 実験

### 3.1 データの取得

計測用の PC で電圧値のデータを取得する。計測システムを図 1 に示す。二人一組でグループ毎に順に使う事。

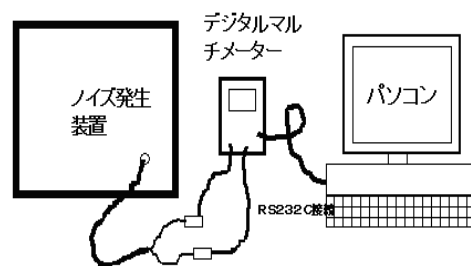


図 1: 計測用 PC の構成図

1. ノイズデータの発生  
ノイズ発生装置はデータに熱的な雑音を加え、ばらつきを生じさせている。ノイズ発生装置の取り扱いには付属の説明を参照すること。
2. デジタルマルチメーターの設定  
ノイズ発生装置のアンプ出力をオシロスコープで確認後、デジタルマルチメーター（Sanwa PC10）に接続する。デジタルマルチメーターのモードは DCV、Auto range で REL でないことに注意する
3. PC でのノイズデータの取り込み  
データの取り込みは次の順で行う。
  - (a) Login する。（TA または担当教員が説明する）
  - (b) スタートからプログラムで PCLink を選んでデータ取得プログラムを起動する。
  - (c) 取り込み間隔を秒間隔で 0 にする（最大スピードに対応）。
  - (d) 設定で Y 軸手動にしてデジタルマルチメーターの値が入るよう適当にセットする。

- (e) 設定で X 軸手動にして X 軸 1000 点にする。
- (f) 編集で全データの削除を行う。
- (g) データ取り込みで取り込み開始を選択する。
- (h) データを 1000 点以上取り込んだらデータの取り込みで 取り込み停止を行う。
- (i) ファイルからファイルへ保存を選ぶ。(配布された) USB メモリを差込み、マイコンコンピュータを選んで USB メモリのドライブを選ぶ。
- (j) 名前をつけて保存する。
- (k) ファイルから終了を選択。USB メモリを取り出して終了する。
- (l) 最後のグループは PC を Logout する。(スタートからシャットダウンを選び、さらにシャットダウンを選ぶ。) モニタ、マルチメータの電源も OFF にすること。

## 3.2 データの解析

ここからは解析用の PC を一人一台用いて実験する。解析のソフトとしては Microsoft EXCEL を使用する。EXCEL の基本的な使い方や用語などについては 5.6 を参照すること。

### 3.2.1 EXCEL の起動とデータの読み込み

1. Excel のアイコンをダブルクリックして起動する。
2. ファイルメニューから開くを選択
3. USB メモリのファイルを表示する。
4. ファイルの種類を全てのファイルにすると、ほしいファイルが見えるのでそれをダブルクリックする。

データは コラム A , B , C が時刻、測定値、測定した物の種類である。読み込んだデータは CSV 形式であるので EXCEL 形式に変えて保存しておくこと。(5.2 参照。)

計算機を用いると必要以上の精度で数値が表示されることがある。表示は書式設定で変更できる。(5.4 参照。値そのものが変わるわけではない。) 元々の数値は実験値なのであるから、結果としてレポートなどに示すときには有効数字がどれほどであるのかを考慮することが重要である。

### 3.2.2 取得したデータの表示

測定された全データを図 2 のような時系列でのグラフに描いてみよう。(5.3 参照。)

### 3.2.3 平均値と不偏標準偏差

測定された電圧値の最大値、最小値を求めてみよう。EXCEL ではそれぞれ max, min の関数が利用できる。

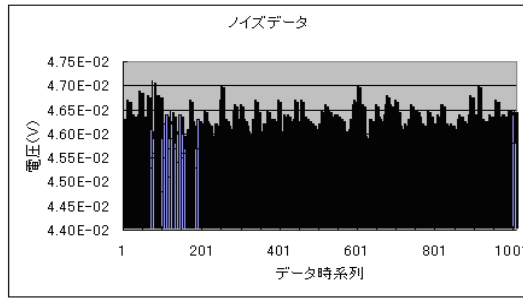


図 2: ノイズの時系列データ

### 3.2.4 電圧値の頻度分布

電圧値がどのように分散しているか頻度分布を調べてみよう。

1. 電圧値を最小値から最大値の間で適当な間隔の区間  $x_i$  に分け、区間配列を設定する。区間の数は 4 以上にとること。必ずしも等間隔でなくともよい。
2. データ配列と区間配列から frequency 関数を使って頻度分布  $O_i$  を計算する。この場合 Ctrl-Shift-Enter とすることを忘れずに。各区間には 5 個以上のデータが入るようにし、そうならない場合には区間を変更すること。
3. 頻度分布をグラフに描く。分布はどのような特徴をもっているか？
4. 頻度分布を中心値  $x_0$ 、幅  $s$ 、総数  $N$  の Gauss 分布  $\frac{N}{\sqrt{2\pi}s} \exp(-\frac{(x-x_0)^2}{2s^2})$  と比べてみよう。Gauss 分布には  $x_0, s$  の二つのパラメタがある。(  $N$  は頻度分布の計算に用いたデータの個数である。)  $x_0$  は頻度が最大になるような  $x$  の値から、 $s$  は半値幅 (FWHM, レポートの作成) を参照) から  $s = \text{FWHM}/2.35$  によって推定する。
5. normdist 関数を用いると各区間に対応する Gauss 分布の面積から期待数  $E_i$  が計算できる。(区間があまり広くなければ、Gauss 分布の面積は中央値  $\times$  幅で代用してもよい。) これを計算してグラフにも追加してみよう。
6.  $O_i$  と  $E_i$  から  $\tilde{\chi}^2 = \frac{1}{d} \sum \frac{(O_i - E_i)^2}{E_i}$  を計算すると、二つの分布がどれだけよく合っているかが判定できる (カイ二乗検定)。  $d$  は区間の数から自由度 (=2) を引いた数である。(データの数が 5 未満であるような区間はこの計算では使わないほうがよい。)  $x_0, s$  の値を変化させて、二つの分布を合わせてみよう。  $\tilde{\chi}^2$  の値が 1 程度なら Gauss 分布はデータをよく表しているといえる。(ただし、  $d$  があまり大きくない場合に判定するには  $\chi^2$  分布の表と比べる必要がある。詳細は参考文献を見よ。)

考察 Gauss 分布のパラメタ  $x_0, s$  は、元のデータの平均値  $m$  (average 関数)、不偏標準偏差  $\sigma$  (stdev 関数) や最大値、最小値とどのような関係になっているだろうか。

### 3.2.5 平均値の精度と扱うデータの個数との関係

平均値  $m$  が使用するデータの個数  $N$  によってどのようにばらつくかを調べよう。

$N$  が 5, 10, 20, 50, 100, 200 の場合に平均値を 5 個程度ずつ計算し、平均値  $m$  のばらつきが  $N$  についてどのように変わるかを調べよう。

### 3.2.6 時間的な変動の存在

$m$  の値に時間的な変動があるかどうか調べよう。 $\chi^2$  法を用いると、 $(x_i, y_i)$  の測定の組に対して、 $y = ax + b$  の関係を満たす最適な  $a, b$  の値を計算することができる。ここでは各時刻  $x_i$  ごとの電圧値としてデータを 100 個毎の 10 組に分けた  $m$  を  $y_i$  とする。 $m$  の値の精度 (誤差)  $\sigma_i$  は 10 個の  $m$  の値の不偏標準偏差によって評価する。 $m$  は  $N (=100)$  個の値の平均値なので、その誤差  $\sigma_i$  は不偏標準偏差  $\sigma$  と  $\sigma_i = \sigma/\sqrt{N}$  の関係にあることから評価する。時刻  $x_i$  はコラム A のデータから決めることができるが、知りたいのは変動であるのでデータは 1 秒に 2 個であることを利用して決めよう。

具体的な計算は以下のような手順で行なう。(計算式については「レポートの作成」の最小二乗法の項を参照すること。)

1.  $1/\sigma_i^2$  の値を計算しておく。
2. sum, sumproduct 関数などを利用して  $\sum(1/\sigma_i^2), \sum(x_i^2/\sigma_i^2), \sum(y_i^2/\sigma_i^2), \sum(x_i y_i/\sigma_i^2), \sum(x_i/\sigma_i^2), \sum(y_i/\sigma_i^2), \Delta$  の値を計算する。
3.  $a, b, \sigma_a, \sigma_b, \chi_{min}^2$  の値を求める。

$\chi_{min}^2/(N - f)$  の値を換算  $\chi^2$  (reduced  $\chi^2$ ) 値という。ここで  $N$  はデータ点の数、 $f$  は適合する曲線 (ここでは直線) の自由度の数 (ここでは 2) である。換算  $\chi^2$  の値は (他に誤差の要因がなければ) 1 程度になることが知られている。1 よりかなり大きい場合は、仮定した適合曲線が適切ではないことを、かなり小さい場合には誤差  $\sigma_i$  を過大評価している可能性が高いことを示唆している。

考察 時間変動はあったとしてよいかどうか、誤差の評価に基づいて 判断しなさい。

## 4 レポート

計算機はあくまで道具である。使い方を覚えるだけが目的ではない。従って、この実験のレポートでは実験方法はテキストにかかれたことを詳細に記述する必要はなく、実際に自分がやったことを簡潔にまとめること。(テキストや他人のレポートの丸写しは減点する。) 実験データの処理方法、誤差の取扱い等について実験から自分で考察したこと、自分でいろいろと調べてみたことを中心に記述すること。単に EXCEL の出力結果 (図や表) を貼り付けただけのようなレポートは不合格とします。

参考文献: 「実験データを正しく扱うために」 化学同人編集部 化学同人 (2007)  
「計測における誤差解析入門」 J.R.Taylor 東京化学同人 (2000)  
「実験法入門」 D.C.Baird ピアソンエデュケーション (2004)

---

## 5 Excel 入門

Excel は表計算ソフトであり、起動すると A,B,C,... の列と 1,2,3,... の行による巨大な表が表示される。列と行を指定して B1 のように決めたものをセルと呼び、各々のセルには数値や文字列などのデータを入力することができる。Excel では種々の関数やデータベースの機能を用いてデータを処理し、解析を行なってグラフや表を作成することなどができる。

## 5.1 関数

この実験で有用な EXCEL の関数として次のようなものがある。

min(数値 1, 数値 2,...)	引数リスト (配列も可) に含まれる最小の数値を返します。
max(数値 1, 数値 2,...)	引数リスト (配列も可) に含まれる最大の数値を返します。
average(数値 1, 数値 2,...)	引数 (配列も可) の (数学的な) 平均値を返します。
stdev(数値 1, 数値 2,...)	引数 (配列も可) の標準偏差の推定値を返します。
sum(数値 1, 数値 2,...)	セル範囲に含まれる数値をすべて合計します。
sumsq(数値 1, 数値 2,...)	引数 (配列も可) の 2 乗の和 (平方和) を返します。
sumxmy2(配列 1, 配列 2)	2 つの配列で対応する配列要素の差を 2 乗しさらにその合計を返します。
sumproduct(配列 1, 配列 2, 配列 3,...)	指定した配列の対応する要素間の積をまず計算し、さらにその和を返します。
frequency(データ配列, 区間配列)	データの頻度分布を縦方向の数値の配列として返します。 区間配列より 1 つ大きい配列を選択し、関数を記述しおわったら ctrl-shift-Enter とします。 結果は区間の最初の値以下のデータの数、1 番目より大きく 2 番目以下のもの、...、区間の最後の値以上の数が入ります。
sqrt(数値)	正の平方根を返します。
exp(数値)	$e$ を底とする数値のべき乗を返します。
power(数値, 指数)	数値のべき乗を返します。^ 演算子を使用してべき乗の指数を表すこともできます。この場合は、5^2 のように指定します。
normdist( $x, n, \sigma, \text{false}$ )	$n$ を中心とし、 $\sigma$ を幅とする Gauss 分布の $x$ に対する値を返します。分布は積分すると 1 になるような確率分布となっています。false を true にすると $x$ を上限とする積分値 (累積値) になります。
poisson( $m, n, \text{false}$ )	$n$ を平均とする Poisson 分布で $m$ 個の事象が起こる確率を返します。

## 5.2 データのファイルへの保存

作業終了時 (あるいは途中で) にはデータをファイルに保存しておく。その際、名前をつけて保存を使ってファイル形式を EXCEL 形式に変えておくこと。(二度目からは上書き保存でもよい。) 元の実験データは CSV 形式なので、そのまま上書き保存してしまうと EXCEL で入力した関数や作成したグラフなどは保存されない!!

1. ファイルから名前を付けて保存を選択。
2. 保存先を保存すべきところに移動する。
3. ファイルの種類を Microsoft Excel ブックにする。
4. 必要ならファイル名を変更する。
5. 保存をクリックする。

## 5.3 グラフの描き方

配列あるいは配列の組を選択しグラフアイコンをクリックすることによりグラフを作成することができる。メニューに従って種々のグラフを作成できる。軸や凡例などの表示形式は後でも変更することができる。実験データの表示では散布図がよく使用される。

1. データのはいつている最初のセルをクリックする。

2. 右のスクロールバーで必要なデータの一番下が見えるところまでスクロールし、シフトを押したまま最後のデータをクリックすることにより必要なデータを全て選択する。(キーボードから範囲を入力してもよい。)
3. グラフアイコンをクリックし、グラフを作成する。タイトル、横軸、縦軸のラベルを設定する。
4. グラフウィザード(メニューバーの下にある)をクリックする。メッセージに従い指定していくことによってグラフを表示することができる。

グラフのウィンドウをクリックして選択すると、メニューからグラフへのデータの追加などができるようになる。

## 5.4 データ書式を変える

数値などの書式変更は、右クリックまたは書式メニューから「セルの書式設定」を選び「ユーザー定義」または「指数」で「0.000E+00」とセットすれば有効数字を意識した表示ができる。

## 6 初心者の為の用語集

**セル** Excelの1つの欄のこと。列(A,B,C...)と行(1,2,3...)で指定される。セルを移動するにはマウスを動かして適当な位置でクリックするか、矢印キーを用いる、Enter,tab,shift+Enter,shift+tabのキーを用いるなどの方法がある。

**クリック** マウスを適当な位置に動かし、左ボタンを短く押すこと。

**右クリック** マウスを適当な位置に動かし、右ボタンを短く押すこと。

**ダブルクリック** マウスを素早く2回クリックすること。

**ドラッグ** マウスを適当な位置に動かし、左ボタンを押したままマウスを移動し、適当な位置で離すこと。最初にセルをクリックし、マウスポインタがセルの右下で十字のマークになった状態でドラッグすると、オートフィルの機能によって多数のセルに自動で入力することもできる。

**選択** マウスをクリック又はドラッグすることにより一つ又は複数のセルを指定すること。一つのセルを選択後、shiftキーを押しながら別のセルを指定すると二つのセルを左上、右下とする領域を一度に選択できる。又、ctrlキーを押しながらマウスをクリックすると現在選択されているものと離れたところにあるセルを同時に選択することができる。

**取消し** 式の入力などのさいに間違えて訳がわからなくなった場合には×のところをクリックすると元に戻る。

**配列** いくつかのセルの列あるいは行をまとめたものを配列という。B1:B1000のように表わす。

**関数** Excelではセルや配列の値を使って様々な計算を行うことができる。結果を表示したいセルをクリックし、上部の入力エリアに=を入力する(あるいはfxをクリックする)と関数が入力できる。関数はsqrt(2.5)のように直接値を記述しても、sqrt(B1)のようにセルを参照してもよい。引数としてmax(B1:B1000)のように配列を指定する(できる)場合もある。関数を組み合わせて複雑な式を作ることにもできる。(式の記法はFortranやBasicに似ている。)

**相対参照と絶対参照** セルのデータを関数などで参照する際、B1のように記述すると入力中のセルに対する相対的な位置として扱われる。定数の値として使う場合などは\$B\$1のように\$をつけて絶対位置で指定するとよい。F4を押すと切り替えができる。又、セルに名前をつけて参照することもできる。