

情報リテラシー第 11 回：最終課題「直線式の場合の最小 2 乗法」

1. はじめに

自然現象や科学計測には直線の式で近似できるものが沢山あり、しばしばその式の係数がわかっていると便利（調べておく必要がある）ことがある（例えば、固体の線膨張率、イオンクロマトグラフ分析における検量線など）。ここでは、対象として物理学実験の放射線計測で行なっているベータ線源と GM（ガイガー・ミュラー）計数管の検出器窓間の距離の関数として、1 分間の計数を測ったときに、良い近似でその計数が距離の 2 乗に反比例することを Excel と C 言語のプログラミングによる最小 2 乗法によって、客観的に確認することを本授業の最終課題として出題する。

2. 直線の式への最小 2 乗法の適用（回帰直線の求め方）

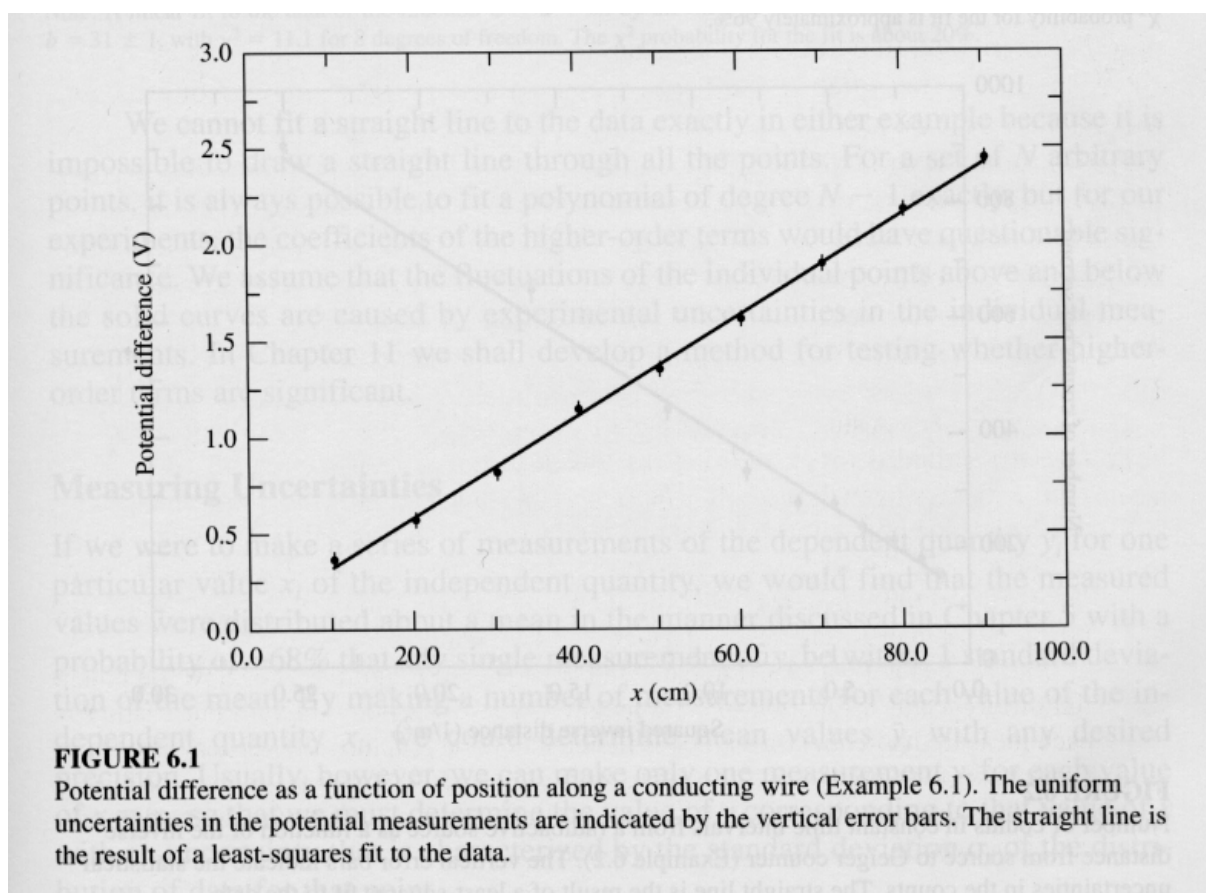


図 1. 直線に乗るデータと最小 2 乗法で求めた直線の例

(P. R. Bevington, D. K. Robinson, “Data Reduction and Error Analysis for the Physical Sciences”,
Third Edition, 2003 より転載)

上の図は説明を読むとわかるように抵抗線の両端にある電圧をかけたときの一端と途中の 1 点間の電位差を測りグラフ化したものを示している。各点の縦棒は期待されるある一定の

誤差棒を示しており、直線はデータから最小2乗法で求めた結果であるとある。このように理論的に完全に直線で表せる場合や近似が良い場合、客観的にこれから述べるこの計算法を使えば直線の関係式（例えば2変数を x, y とすれば $y=a+bx$ と表せるときの係数 a, b を求めることに相当する）を求められるが、短時間の実験や Excel などが側で使えない環境では、目分量で全てのデータを誤差内で繋げられるような直線をグラフ上で直線定規を当てて関係式を出していくことが多かったと思う。しかし、それは多少、主観が入ってしまい、数値読み取り誤差も入り、同じデータでも人によってしばしば結果が異なってしまう。

ところが、最小2乗法を用いれば完全に客観的かつユニークに直線（回帰直線）が決定できる。ここでは最も簡単な例として、各データの誤差が0または一定（決められない場合も含めて）のときを扱おう。

N 個の観測値 (x_i, y_i) ($i=1, 2, \dots, N$) があつたときに（ここで x は測定条件の変数で観測者が決められる一般に誤差の無い変数であり、 y は実験や観測で得られる物理量など誤差を含む測定結果である）、それらに対して直線の式 $y=a+bx$ をあてはめることを考える。最小2乗法は残差 $e_i=y_i-(a+bx_i)$ の2乗和

$$\chi^2 \equiv \sum_{i=1}^N e_i^2 \quad (1)$$

χ^2 （カイジジョウと読む）が最小になるように、係数 a, b を決める方法である。このように決めた直線をしばしば回帰直線ともいう。求める係数 a, b は次の偏微分式が0になることを満たすように決める。

$$\frac{\partial}{\partial a} \chi^2 = \frac{\partial}{\partial a} \left[\sum_{i=1}^N (y_i - a - bx_i)^2 \right] = -2 \sum_{i=1}^N (y_i - a - bx_i) = 0 \quad (2)$$

$$\frac{\partial}{\partial b} \chi^2 = \frac{\partial}{\partial b} \left[\sum_{i=1}^N (y_i - a - bx_i)^2 \right] = -2 \sum_{i=1}^N [x_i (y_i - a - bx_i)] = 0 \quad (3)$$

この連立方程式の解は、次のように与えられる。

$$a = \frac{1}{\Delta} (\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i) \quad (4)$$

$$b = \frac{1}{\Delta} (N \sum x_i y_i - \sum x_i \sum y_i) \quad (5)$$

ここで、

$$\Delta = N \sum x_i^2 - (\sum x_i)^2 \quad (6)$$

であり、 \sum は全て $\sum_{i=1}^N$ を表している。式の形からこれらは表計算によって求めるのに適していることがわかる。また、別に次のように表記することも出来る。

次の \bar{x} と \bar{y} の平均値、 $x - \bar{x}$ の平方和、 $x - \bar{x}$ と $y - \bar{y}$ の積和、

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i, \quad S_x = \sum_{i=1}^N (x_i - \bar{x})^2, \quad S_{xy} = \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad (7)$$

を使うと、 a, b は次の式によって求められる。

$$b = \frac{S_{xy}}{S_x}, \quad a = \bar{y} - b\bar{x} \quad (8)$$

(4), (5), (8)式の導出及びお互いに一致していることを確認することは各自に任せる。

3. 最終課題

本講義の HP に Excel ファイルで実際の放射線計測で得た 1 分間の計数値 $Y(r)$ (3 回の平均値にバックグランド補正を施したもの) が線源と検出器の窓 (円形) までの距離 r の関数として 2 組の本学科 1 年生によって得られたものを示している。原理的に、 $Y(r)$ は線源を点状にみなせるため、距離 r の位置における検出器が見込む立体角 $\Omega(r)$ に比例する。 r が検出器窓の半径 a より十分大きいとき ($r \gg a$)、

$$\Omega(r) \cong \frac{\pi \cdot a^2}{r^2} \quad (9)$$

で近似できるため、 $Y(r)$ は線源の放射能や検出器の検出効率などを含む定数を A として次式で表せる。

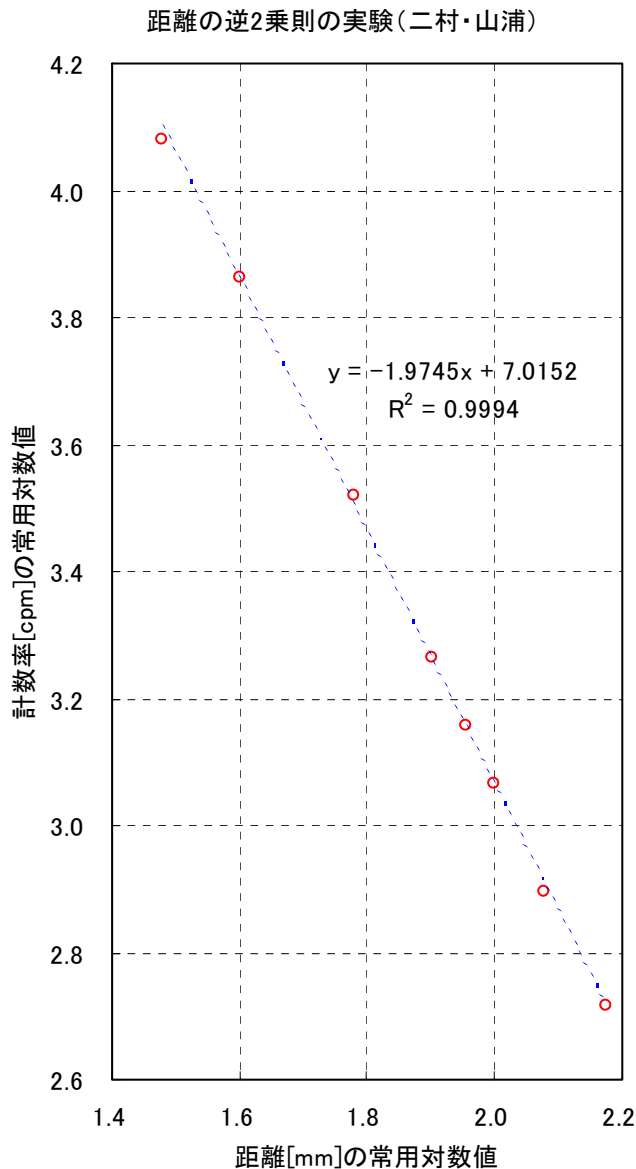
$$Y(r) = \frac{A \cdot \pi a^2}{r^2} \quad (10)$$

式(10)の両辺の対数をとれば、

$$\log Y(r) = \log(A\pi a^2) - 2\log r \quad (11)$$

この式は \log が常用対数 (10 を底とする対数) でも自然対数 (ネピア数 e を底とする対数) でも成り立つがここでは常用対数を使うことにする。ここで $\log Y(r)$ を y 、 $\log r$ を x とおけば、

$\log(A\pi a^2)$ は定数であるから、先に考えた直線の式 $y=a+bx$ で $b=-2.0$ の場合に当てはまる



ことがわかる。物理学実験では両対数グラフを用いてデータをプロットし、データが直線にのることを確かめた上で、実測で傾きが -2 に近くなることを確認している。その Excel を使って示したプロット例を1組のデータについて示している(図2)。

ここでは Excel に組み込まれてる近似曲線の機能を各対数値のデータの組み合わせに適用して直線での近似曲線を求めている。

つまり、グラフを作成しマウスで図を選択した後に、グラフ→近似曲線の追加で直線を選んで実行させる。結果の式や R^2 乗値(相関係数の2乗値; この値が1に近いほど2つ変数の相関は高いことを示す)はオプションタブでチェックを入れれば表示させることができる(図2)。式から x の係数は -1.9745 と期待される -2.0 に非常に近く、データと式との一致も非常に良い(相関が高い)ことがわかる。

図2. ベータ線源からの GM 計数管の入射窓までの距離を変えたときの計数率[cpm]の変化

課題1 もう一組の須山・吉田組のデータについても図2に示したような Excel での近似曲線を求めよ。

課題2 Excel の近似曲線の機能を用いずに、2節で説明した最小2乗法を適用して、2組のデータについて係数 a , b を求め、Excel の近似曲線で求めた値と一致することを示せ。

課題3 C 言語でデータを配列の初期化で代入し、for 文などを用い、最小2乗法のプログラムを作成し、2つのデータの組についてやはり係数 a , b を求め、上記2つの結果と一致す

ることを示せ。但し、桁落ちを避けるために計算に用いる変数は全て倍精度実数すなわち **double** で宣言せよ。レポートにはプログラムを **Word** 内にコピーしたものを入れ、結果については、スクリーンコピー(**Alt + Prt Sc** キーでコピー、さらに編集で貼り付け; **C 言語入門(1)** の付録で説明済み)で画面上の結果を貼り付けること。

以上の課題を9月11日(金) 17:00までに各自の担当教官にメールの添付ファイルで送付すること。課題1, 2は一つの **Excel** にまとめて、課題3は **Word** のファイルに説明文も加えて提出すること。

質問があれば、随時、担当の先生方を中心に問い合わせること。これで、今年の情報リテラシーは終わります。＼(^0^)/

Tips: 既に気がついている人も多いと思うが、コマンドプロンプトの画面では、↑キーを入力することで、前に入力したコマンドをたどって呼び出すことが出来る。逆に↓キーで以降に入力したコマンドに戻っていく。これはプログラムを修正しながらコンパイル・実行を繰り返しながら開発していくときなど非常に便利である。