地球物理学学生実験　誤差について

2001/10/4　**中原　恒**

目次

１．誤差の概念

1.1 測定と誤差

測定とは,「基準となる量（単位）との比較によって，物理量の大きさを数値で表現すること」である．測定には，直接測定と間接測定の 2 種類ある．直接測定とは「基準となる量との直接比較により物理量を求めること」であり，間接測定とは「直接測定した物理量の関数として間接的に物理量を求めること」である．

（例）直接測定… ものさしで長さをはかる，天秤で重さをはかる，など．
　　　間接測定… 円の半径を測定して面積を求める，直方体の３辺の長さと質量を測定し密度を求める，など．

一般に，測定された物理量（測定値）はさまざまな要因により真の値とは異なる．測定値と真の値との差は誤差と呼ばれる．しかし，実際のところ真の値は分からないので，測定値から真の値を推定し，その値との差を用いて誤差を評価することになる．

1.2 誤差の種類とその要因

誤差は大別して，系統誤差と偶発誤差（偶然誤差ともいう）の２種類に分類される．

(1)系統誤差

結果に一定の偏りを与える誤差．この誤差は，その原因がわかればあらかじめ避けたり，後で補正したりして除きうるものである．

（要因）１．根本理論（モデル）が誤っていること．
　　　　２．測定機器の構成や動作が不完全であること．
　　　　３．測定者の一定の癖．

(2）偶発誤差

全く偶然に不定な関係で生じる誤差．以下に挙げる要因のうち測定者の過失による部分は原則的に避けられる．一方，測定機器の精度限界による部分は除き得ない．この部分は必然的偶発誤差と呼ばれることもあり，誤差論の対象となるものである．

（要因）１．測定者の過失（うっかりミス）．
　　　　２．測定機器の精度限界．

1.3 有効数字

有効数字とは，測定値の信頼できる桁数までのことである．測定結果がどれくらい精密に求められたかを知るには，有効数字を見れば良い．そのため，得られた測定値を必要以上に小さい桁まで記してはいけない．測定機器の精度などを考慮し，どの桁までが信頼できるのかを十分に吟味する必要がある．また，四則演算による有効数字の変化に

も留意する必要がある．

1.4 精度

精度という言葉は，正確さ（accuracy）と精密さ（precision）の２つの意味を混同して使われていることが多いので気をつけたい．

正確さ (accuracy)

…測定値がどれくらい真の値に近いか，つまり系統誤差（偏り）の大小を示す．

精密さ (precision)

…測定値が有効数字で何桁まで求められるか，つまり偶発誤差（ばらつき）の大小を示す．

２．誤差の分布（統計的性質）

前節で述べたように, 測定値には何らかの誤差が含まれているため, 真の値から偏り, かつばらついている．このうち系統誤差（偏り）はすでに取り除かれて，必然的偶発誤差のみが含まれているものとする．以下では，誤差とは必然的偶発誤差のことを指すものとする．

2.1 母集団と標本

起こりうるすべての測定値からなる集合を母集団といい, それらの測定値が従う分布を母集団分布という．我々がある測定を n 回行うことは，母集団から n 個の測定値を得ることに対応する．この時，n 個の測定値の集合を，大きさ n の標本という．測定の目的は,「有限個の標本から母集団の性質（すなわち分布）を推定すること」と言える．

2.2 分布と確率

測定により母集団から測定値を取り出すものとする. 測定値が離散量（とびとびの値）の場合，測定値が x_i となる確率（頻度）を $P(X = x_i)$ と表わす．一方，測定値が連続量の場合，測定値が $[x, x+dx]$ である確率 $P(x \le X \le x+dx)$ を $f(x)dx$ で表わし，関数 $f(x)$ を確率密度関数と呼ぶ.

確率は，正またはゼロで，総和が１という性質がある：

$$P(X = x_i) \ge 0, \quad \sum_i P(X = x_i) = 1,$$

$$f(x) \ge 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1. \tag{1}$$

また，測定値が x よりも小さい確率 $F(x) = P(X \le x)$ を確率分布関数と呼ぶ．すなわち

$$F(x) = \sum_{t \le x} P(X = t),$$

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{2}$$

である．

2.3 平均と分散

ここでは，分布を特徴づける量である平均と分散を定義する．

母集団に対して，平均（期待）値（母平均と呼ばれる）は，

$$m \equiv \int_{-\infty}^{\infty} x f(x) dx, \tag{3}$$

分散（母分散と呼ばれる）は，

$$s^2 \equiv \int_{-\infty}^{\infty} (x - m)^2 f(x) dx \tag{4}$$

で定義される．分散は，平均のまわりでの測定値のばらつき（誤差）の大きさを示す量である．

一方，大きさ n の標本（$x_1, x_2, \cdots, x_n$）に対して，平均（標本平均）は

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \tag{5}$$

分散（標本分散）は

$$s^2 = \overline{(x - \bar{x})^2} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{6}$$

で定義される．しかし，n が小さい場合は次の不偏分散を用いる方が良い：

$$u^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 . \tag{7}$$

2.4 母集団分布

母集団に属する要素が従う分布を母集団分布という．ここでは，代表的な例として二項分布と正規（ガウス）分布を挙げておく．

(a)二項分布

「コインを投げて表が出るか裏が出るか？」や「さいころの目が奇数か偶数か？」など二者択一の問題は二項分布によってあらわされる．

ここでは，コインを n 回投げて k 個が表になる確率を求めてみよう．コインが表になる確率を p，裏になる確率を q とすると，$p+q=1$ である．

n 個の中から k 個を選び出す場合の数は

$$ {}_nC_k = \frac{n!}{(n-k)!k!} \equiv \begin{pmatrix} n \\ k \end{pmatrix} \tag{8} $$

である．その各々の場合について，確率は

$$ p^k q^{n-k} \tag{9} $$

となる．従って，n 個のうち k 個が表になる確率は $W(n, k)$ は

$$ W(n,k) = p^k q^{n-k} \frac{n!}{(n-k)!k!} \tag{10} $$

と表わされる．この式をあらゆる k に関して和をとると，

$$ \sum_{k=0}^{n} W(n,k) = \sum_{k=0}^{n} p^k q^{n-k} \frac{n!}{(n-k)!k!} = (p+q)^n = 1 \tag{11} $$

である．

また，k の平均（期待値）は

$$ m = \sum_{k=0}^{n} kW(n,k) = np \tag{12} $$

であり，分散は

$$ s^2 = \sum_{k=0}^{n} (k-m)^2 W(n,k) = npq \,. \tag{13} $$

で与えられる．

(b)正規分布（ガウス分布）

平均 m，分散 s^2 の正規分布は N（m，s^2）と表わされ，その確率密度関数は

$$ f(x) = \frac{1}{\sqrt{2p}\, s} \exp\left(-\frac{(x-m)^2}{2s^2} \right) \tag{14} $$

である．このとき，x の平均（期待値）は

$$ \bar{x} = m \tag{15} $$

であり，分散は

$$ \overline{(x-m)^2} = s^2 \tag{16} $$

である．

2.5 標本分布

母集団から取り出した大きさ n の標本（$x_1, x_2, \cdots, x_n$）からなる確率変数の組（$X_1, X_2, \cdots, X_n$）は標本変量と呼ばれる．標本変量あるいは標本変量の関数が従う分布は標本分布と呼ばれる．標本分布の例としては，c^2 分布，t 分布，F 分布などがある．

ここで，ひとつ例を挙げる．

大きさ n の標本について，確率変数 X_i が正規分布 $N(m, \sigma^2)$ に従う時，標本平均 $\overline{X}$，不偏分散 U^2 を用いた確率変数

$$T = \frac{\overline{X} - m}{U / \sqrt{n}} \tag{17}$$

は，自由度 n-1 の t 分布に従う．そのため，t 分布の $T=-t_\alpha$ から t_α までの積分が(1-α)になる様な t_α を t 分布表から得ると，標本平均と不偏分散を用いて，母平均 m の区間推定値を（1-α）の信頼係数で

$$[\overline{x} - \frac{u}{\sqrt{n}} t_a, \overline{x} + \frac{u}{\sqrt{n}} t_a]$$

と求めることができる．

2.6　誤差の分布

ここでは誤差の分布を調べる．今，誤差は次の性質をもつものと仮定しよう．

１．誤差は総数 n 個（n は大きな数）の素誤差（「Hagen の素誤差」と呼ばれる）の集積からなる．

２．素誤差は小さな絶対値 e をもち，正負両面に対して同じ生起確率をもつ．

この場合，「表に $+e$，裏に $-e$ と書いた n 個のコイン投げで，表と裏の数の和を求める問題」と等価になる．そのため，誤差は二項分布の $n \to \infty, e \to 0$ の極限であり，これは正規分布になる．以上より，誤差が従う分布は正規分布であることが分かる．

最後にこれまで出てきた重要な用語を図１にまとめて示しておく．

３．間接測定における誤差伝播

3.1　間接測定量の最大誤差

間接測定量 R は，次式のように直接測定量 X, Y, Z の関数であるとする：

$$R = f(X, Y, Z). \tag{18}$$

このとき，R を平均のまわりで Taylor 展開すると，平均からのずれ dR は

$$dR = \frac{\partial f}{\partial X} dX + \frac{\partial f}{\partial Y} dY + \frac{\partial f}{\partial Z} dZ \tag{19}$$

と表わされる．このとき，

$$| dR | \le \left|\frac{\partial f}{\partial X}\right|\left|dX\right| + \left|\frac{\partial f}{\partial Y}\right|\left|dY\right| + \left|\frac{\partial f}{\partial Z}\right|\left|dZ\right| \qquad (20)$$

が成り立つので，上の不等式の右辺から最大誤差を見積もることができる．

3.2 間接測定値の分散

誤差$d\,X, d\,Y, d\,Z$ は，平均ゼロ，分散$s_X{}^2$, $s_Y{}^2$, $s_Z{}^2$であり，互いに独立であるとすると，

$$s_R{}^2 = (\frac{\partial f}{\partial X})^2 s_X{}^2 + (\frac{\partial f}{\partial Y})^2 s_Y{}^2 + (\frac{\partial f}{\partial Z})^2 s_Z{}^2 \qquad (21)$$

の関係がある．これは直接測定量の誤差が間接測定量にどのように影響するかを示したもので，誤差の伝播法則と呼ばれる．

４．最小二乗法

4.1 最小二乗法

n 個の測定値の対（x_1 , y_1），…，（x_n, y_n）が得られた時，x と y との関数関係を推定することを考えよう．ここで，x の誤差は y の誤差に比べて小さいものと仮定する．

x と y の間には$y = f(x)$ なる関係式が存在し, 関数$f(x)$は m ($<n$) 個のパラメタ a_i(i=1, …，m) で規定されるものとする．このとき次の基準を満たすようにパラメタ a_i を推定する：

$$\sum_{i=1}^{n}\left(\frac{y_i - f(x_i)}{s_i}\right)^2 \to \min\ . \qquad (22)$$

ここで s_i は y_iの標準偏差を表わす．これは,「測定値 y_i と推定値 $f(x_i)$ との差（測定値 y_i の誤差）の二乗和を最小にする」というものであり，この基準にもとづく推定法が最小二乗法と呼ばれる．

4.2 最小二乗法と最尤法

最小二乗法は,「測定値 y_i は正規分布に従う」との条件の下で，最尤法と一致する．

今，y_i が平均m_i，分散$s_i{}^2$に従うとすると，y_i の確率密度関数は,

$$p(y_i) = \frac{1}{\sqrt{2p}\,s_i}\exp\left\{-\frac{(y_i - m_i)^2}{2s_i{}^2}\right\} \qquad (23)$$

で表わされる．y_1 ,… , y_nが互いに独立であると仮定すると，(y_1, …, y_n）の組み合わせが得られる確率密度関数（尤度関数とも呼ばれる）は

$$p(y_1, \cdots, y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2p}\, s_i} \exp\left\{ -\frac{(y_i - m_i)^2}{2s_i^{\ 2}} \right\}$$

$$= \left(\frac{1}{\sqrt{2p}} \right)^n \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \frac{(y_i - m_i)^2}{s_i^{\ 2}} \right\} \prod_{i=1}^{n} \frac{1}{s_i} \qquad (24)$$

で与えられる．ここで，$m_i = f(x_i)$ である．この尤度関数を最大にするように関数 $f(x)$ のパラメタ a_i を決定するのが最尤法であり，その条件は次の m 個の式である．

$$\frac{\partial p(y_1, \cdots, y_n)}{\partial a_j} = 0 \quad (j = 1, \cdots, m). \qquad (25)$$

次のように，尤度関数の対数の偏微分がゼロの条件でもよく，こちらの方が計算しやすい：

$$\begin{aligned} \frac{\partial \ln p(y_1, \cdots, y_n)}{\partial a_j} &= \frac{1}{p(y_1, \cdots, y_n)} \frac{\partial p(y_1, \cdots, y_n)}{\partial a_j} \\ &= \sum_{i=1}^{n} \frac{y_i - f(x_i)}{s_i^{\ 2}} \left(\frac{\partial f(x_i)}{\partial a_j} \right) = 0 \quad (j = 1, \cdots, m) \end{aligned} \qquad (26)$$

重みを $w_i = 1/s_i^{\ 2}$ とおくと

$$\sum_{i=1}^{n} w_i \{ y_i - f(x_i) \} \left(\frac{\partial f(x_i)}{\partial a_j} \right) = 0 \quad (j = 1, \cdots, m) \qquad (27)$$

とも書ける．結局，m 個の式から m 個のパラメタを求めることになるので原理的に解ける．この式は正規方程式と呼ばれる．

尤度関数 (24)を最大にすることは，指数関数部の引数の絶対値を最小にすることに対応するので，(22) に等しい．これより，測定値の分布が正規分布である時は，最小二乗法は最尤法と等しいことが分かる．

4.3 最小二乗法による直線のあてはめ

（1） データの分散が等しくないとき

ここでは，n 個の測定値の対 (x_1, y_1)，…，(x_n, y_n) に対して，直線 $f(x) = a_1 + a_2 x$ をあてはめることを考えよう．この場合，$\partial f(x) / \partial a_1 = 1$, $\partial f(x) / \partial a_2 = x$ であるから，

$1 / s_i^{\ 2} = w_i$ とおいて，(27)式は

$$\sum_{i=1}^{n} w_i y_i = a_1 \sum_{i=1}^{n} w_i + a_2 \sum_{i=1}^{n} w_i x_i \qquad (28)$$

$$\sum_{i=1}^{n} w_i x_i y_i = a_1 \sum_{i=1}^{n} w_i x_i + a_2 \sum_{i=1}^{n} w_i x_i^{\ 2} \tag{29}$$

となる．この正規方程式を解くと

$$a_1 = \frac{1}{\Delta}\begin{vmatrix} \sum w_i y_i & \sum w_i x_i \\ \sum w_i x_i y_i & \sum w_i x_i^{\ 2} \end{vmatrix} \tag{30}$$

$$a_2 = \frac{1}{\Delta}\begin{vmatrix} \sum w_i & \sum w_i y_i \\ \sum w_i x_i & \sum w_i x_i y_i \end{vmatrix} \tag{31}$$

が得られる．ただし

$$\Delta = \begin{vmatrix} \sum w_i & \sum w_i x_i \\ \sum w_i x_i & \sum w_i x_i^{\ 2} \end{vmatrix} \neq 0 \tag{32}$$

である．推定した係数 a_1, a_2 の分散は誤差の伝播法則より，

$$s_{a_1}^{\ 2} = \frac{1}{\Delta}\sum_{i=1}^{n} w_i x_i^{\ 2} \tag{33}$$

$$s_{a_2}^{\ 2} = \frac{1}{\Delta}\sum_{i=1}^{n} w_i \tag{34}$$

となる．

(2) データの分散が等しい時

全データが等しい分散$s^{\ 2}$をもつすると，直線の係数は

$$a_1 = \frac{1}{\Delta}\begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^{\ 2} \end{vmatrix} \tag{35}$$

$$a_2 = \frac{1}{\Delta}\begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix} \tag{36}$$

と推定される．ただし

$$\Delta = \begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^{\ 2} \end{vmatrix} \neq 0 \tag{37}$$

である．推定した係数 a_1, a_2 の分散は誤差の伝播法則より，

$$s_{a_1}^{\ 2} = \frac{s^{\ 2}}{\Delta}\sum_{i=1}^{n} x_i^{\ 2} \tag{38}$$

$$s_{a_2}^{2} = n\frac{s^2}{\Delta} \tag{39}$$

となる．ただし n が小さい時は，s^2 に代えて不偏分散

$$u^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - a_1 - a_2 x_i)^2 \tag{40}$$

を用いた方がよい．例えば，測定値の誤差を明示するには，測定値 y_i を中心として上下に長さ u の誤差棒をひくと良い．

参考文献

一瀬正巳，誤差論，培風館.

粟谷隆，データ解析－アナログとディジタル－，学会出版センター.

吉澤康和，新しい誤差論－実験データ解析法－，共立出版.

中川徹・小柳義夫，最小二乗法による実験データ解析，東大出版会.

鈴木義也・他，統計学概説，培風館.

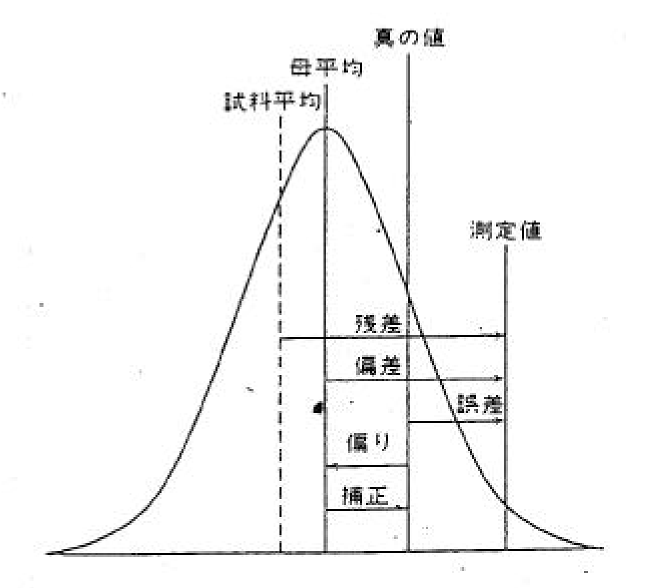

図 1，用語の定義（「物理学辞典」（培風館）より抜粋）