誤差と最小二乗法

1　誤差の種類

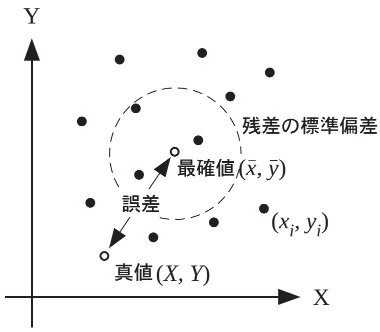
誤差のない計測というものはあり得ない．常に計測値には誤差がつきものである．誤差とは，真値から計測値を差し引いたものである．この誤差は，大きく次の３つに分類することが出来る．

1. **過失誤差**
2. **系統的誤差**
3. **偶然誤差**

過失誤差は，過誤とも呼ばれ，数値の読み取りミスや視準ミス等によって発生するものである．ミスによる誤差は，調整のしようがなく，再測せざるを得ない．系統的誤差は，誤差の発生機構が解っており，ある程度補正が可能な誤差である．温度による計測値のズレやレンズの歪みや大気の影響などによる誤差がこれに当たる．これらの誤差は，**キャリブレーション**によりある程度低減させることが出来る．キャリブレーションとは，誤差とその発生要因との関係を求め，計測値に対して補正することをいう．偶然誤差は，過失誤差や系統誤差以外の誤差であり，確率統計的な手段で誤差の調整が可能である．例えば，同じ対象を何度も測り，平均値を算出することで真値に近づけることが可能である．この平均値は，真値に近いものの真値とはいえないことから**最確値**と呼んでいる．そして，この最確値と計測値との差は，**残差**と呼ばれ，真値からの差の誤差とは区別している．

2　平均二乗誤差

下図は，誤差と残差の概念をグラフで示したものである．ある XY 座標を計測した結果を黒丸で示している．n 個の計測結果が $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ と得られた場合，これらの点を使って最確値 $(\bar{x}, \bar{y})$ を求めたとしても，図に示したように最確値は真値 (X, Y) とは異なる．この図には極端な例を示しているが，このように最確値が真値と大きく異なる場合がある．このように誤差を調整しても真値との差が非常に大きい場合は，系統的な誤差が含まれていると考えるべきであろう．



計測結果の評価を行う場合，真値が解らないことがほとんどである．そこで，計測値の分散や標準偏差を用いて評価することがしばしばある．X 座標における残差の標準偏差 σ_x は，以下の式で表すことが出来る．

$$\sigma_x = \sqrt{\sum_{i=1}^{n}(\bar{x} - x_i)/n} \tag{1}$$

しかし，この標準偏差は，残差の散らばりを示しており，先の図に示したように，誤差を評価していることにはならない．そこで計測結果を評価する場合は，別の方法や異なる機器，異なる位置から計測し，その結果を踏まえて評価した方が良い．例えば，高精度の機器を用いて計測し，それを真値と仮定するのも一つの方法である．例えば，高精度の機器で測られたデータを真値と仮定し，その真値との差を評価する．**平均二乗誤差**は，RMSE(Root Mean Square Error) と呼ばれ，その評価指標の一つである．式で表すと以下のようになる．

$$RMSE = \sqrt{\sum_{i=1}^{n}(X - x_i)/n} \tag{2}$$

計算式は，標準偏差の計算とほぼ同じで，最確値の値が真値となったにすぎない．計測結果の評価は，このような平均二乗誤差を用いることが好ましい．

3 最小二乗法

3.1 同一区間を複数回計測した場合

求める最確値を X とし，n 回計測した計測値を $(x_1, x_2, \cdots x_n)$ とすると，各計測値の残差は $X - x_i$ と表すことが出来る．**最小二乗法**は，この残差の二乗和が最小となる X を求めることである．残差の二乗和の関数 Φ は，以下の式で表すことが出来る．

$$\Phi = \sum_{i=1}^{n}(X - x_i)^2 \tag{3}$$

この誤差関数 Φ は，下向きに凸の形をしているので Φ の最小値を求めるには，X で微分し，それが 0 となる X を求めれば良い．したがって，以下の式を得る．

$$\frac{d\Phi}{dX} = 2\sum_{i=1}^{n}(X - x_i) = 0 \tag{4}$$

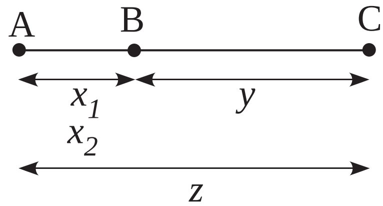
$$nX - \sum_{i=1}^{n} x_i = 0 \tag{5}$$

$$X = \sum_{i=1}^{n} x_i/n \tag{6}$$

つまり，平均値を求める式と同じ式が得られたことになる．

3.2　複数の区間を複数回計測した場合

下図のように，AB と BC の二区間長について，AB 間が x_1, x_2，BC 間が y，AC 間が z という計測結果が得られたとき，AB と BC の長さの最確値を最小二乗法を使ってどのように求めるか考える．



まず，AB と BC の長さの最確値を X, Y とおく．すると残差の二乗和の関数は，以下の式で表すことが出来る．

$$\Phi = (X - x_1)^2 + (X - x_2)^2 + (Y - y)^2 + (X + Y - z)^2 \tag{7}$$

先と同様に Φ の最小値を求めるが，このとき変数が X, Y の二個あるので偏微分により最小値を求める．すなわち，Φ を X と Y とでそれぞれ偏微分し，それが 0 となる X, Y を計算する．

$$\begin{cases} \dfrac{\partial \Phi}{\partial X} = 2(X - x_1) + 2(X - x_2) + 2(X + Y - z) & = 0 \\ \dfrac{\partial \Phi}{\partial Y} = 2(Y - y) + 2(X + Y - z) & = 0 \end{cases} \tag{8}$$

これを整理すると，以下の式を得る．

$$\begin{cases} 3X + Y & = x_1 + x_2 + z \\ X + 2Y & = y + z \end{cases} \tag{9}$$

この連立方程式を解けば，最確値 X, Y が求まる．