最小二乗法による線形モデルのパラメータ決定

中野武雄

2005 年 6 月 4 日

概 要

線形最小二乗法の定義と、そのモデルパラメータの決定法について述べる。得られたモデルパラメータの誤差の求め方についても記し、最後にいくつかのケースに関して具体的な表式を示す。

1 線形最小二乗法

1.1 測定と解析モデルの定義

ある実験において n 回の測定が行われた場合を考える。一つ一つの測定に際しては、l 個の測定条件が (数値によって) 指定されており、1 つの測定値が得られるものとする。i 番目の測定 $(1 < i \leq n)$ における条件パラメータの組を $x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(l)}$ のように表し、得られる測定値を y_i とする。測定値 y_i は、おのおの分散 σ_i を持って、真の値 ${y_i}^0$ の周りを正規分布しているものとする。また $x_i^{(j)}$ には誤差は含まれないものとする。

このとき、${y_i}^0$ と $x_i^{(j)}$ が、あらかじめ導入したモデル

$$
{y_i}^0 = f(x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(l)}; \theta_1, \theta_2, \ldots, \theta_m) \tag{1}
$$

によって誤差なく関係付けられるものと仮定したとき、このモデルのパラメータ $\theta_k \quad (k = 1 \sim m)$ を決定することを考える。パラメータ推定値はベクトルの形式に

$$
\hat{\boldsymbol{\theta}} \equiv \left\{ \hat{\theta_1}, \ldots, \hat{\theta_m} \right\} \tag{2}
$$

と書き、この推定値からモデルを用いて計算される値を

$$
\begin{aligned}
\hat{y_i} &= f(x_i^{(1)}, \ldots, x_i^{(l)}; \hat{\theta_1}, \ldots, \hat{\theta_m}) \quad (i = 1 \sim n) \\
&\equiv f_i(\hat{\theta_1}, \ldots, \hat{\theta_m}) \equiv f_i(\hat{\boldsymbol{\theta}})
\end{aligned} \tag{3}
$$

のように表わすものとする。

最小二乗法とは、このパラメータ推定法のひとつであり、具体的には以下の最小二乗条件を満たすように $\hat{\boldsymbol{\theta}}$ を定める方法である。

$$
S(\hat{\boldsymbol{\theta}}) \equiv \sum_{i=1}^{n} \frac{[y_i - f_i(\hat{\boldsymbol{\theta}})]^2}{{\sigma_i}^2} = \min \tag{4}
$$

ここで、測定値と計算値との差

$$
v_i \equiv y_i - \hat{y_i} = y_i - f_i(\hat{\boldsymbol{\theta}}) \quad (i = 1 \sim n) \tag{5}
$$

を残差と呼ぶ。また次の式で各測定値 y_i の重み w_i を定義する。

$$w_i \equiv {\sigma_0}^2/{\sigma_i}^2 \tag{6}$$

ここで σ_0 は任意に選べる比例定数であり、測定 i における分散 ${\sigma_i}^2$ が各測定についてそれぞれ既知である場合には、単純に 1 と置いて構わない。もし ${\sigma_i}^2$ が未知のときは、代表的と思われる値を設定しておき、式 6 に従って各測定値における分散の逆数の相対値を重みとすれば良い。

式 5 と式 6 を用いると、式 4 の最小二乗条件は、

$$S'(\boldsymbol{\theta}) \equiv {\sigma_0}^2 S(\boldsymbol{\theta}) = \sum_{i=1}^{n} v_i(\boldsymbol{\theta})^2 w_i = \min \tag{7}$$

と書き直すことができる。

最小二乗法がパラメータ $\hat{\boldsymbol{\theta}}$ の最尤推定法であるためには、以下の前提が必要となる。

1. 測定値の誤差 $\epsilon_i = y_i - {y_i}^0$ は不偏である。すなわち、$\langle \epsilon_i \rangle = 0$ である。
2. 測定値の誤差の分散は既知である。
3. 各測定は互いに独立であり、共分散は 0 である。
4. 誤差の分布系は正規分布である。

実際には条件 2 が満たされることは困難であるが、最小二乗法では測定値の誤差の相対値のみでモデルパラメータを決定することができる。ただしその場合、モデルの有意性やパラメータの決定誤差を求めることはできない。

1.2 線形モデル

線形モデルとは、式 1 において、理論式 f が**モデルのパラメータ** θ_k に対する一次結合

$$\begin{aligned}
&f(x_i^{(1)}, \ldots, x_i^{(l)}; \theta_1, \ldots, \theta_m) \\
&\quad \equiv f_i(\theta_1, \ldots, \theta_m) \\
&\quad = X_{i1}\theta_1 + \cdots + X_{im}\theta_m \\
&\quad = \sum_{j=1}^{m} X_{ij}\theta_j \quad (i = 1 \sim n)
\end{aligned} \tag{8}$$

で表されるものをいう。ここで各係数 X_{ij} は、条件パラメータ $x_i^{(1)}, \ldots, x_i^{(l)}$ を使って表される既知の定数であり、モデルパラメータ θ_i には依存しないものとする。つまり『線形』モデルとは、あくまでモデルのパラメータに対しての意味であることに注意する。[1]

この線形モデルを最小二乗条件である式 4 に代入すると、

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[y_i - \sum_{j=1}^{m} X_{ij}\theta_j \right]^2 / {\sigma_i}^2 \tag{9}$$

[1]例えばフーリエ展開で展開係数を計算するのも線形最小二乗問題の一種であると言える。

が得られる。式 9 が各モデルパラメータ θ_j に関して最小になるためには、各パラメータに対する偏微分が 0 になれば良い。

$$0 = \frac{\partial S(\boldsymbol{\theta})}{\partial \theta_j} = -2\sum_{i=1}^{n}\left[y_i - \sum_{j'=1}^{m} X_{ij'}x_{j'}\right]\cdot X_{ij}/{\sigma_i}^2 \quad (j = 1 \sim m) \tag{10}$$

これを未知のモデルパラメータ $\theta_{j'}$ に対して整理すると、

$$\sum_{j'=1}^{m}\left(\sum_{i=1}^{n} X_{ij'}X_{ij}/{\sigma_i}^2\right)\theta_{j'} = \sum_{i=1}^{n}\left(X_{ij}/{\sigma_i}^2\right)y_i \quad (j = 1 \sim m) \tag{11}$$

となる。これは $\theta_{j'}$ に関する以下のような連立一次方程式になる。

$$\left\{\begin{array}{ccccccccc}
A_{11}\theta_1 & + & A_{12}\theta_2 & + & \cdots & + & A_{1m}\theta_m & = & a_1 \\
A_{21}\theta_1 & + & A_{22}\theta_2 & + & \cdots & + & A_{2m}\theta_m & = & a_2 \\
 & \vdots & & & & & & & \\
A_{m1}\theta_1 & + & A_{m2}\theta_2 & + & \cdots & + & A_{mm}\theta_m & = & a_m
\end{array}\right. \tag{12}$$

ただしここで、

$$A_{jj'} \equiv \sum_{i=1}^{n} X_{ij'}X_{ij}/{\sigma_i}^2 \quad (j, j' = 1 \sim m) \tag{13}$$

$$a_j \equiv \sum_{i=1}^{n}\left(X_{ij}/{\sigma_i}^2\right)y_i \quad (j = 1 \sim m) \tag{14}$$

である。これらの量は、既知の値 (測定条件パラメータと測定値) で表されていることに注意。式 10 を正規方程式 (normal equation) と呼び、この方程式の解が最小二乗法で決定されるパラメータ推定値 $\hat{\boldsymbol{\theta}}$ となる。

1.3 行列表式

以上の表式を簡潔に表わすためには、行列形式を用いるのが便利である。測定値や測定パラメータなど、これまで出てきた組になっている量を、今後列ベクトル (縦ベクトル) で表わす。ベクトルは小文字の太字、行列は大文字の太字で表し、ベクトルや行列の転置には ˜ を用いる。

測定値のベクトルを $\mathbf{y}$ とし、その真値を $\mathbf{y_0}$ とおいて、誤差ベクトル $\boldsymbol{\epsilon}$ を次のように定義する。これらのベクトルの次元は測定回数である n となる。

$$\boldsymbol{\epsilon} \equiv \boldsymbol{y} - \boldsymbol{y_0} \tag{15}$$

さらに、この n 個の測定に関する誤差行列 $\mathbf{\Sigma}$ を

$$\mathbf{\Sigma} \equiv \langle \boldsymbol{\epsilon}\tilde{\boldsymbol{\epsilon}} \rangle \tag{16}$$

と定義する。すなわち $\mathbf{\Sigma}_{ii'} \equiv \langle \epsilon_i \epsilon_{i'} \rangle \quad (i, i' = 1 \sim n)$ である。この誤差行列の対角成分を分散、非対角成分を共分散と呼ぶ。

最小二乗法における、各測定値の誤差に関する要請から、誤差ベクトルと誤差行列は次の性質を持っていることがわかる。

$$\langle \boldsymbol{\epsilon} \rangle = \mathbf{0} \tag{17}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} {\sigma_1}^2 & & & 0 \\ & {\sigma_2}^2 & & \\ & & \ddots & \\ 0 & & & {\sigma_n}^2 \end{bmatrix} \tag{18}$$

さらに、式 6 で定義されていた、重み w_i を対角成分とする重み行列 $\mathbf{W}$ を、誤差行列 $\boldsymbol{\Sigma}$ の逆行列を用いて

$$\boldsymbol{W} \equiv {\sigma_0}^2 \boldsymbol{\Sigma}^{-1} \tag{19}$$

と定義する。測定値間に相関があるとき (共分散があり、$\boldsymbol{\Sigma}$ が非対角行列となるとき) も、式 19 はそのまま成立し、この後の表式を用いれば良い。

パラメータの列ベクトル (m 元) を $\boldsymbol{\theta}$、モデル理論式の列ベクトル (n 次元) を $\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\theta}) \equiv \boldsymbol{f}(\boldsymbol{\theta})$ と書くことにする。すると最小二乗問題は、誤差 $\boldsymbol{\sigma}$ (より一般的には誤差行列 $\boldsymbol{\Sigma}$) のもとで、

$$\boldsymbol{y} \cong \boldsymbol{f}(\boldsymbol{\theta}) \tag{20}$$

に対する最適パラメータ $\hat{\boldsymbol{\theta}}$ を求めることであると表現できる。

$\hat{\boldsymbol{\theta}}$ の解は、式 4 を一般化した以下の最小二乗条件

$$S(\boldsymbol{\theta}) \equiv \tilde{\boldsymbol{v}}(\boldsymbol{\theta}) \boldsymbol{\Sigma}^{-1} \boldsymbol{v}(\boldsymbol{\theta}) = {\sigma_0}^{-2} \tilde{\boldsymbol{v}}(\boldsymbol{\theta}) \boldsymbol{W} \boldsymbol{v}(\boldsymbol{\theta}) = \min \tag{21}$$

を満たすものとして定義されることになる。

ここから線形モデルの話になる。式 8 で表される線形モデルは、行列の表現では

$$\boldsymbol{f}(\boldsymbol{\theta}) \equiv \boldsymbol{X}\boldsymbol{\theta} \tag{22}$$

と表される。ここで $\boldsymbol{X}$ は式 8 の係数 X_{ij} を要素とする $n \times m$ 次の長方行列で、ヤコビアン行列と呼ばれる。一般に m 個のパラメータを決定するにはそれ以上の回数の測定が必要であるから、$n > m$ であることに注意する。

より一般的には、ヤコビアン行列の要素 X_{ij} は、モデル $\boldsymbol{f}_i(\boldsymbol{\theta})$ のパラメータ θ_i に対する偏微分係数であると定義されており、

$$X_{ij} \equiv \frac{\partial f_i(\boldsymbol{\theta})}{\partial \theta_j} \quad (i = 1 \sim n, j = 1 \sim m) \tag{23}$$

である。これは線形モデルの場合には、測定の条件パラメータととモデルとから既知となる定数である。

最小二乗条件 (式 9) を微分して得られる正規方程式は、式 10 を行列形式で表し、

$$\left(\tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{X}\right)\boldsymbol{\theta} = \tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{y} \tag{24}$$

と書ける。あるいは、式 12 の形式で書けば、

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a} \tag{25}$$

となる。ただしここで、

$$\boldsymbol{A} \equiv \tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{X} \tag{26}$$

$$\boldsymbol{a} \equiv \tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{y} \tag{27}$$

である。

したがって最小二乗解 $\hat{\boldsymbol{\theta}}$ は、

$$\hat{\boldsymbol{\theta}} = \boldsymbol{A}^{-1}\boldsymbol{a} = (\tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{X})^{-1}\tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{y} \tag{28}$$

である。ここで、$m \times n$ の行列 $\boldsymbol{C}$ を

$$\boldsymbol{C} \equiv \boldsymbol{A}^{-1}\tilde{\boldsymbol{X}}\boldsymbol{W} = \left(\tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{X}\right)^{-1}\tilde{\boldsymbol{X}}\boldsymbol{W} \tag{29}$$

のように定義すれば、結局式 28 は

$$\hat{\boldsymbol{\theta}} = \boldsymbol{C}\boldsymbol{y} \tag{30}$$

となり、結局線形最小二乗法の解 $\hat{\boldsymbol{\theta}}$ は、測定値 y の線形結合で表されることがわかる。

2 モデル有意性の検定

線形最小二乗解を求める際に利用したモデルが妥当であるかどうかについては、各測定値の分散がわかっている場合は χ^2 検定によって検定することができる。(To be written: 参考文献の §3.5 を参照)

3 決定パラメータの誤差

前提として、式 15 で定義される誤差ベクトル $\boldsymbol{\epsilon}$ は既知とする。また測定値の真値 $\boldsymbol{y}^0$ と、パラメータ推定値の真値 $\boldsymbol{\theta}^0$ は $\boldsymbol{y}^0 = \boldsymbol{X}\boldsymbol{\theta}^0$ という関係にあるとする。このとき

$$\boldsymbol{C}\boldsymbol{y}^0 = \boldsymbol{C}\boldsymbol{X}\boldsymbol{\theta}^0 = (\tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{X})^{-1}\tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta}^0 = \boldsymbol{\theta}^0 \tag{31}$$

だから、式 30 に $\boldsymbol{y}^0$ を用いれば $\boldsymbol{\theta}^0$ が得られることがわかる。

このときパラメータ推定値 $\hat{\boldsymbol{\theta}}$ の真値に対するずれ $\delta\boldsymbol{\theta}$ は、

$$\delta\boldsymbol{\theta} \equiv \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = \boldsymbol{C}\boldsymbol{y} - \boldsymbol{C}\boldsymbol{y}^0 = \boldsymbol{C}\boldsymbol{\epsilon} \tag{32}$$

となる。ちなみに最小二乗法の適用条件 (式 17) より $\langle\delta\hat{\boldsymbol{\theta}}\rangle = 0$ となるので、パラメータ推定値の期待値も 0 となり、これが不偏推定量になっていることも理解できる。

いまこのパラメータ推定値の分散・共分散からなる誤差行列 $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ を次のように定義する。

$$(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}})_{jj} \equiv {\sigma_{\hat{\boldsymbol{\theta}}_j}}^2 = \langle\langle\hat{\theta}_j - {\theta_j}^0\rangle^2\rangle \quad (j = 1\ldots m) \tag{33}$$

$$(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}})_{jj'} \equiv \langle(\hat{\theta}_j - {x_j}^0)(\hat{\theta}_{j'} - {x_{j'}}^0)\rangle \quad (j \neq j', j, j' = 1\ldots m) \tag{34}$$

これらをまとめて表し、32 を代入すると、

$$\begin{aligned}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} &\equiv \langle(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\widetilde{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)}\rangle \\ &= \langle\boldsymbol{C}\boldsymbol{\epsilon}\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{C}}\rangle = \boldsymbol{C}\langle\boldsymbol{\epsilon}\tilde{\boldsymbol{\epsilon}}\rangle\tilde{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{\Sigma}\tilde{\boldsymbol{C}}\end{aligned} \tag{35}$$

が得られる。この関係は、測定値からパラメータへの誤差伝播そのものといえる。

式 35 に 19 と 29 を代入して整理すると ($\tilde{\boldsymbol{W}} = \boldsymbol{W}$ に注意)、

$$\begin{aligned}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} &\equiv \left(\tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{X}\right)^{-1}\tilde{\boldsymbol{X}}\boldsymbol{W}\left({\sigma_0}^2\boldsymbol{W}^{-1}\right)\boldsymbol{W}\boldsymbol{X}\left(\tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{X}\right)^{-1} \\ &= {\sigma_0}^2\left(\tilde{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{X}\right)^{-1} = \left(\tilde{\boldsymbol{X}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1} \end{aligned} \tag{36}$$

と書ける。すなわち、パラメータ推定値の誤差は、測定点の取り方とモデル ($\boldsymbol{X}$) 及び測定精度 ($\boldsymbol{\Sigma}$) から決まり、測定値そのもの ($\boldsymbol{y}$) には依存しない。

3.1 測定誤差の分散がすべて等しい場合

最小二乗法が適用できるためには 1.1 で与えた 4 つの条件が必要であったが、さらに特殊な場合として、

5. 各測定値の分散がすべて等しい ($\sigma_i = \mathrm{const} \equiv \sigma$)。

ことを期待できる場合も多い。この場合、測定数がある程度大きければ、最小二乗解 $\hat{\boldsymbol{\theta}}$ を使って

$$\sigma^2 = \left\langle \left(y_i - \sum_{j=1}^{m} X_{ij}\hat{\theta}_j \right)^2 \right\rangle \tag{37}$$

とみなせる。このとき誤差行列は $\Sigma = \sigma^2 I$ (I は単位行列) となり、これを式 35 に代入すればパラメータの推定誤差が得られる。

4 ケーススタディ

4.1 $y = \theta x$

このときモデルパラメータはひとつだけだから $n = 1$ で $\boldsymbol{\theta} = \theta$。いま m 回の測定を行って $(x_i, y_i)\quad(j = 1 \ldots m)$ というデータセットが得られたとすると、ヤコビアン行列 $\boldsymbol{X}$ は

$$\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \tag{38}$$

となる。ここで式 19 のような重み行列を用い、式 30 に代入すれば θ が得られる。

式 18 のように、誤差行列が対角成分しか持たない (つまり測定間の共分散が 0) ならば、式 30 は

$$\theta = \frac{\sum_{j=1}^{m}\left(\frac{x_j y_j}{{\sigma_j}^2}\right)}{\sum_{j=1}^{m}\left(\frac{{x_j}^2}{{\sigma_j}^2}\right)} \tag{39}$$

となる。またこのパラメータ推定値の誤差は、式 32 を用いて

$$\sigma_\theta = \frac{1}{\sum_{j=1}^{m}\left(\frac{{x_j}^2}{{\sigma_j}^2}\right)} \tag{40}$$

となる。

4.2 $y = \theta_1 + \theta_2\ \mathbf{x}$

同様に考え、このときのヤコビアン行列は

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \tag{41}$$

である。

このケースを GNU octave (http://www.octave.org) によって実装した例を示す。以下のスクリプトでは、データセットを標準入力から与えると、$\boldsymbol{\theta}$ ベクトルとその誤差行列とが出力される。Tx を変更すれば、他の場合に適用するのも容易だろう。なお、誤差行列の成分は分散であり、標準偏差ではない点に注意すること。

なお octave には gls, ols という自前の最小二乗フィット関数もあるので、そちらを利用するのも良いだろう。ただしこれらの関数ではパラメータの誤差は評価されない。

```
#!/usr/bin/octave -qf
line=fgets(stdin);
while ( !feof(stdin) )
     [xi, yi, dyi] = sscanf(line, "%f %f %f", "C");
     x(i)  = xi;
     y(i)  = yi;
     dy(i) = dyi;
     i++;
     line=fgets(stdin);
endwhile

[rx, cx] = size(x);
Tx = [ones(rx,1), x];

EM = diag(dy.^2);
WM = inv(EM);
C = inv(Tx' * WM * Tx) * Tx' * WM;

Theta = C * y
vMat = inv(Tx' * WM * Tx)
```

4.3 片対数・両対数プロットの場合

測定値 $\tilde{y_i}$ に測定誤差 $\delta\tilde{y_i}$ がある場合、$y_i \equiv \log_{10}\tilde{y_i}$ の誤差 σ_i は、誤差伝播を用いて

$$\begin{aligned} \sigma_i &= \frac{\partial y_i}{\partial\tilde{y_i}}\delta\tilde{y_i} \\ &= \frac{1}{\log_e 10}\frac{\delta\tilde{y_i}}{\tilde{y_i}} \end{aligned} \tag{42}$$

となる。この y_i と σ_i を用いて、ここまで述べてきた計算を行えばよろしい。絶対誤差 $(\delta\tilde{y_i})$ が一定ならば、小さな測定値 $\tilde{y_i}$ では y_i が小さくなるために係数決定に寄与する重みは小さくなる。一方相対誤差 $(\delta\tilde{y_i}/y_i)$ が一定ならば、それぞれの測定点は同じ重みを持つことがわかるであろう。

参考文献

[1] 中川 徹, 小柳義夫「最小二乗法による実験データ解析」東京大学出版会 UP 応用数学選書 7 (1982)