２．最小２乗法による回帰曲面

応答曲面法は関数形状に関係無く，近似関数を求めて最適化に使用する方法であるが，線形関数または線形化変換可能な関数は最小２乗法を用いることで容易にその関数の係数を統計的に推定することが可能であり，またその近似関数の統計的評価が可能である．本章では，最小２乗法を用いた回帰曲面の作成手法に関する統計的手法を述べる．

２．１ 最小２乗法（Least Squares Method）

簡単な例として変数が１個（ｘ）で応答ｙを直線で近似する場合を考える．

$$y = \beta_0 + \beta_1 x \tag{2.1.1}$$

ここで，βは未知係数である．

実験を行い，ｎ個のデータの組（x_i，y_i）（i=1,…,n）を取得した．これを式(2.1.1)の直線で近似する．

ｎ個のデータを代入して，ｎ個の式が得られる．

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_n + \varepsilon_n \end{aligned} \tag{2.1.2}$$

ここで，ε_iは誤差である．誤差２乗和 SSE（Square Sum of Errors）は次式で表される．

$$SSE = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{2.1.3}$$

SSE はβの２次関数であり，βには区間の制限が無いので，SSE をβの関数とみなせば，SSE を最小にするβは，∂SSE/∂β＝０が必要十分である．

$$\frac{\partial SSE}{\partial \beta_0} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \tag{2.1.4}$$

$$\frac{\partial SSE}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \tag{2.1.5}$$

式(2.1.4)，(2.1.5)から，次式の連立方程式が得られる．

$$\begin{aligned} \sum_{i=1}^{n} y_i &= n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i &= \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 \end{aligned} \tag{2.1.6}$$

式（2.1.6）からβが求められる．

$$\begin{aligned} \beta_1 &= \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} \\ \beta_0 &= \frac{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} \end{aligned} \tag{2.1.7}$$

誤差の２乗和の最小化と距離の最小化，最大誤差の最小化のそれぞれの違いについてｙ＝βという簡単な直線を例にとって考えてみる．

例として図 2.1.1 に示す３点（0,0），(1,1)，(2,0)を通る y=βを考える．式(2.1.6)から，β_1=0 とすればｙ＝βの場合が得られ，それは単に平均値となる．

$$y = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{3} \qquad (2.1.8)$$

y=βとの誤差（距離）の和 L は次式で表される．

$$L = 2\mathrm{b} + (1 - \mathrm{b}) = 1 + \mathrm{b} \qquad 0 \le \mathrm{b} \le 1 \qquad (2.1.9)$$

式（2.1.9）を最小化するβは明らかにβ_L＝0 となる．

最大誤差は次式で表される．

$$M = \max(\mathrm{b}, 1 - \mathrm{b}) \qquad (2.1.10)$$

最大誤差を最小にするには，β_M=0.5 とすればよい．

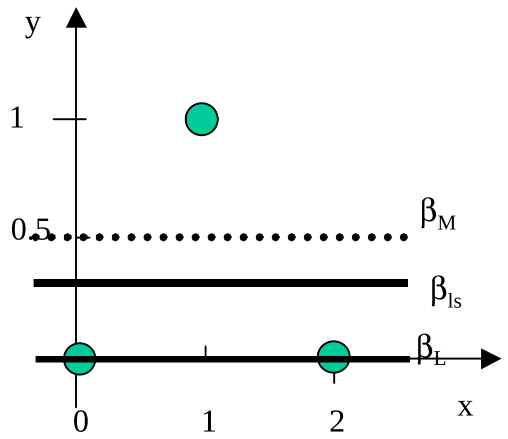

図 2.1.1　y=β の近似（誤差２乗和最小，誤差和最小
最大誤差最小）

図 2.1.1 から明らかに，誤差の和の最小化は多数の類似データに大きく影響を受け，最大誤差最小化は少数の値の異質データの影響を強く受ける．誤差２乗和の最小化はちょうどデータの平均となっている．このことから，誤差２乗和の最小化が多く用いられている．

２．２ マトリックス表現[1][2]

応答関数として２次多項式（Quadratic Polynomials）を採用した場合，応答曲面は次式となる．

$$y = \mathrm{b}_0 + \sum_{i=1}^{k} \mathrm{b}_i x_i + \sum_{i=1}^{k} \mathrm{b}_{ii} x_i^2 + \sum_{i<j} \mathrm{b}_{ij} x_i x_j \qquad (2.2.1)$$

簡単化のため，変数が２つの場合を例に取ると，式(2.2.1)は次式となる．

$$y = \mathrm{b}_0 + \mathrm{b}_1 x_x + \mathrm{b}_2 x_2 + \mathrm{b}_3 x_1^2 + \mathrm{b}_4 x_2^2 + \mathrm{b}_5 x_1 x_2 \qquad (2.2.2)$$

式（2.2.2）において，$x_1^2=x_3, x_2^2=x_4, x_1x_2=x_5$ と変数を置き換えることで多変数の２次多項式は多変数１次式に変換できる．全く同様にして，３次多項式，４次多項式など

の高次多項式も線形化可能である．

回帰式の係数βの推定に用いる実験点の組の総数をn，変数（変換後の変数）の数をkとすると線形回帰モデルは行列表示で次式になる．

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (2.2.3)$$

$$\mathrm{y} = \begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{Bmatrix} \qquad \mathrm{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$\beta = \begin{Bmatrix} \mathrm{b}_0 \\ \mathrm{b}_1 \\ \vdots \\ \mathrm{b}_k \end{Bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{Bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{\mathrm{n}} \end{Bmatrix}$$

誤差２乗和をＬとすると次式が得られる．

$$L = \boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathbf{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^{\mathbf{T}}\mathbf{y} - 2\boldsymbol{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\boldsymbol{\beta} \qquad (2.2.4)$$

前節と同様に，誤差２乗和 L を最小にすることから，係数βの不偏推定量bに関しての最小２乗の正規方程式(least squares normal equations)が求められる．

$$\left.\frac{\partial L}{\partial \boldsymbol{\beta}}\right|_{b} = -2\mathbf{X}^{\mathbf{T}}\mathbf{y} + 2\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{b} = 0 \qquad (2.2.5)$$

式(2.2.5)から，未知係数βの最小２乗推定量b(Least squares estimations of β)が次式で得られる．

$$\mathrm{b} = \left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}\mathbf{X}^{T}\mathbf{y} \qquad (2.2.6)$$

式（2.2.6）で得られるbの分散共分散行列(variance covariance matrix) V(b)＝cov($\mathrm{b_i}$,$\mathrm{b_j}$)は$\mathbf{C} = \mathbf{X}^{T}\mathbf{X}$ とおくと次式で表される．

$$\begin{aligned} \mathrm{V(b)} &= \mathrm{cov}(\mathrm{b_i}, \mathrm{b_j}) = \mathrm{V}(\mathbf{Cy}) = \mathbf{C}^{\mathbf{T}}\sigma^2\mathbf{C} = \mathbf{C}^{\mathbf{T}}\mathbf{C}\,\sigma^2 \\ &= (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\sigma^2 = \sigma^2(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1} \end{aligned} \qquad (2.2.7)$$

ここで，σ^2は応答yの誤差分散である．σ^2の最尤推定値$\hat{\sigma}^2$ はnが十分大きい場合，残差平方和 SSE から次式で計算できる．

$$\hat{\sigma}^2 = \frac{SSE}{n-p} \qquad (2.2.8)$$

ただし，SSE は残差２乗和で次式で定義される．

$$\mathrm{SSE} = \mathrm{y^T y} - \mathrm{b^T X^T y} \qquad (2.2.9)$$

２．３　回帰モデルの検定と最適回帰モデル

回帰したモデルの有効性の検定を統計的に行うことが可能である．今，回帰モデルが無意味であるという仮説を立てる．

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \tag{2.3.1}$$

式(2.3.1)の仮説の棄却は少なくとも回帰モデルの一つの係数はゼロでないこと，つまり回帰モデルの変数が少なくとも一つは応答変化に寄与していることを意味する．

この仮説検定は次式で定義される F 値から F 検定を用いて行われる．

$$F_0 = \frac{\dfrac{SSR}{k}}{\dfrac{SSE}{n-k-1}} \tag{2.3.2}$$

ここで，SSR は回帰２乗和（Regression Sum of Squares）であり，次式で定義される．

$$SSR = \mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{2.3.3}$$

もしも，F_0 の値が F 分布の $F_{\alpha,k,n-k-1}$ を超える場合には，H_0 の仮説が棄却される（F 分布については付録参照）．つまり，回帰が意味があることを意味する．

この検定は回帰が意味があることだけを告げており，回帰モデルの適切さについては何も答えていない．一般には機械工学分野において応答曲面を適用する場合には既に変数はほぼ適切に明らかになっている場合が多く，前述の検定は必要ない場合が多い．もちろん農業や品質工学分野，社会学などで，変数が明らかでない場合には必要不可欠な検定であることは注意しなければならない．

回帰モデルが適切かどうかの判定は一般には決定係数（coefficient of multiple determination）を用いる．決定係数 R^2 は次式で定義される．

$$R^2 = \frac{SSR}{Syy} = 1 - \frac{SSE}{S_{yy}} \tag{2.3.4}$$

ここで，Syy は応答 y の平均値まわりの変動（total sum of squares）で，次式で定義される．

$$S_{yy} = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{2.3.5}$$

決定係数とは，回帰モデルの適合性を示す指標となっている．例えば，回帰式が完全に応答に一致していれば，式(2.3.3)の SSR と式(2.3.5)の Syy は完全に同一となり，決定係数 R^2 は１となる．Syy=SSR+SSE の関係があるので，残差があれば必ず決定係数は１より小さくなる．つまり，０－１の間の数を決定係数はとる．ただし，変数を多くすれば残差は減少するので，決定係数の値が高いものが良い回帰モデルと断定はできない．回帰モデルの良否の比較には，単位自由度あたりの残差を比較する必要があり，一般には自由度調整済み決定係数 R^2_{ad} が使用される．

$$R_{ad}^2 = 1 - \frac{SSE/(n-k-1)}{S_{yy}/(n-1)} \tag{2.3.6}$$

自由度調整済み決定係数は不要な変数が追加されたときには低下する．

回帰モデル中の各係数は回帰係数のｔ検定でその有意性判定をすることができる．

回帰式のｊ番目の係数β_j=０つまりｊ番目の変数が回帰式に寄与していないという仮説をたて，検定を実施する．この場合のｔ値は次式で表される．

$$t_0 = \frac{b_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \tag{2.3.7}$$

ここで C_{jj} は式（2.2.7）の正方行列（X^TX）$^{-1}$ のｊｊ成分である．この仮説係数β_j=０はｔ$_0$の絶対値がｔ$_{\alpha/2,\ n-k-1}$より大きいときに否決される（変数が有効と判定される）．ｔ$_{\alpha/2,\ n-k-1}$は自由度 n-k-1 信頼係数 1-αのｔ分布の値である．ｔ分布の値は付録にあるので参照されたい．

例えば，実験点数 16 個（n=16），未知係数の総数 6（k=6）の場合，自由度は 16-6-1＝9 であるので，95%検定（α=0.95）の場合，ｔ値は付録表から 2.26 となる．

各回帰モデル中の係数のｔ$_0$は一般的には回帰モデルに依存して異なるため，棄却された変数を削除したり，別の変数を追加するごとに各変数のｔ$_0$は変化する．このため，ｔ$_0$は削除や追加によって回帰モデル変更するごとに計算しなおす必要がある．最適な回帰モデルを求めるには，変数を増減してゆく増減法（step wise）の他に，変数を増加させる方法や減少させていく方法がある．以下に簡単な例として減少法を示す．

［例］下表２で与えられる２変数 x1，x2 の応答ｙの応答曲面を２次多項式で求める．

表２.3.1　実験結果の例

No.	x1	X2	y
1	-1	-1	16.2
2	-1	1	7.5
3	-1	0	10.2
4	-1	0.5	13.2
5	1	-1	-4.2
6	1	1	15.6
7	-1	0	10.2
8	1	0	5.2
9	0	0	4.8
10	0	1	6.0
11	0	0.5	5.0
12	0.5	0.5	9.5

２変数の多項式は次式となる．

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 \tag{2.3.8}$$

式(2.2.6)から，それぞれの係数を計算し，式(2.3.7)から各項のｔ$_0$値を計算すると，各係数は表 2.3.2 になる．ここで，自由度調整済み決定係数 =0.8655 である．

自由度は 12-5-1＝６であるのでｔの閾値は 2.45 となる．絶対値を比較すると b4 が不要と判断される．そこで，　の項を削除した回帰モデルを用いて再び回帰係数の推定値を計算する．自由度調整済み決定係数 =0.8845 で各係数の計算結果を下表 2.3.2 に示す．

表2.3.2　最小２乗法で推定された係数

	推定値	t 値
b0	4.37	3.42
b1	-2.97	-3.96
b2	3.04	3.2
b3	4.50	3.09
b4	-0.12	1.35
b5	6.86	6.89

自由度は 12-5-1＝６であるので t_0 の閾値は 2.45 となる．絶対値を比較すると b4 が不要と判断される．そこで，　x_2^2 の項を削除した回帰モデルを用いて再び回帰係数の推定値を計算する．自由度調整済み決定係数 =0.8845 で各係数の計算結果を下表 2.3.3 に示す．

表2.3.3　修正された係数

	推定値	t 値
b0	4.33	3.89
b1	-2.98	-4.35
b2	3.03	3.47
b3	4.48	3.39
b5	6.86	7.44

表 2.3.3 から，b4 削除によって全体の t 値が変化していることがわかる．自由度調整済み決定係数は 0.8655 から 0.8845 に上昇している．自由度は 12-4-1＝７であるので，$t_{0.025,7}$＝2.36 となり，全ての係数は採用され，回帰式が決定される．

減少法は簡単であるが，多くの項が削減されてしまう場合，削減後に一度削減した項を追加すると回帰が向上する場合もあり，真の最適回帰モデルを提示する保証はない．比較的高次の項が無く，さらに削除する項数が少ない場合には有効である．一般的には少ない項数から開始して増減させる増減法が多く用いられる．

各係数の推定値 b の信頼区間（confidence interval）は t 分布を用いることで計算される．100(1-α)%の信頼区間は次式で表される．

$$b_j - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \le \beta_j \le b_j + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \qquad (2.3.9)$$

ある点 x_0 における応答 y(x_0)の平均値 μ(x_0)の信頼区間は次式で表される．

$$\hat{y}(x_0) - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 x_0^T (X^T X)^T x_0} \le \mu(x_0) \le \hat{y}(x_0) + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 x_0^T (X^T X)^T x_0} \qquad (2.3.10)$$

ここで， $\hat{y}(x_0)$ は応答曲面による点 x_0 の応答である．

また，点 x_0 における応答 y(x_0)の予測値 $y_0(x_0)$の信頼区間は次式で表される．

$$\hat{y}(x_0) - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 (1 + x_0^T (X^T X)^T x_0)} \le y_0(x_0) \le \hat{y}(x_0) + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 (1 + x_0^T (X^T X)^T x_0)} \qquad (2.3.11)$$

式(2.3.10)と式(2.3.11)の差異は推定する値が平均値か予測値かの違いに起因している．一般に，ある点での予測値は信頼区間が広くなってしまう．

同一の応答に対して，２つの実験値の集団から２つの応答曲面が得られ，それらに差異があるかどうかの検定も統計的に行うことができる．２つの集団の実験個数をそ

れぞれ n_1，n_1 とし，変数の数をｋ個，p=k+1 とする．

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$$

$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2$$

仮説は $\beta_1 = \beta_2$ となる．この仮説の検定に使用する統計量 W は次式で表される．

$$W = \left(\frac{SSE_0}{SSE_{12}} - 1\right)\left(\frac{n-2p}{p}\right) \qquad (2.3.12)$$

ここで，SSE_0 は n_1, n_2 をすべて用いた時の応答曲面の誤差平方和であり，また $n=n_1+n_2$ である．SSE_{12} は次式で表される．

$$SSE_{12} = (n-p)(SSE_1 + SSE_2) \qquad (2.3.13)$$

ここで，SSE_1，SSE_2 はそれぞれ応答曲面１および２の誤差平方和である．W が $F_{\alpha,p,n-2p}$ の値を超える場合には仮説が棄却され，２つの応答曲面は異なるものと判断される．

２．４　変数変換[1]

これまでは線形関数かまたは多項式で線形関数に容易に置換可能な関数を取り扱ってきた．実際にはこれらの関数ばかりではなく，非線形関数が必要な場合も多い．非線形関数を用いる場合には，変数変換を行うことで，線形関数に変換が可能であるものがある．例として表 2.4.1 にこれらを列挙する．

表 2.4.1　変数変換により線形化可能な関数の例[1]

関　数	変数変換	線形化後関数
$y = a + b\log x$	$x' = \log x$	$y = a + bx'$
$y = ax^b$	$y' = \log y$，$x' = \log x$	$y' = \log a + bx'$
$y = \dfrac{x}{ax+b}$	$y' = \dfrac{1}{y}$，$x' = \dfrac{1}{x}$	$y' = a + bx'$
$y = ae^{bx}$	$y' = \log y$	$y' = \log a + bx$
$y = \dfrac{e^{a+bx}}{1+e^{a+bx}}$	$y' = \log\left(\dfrac{y}{1-y}\right)$	$y' = a + bx$

ここにあげた以外にもさまざまな変換が考えられる．このような変数変換で線形化可能な関数は先に延べた正規方程式を用いた最小２乗法で係数を決定できる．

参考文献

[1]佐和隆光，「回帰分析」，朝倉書店(1979)
[2]田島稔，小牧和雄，「最小２乗法の理論とその応用改訂版」，東洋書店,(1996)