

1 線形モデルの最小2乗解

Y, X, β をそれぞれ $n \times p, n \times q, q \times p$ の行列として、

$$Y = X\beta + \varepsilon = \hat{Y} + \varepsilon \quad (1)$$

$$Y \approx X\beta = \hat{Y} \quad (2)$$

というモデルがあった場合、最小2乗基準

$$\text{RSS} = \sum_{i=1}^n \sum_{j=1}^p (y_{ij} - \hat{y}_{ij})^2 = \text{tr}((Y - \hat{Y})'(Y - \hat{Y})) \quad (3)$$

$$= \text{tr}[(Y - X\beta)'(Y - X\beta)] \quad (4)$$

$$= \text{tr}(Y'Y) - \text{tr}(\beta'X'Y) + \text{tr}(Y'X\beta) + \text{tr}(\beta'X'X\beta) \quad (5)$$

$$= \text{tr}(Y'Y) - 2\text{tr}(Y'X\beta) + \text{tr}(\beta'X'X\beta) \quad (6)$$

を最小とする β , すなわち $\hat{\beta}$ は、上式の β に関する偏微分を 0 とおいた方程式

$$2X'X\beta - 2X'Y = \mathbf{0} \quad (7)$$

を解くことにより、

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (8)$$

$$\hat{\beta}' = Y'X(X'X)^{-1} \quad (9)$$

で与えられる。

また、モデルがフルランクでない場合、すなわち、 $\text{rank}(X) < q$ の場合、 β の最小2乗解は一意には定まらないが、任意のベクトル c を用いて

$$\hat{\beta} = (X'X)^{-}X'Y + (I - (X'X)^{-}X'X)c \quad (10)$$

と書ける。

なお、いずれの場合にも、 Y の予測値 \hat{Y} は

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-}X'Y = P_X Y \quad (11)$$

となり、一意に定まる。また、残差行列は

$$Y - \hat{Y} = Q_X Y \quad (12)$$

と書ける。

1.1 重み付き最小2乗法

$$\text{RSS} = \sum_{j=1}^p \sum_{i=1}^n w_{ii}(y_{ij} - \hat{y}_{ij})^2 \quad (13)$$

$$= \text{tr}((Y - X\beta)'W(Y - X\beta)) \quad (14)$$

を最小とする β は

$$\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y} \quad (15)$$

として与えられる。

また、

$$\text{RSS} = \sum_{j=1}^p v_j \sum_{i=1}^n (y_{ij} - \hat{y}_{ij})^2 \quad (16)$$

$$= \text{tr}((\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)\mathbf{V}) \quad (17)$$

を最小とする β は

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (18)$$

として与えられる。

1.2 残差の2乗和の分解 1

$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ とした場合、RSS は以下のように書き表される。

$$\text{RSS} = \text{tr}[(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)] \quad (19)$$

$$= \text{tr}((\beta - \hat{\beta})'\mathbf{X}'\mathbf{X}(\beta - \hat{\beta})) + \text{tr}[(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})] \quad (20)$$

$$= \text{tr}((\beta - \hat{\beta})'\mathbf{X}'\mathbf{X}(\beta - \hat{\beta})) + \text{tr}(\mathbf{Y}'\mathbf{Q}_X\mathbf{Y}) \quad (21)$$

また、上式において、

$$\mathbf{Y} \leftarrow \mathbf{W}^{1/2}\mathbf{Y} \quad (22)$$

$$\mathbf{X} \leftarrow \mathbf{W}^{1/2}\mathbf{X} \quad (23)$$

とすることにより、以下を得る。

1.3 残差の2乗和の分解 2

$\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$ とした場合、RSS は以下のように書き表される。

$$\text{RSS} = \text{tr}((\mathbf{Y} - \mathbf{X}\beta)'\mathbf{W}(\mathbf{Y} - \mathbf{X}\beta)) \quad (24)$$

$$= \text{tr}((\beta - \hat{\beta})'\mathbf{X}'\mathbf{W}\mathbf{X}(\beta - \hat{\beta})) + \text{tr}(\mathbf{Y}'\mathbf{W}^{1/2}\mathbf{Q}_X\mathbf{W}^{1/2}\mathbf{Y}) \quad (25)$$

2 線形制約

RSS を、 r 個の線形制約条件

$$\mathbf{L}\beta = \mathbf{C} \quad (26)$$

の下で最小化したい場合を考える。ただし、 \mathbf{L} ならびに \mathbf{C} は、それぞれ、 $r \times q$, $r \times p$ の行列であり、 $\text{rank}(\mathbf{L}) = nr(\mathbf{L}) = r$ とする。

制約条件の例としては、 β を 4×1 のベクトルとして、 $\beta_1 = \beta_2$, & $\beta_2 = \beta_3$ 等があるが、これは以下に示すような階数 2 の行列

$$L = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix}, \quad \text{and} \quad c = \mathbf{0} \quad (27)$$

を用いて書き表すことが出来る。

2.1 ラグランジュ乗数法

M を $r \times p$ の未定乗数として、

$$\text{RSS1} = \text{tr} \left[(\mathbf{Y} - \mathbf{X}\beta)' (\mathbf{Y} - \mathbf{X}\beta) \right] + 2\text{tr} (M(L\beta - C)) \quad (28)$$

$$= \text{tr} (\mathbf{Y}'\mathbf{Y}) - 2\text{tr} (\mathbf{Y}'\mathbf{X}\beta) + \text{tr} (\beta\mathbf{X}'\mathbf{X}\beta) + 2\text{tr} (ML\beta) - 2\text{tr} (MC) \quad (29)$$

を β と M の両方に関して最小化する。

Eq.(28) の β と M に関する偏微分を 0 と置いた方程式は

$$\frac{\partial \text{RSS1}}{\partial \beta} = 2(\mathbf{X}'\mathbf{X}\beta - \mathbf{X}'\mathbf{Y} + L'M') = \mathbf{O} \quad (30)$$

$$\frac{\partial \text{RSS1}}{\partial M} = L\beta - C = \mathbf{O} \quad (31)$$

となり、制約付きの解はこの連立方程式の解となっている。なお、未定乗数に関する偏微分部分、Eq.(31) に制約式そのものが出る。

Eq.(30) を、 β に関して解くと、

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y} - L'M') \quad (32)$$

を得るが、ここで、式の簡略化のため

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (33)$$

を用いると、

$$\beta = \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}L'M \quad (34)$$

となる。この両辺の左から L を掛けて制約条件を用いると、

$$L\beta = L\hat{\beta} - L(\mathbf{X}'\mathbf{X})^{-1}L'M = C \quad (35)$$

したがって、

$$M = (\hat{\beta}'L' - C')(L(\mathbf{X}'\mathbf{X})^{-1}L')^{-1} \quad (36)$$

となるが、これを Eq.(34) に代入すると以下の解を得る。

$$\beta = \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}L'M \quad (37)$$

$$= \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}L'(\hat{\beta}'L' - C')(L(\mathbf{X}'\mathbf{X})^{-1}L')^{-1} \quad (38)$$

$$= \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}L'(L(\mathbf{X}'\mathbf{X})^{-1}L')^{-1}(L\hat{\beta} - C) \quad (39)$$

なお、この解が確かに制約条件を満たすことは、左から L を掛けて確かめることが出来る。

$$L\beta = L\hat{\beta} - L(\mathbf{X}'\mathbf{X})^{-1}L'(L(\mathbf{X}'\mathbf{X})^{-1}L')^{-1}(L\hat{\beta} - C) \quad (40)$$

$$= L\hat{\beta} - (L\hat{\beta} - C) = C \quad (41)$$

2.1.1 ラグランジュの未定乗数法（補足）

例 1

$$f(x, y) = x^2 + y^2 - 3x - 4y \quad (42)$$

は、 $x = 1.5, y = 2$ において最小値 -6.25 を取る。この関数を、制約条件

$$x - y = 1 \quad (43)$$

の下で最小化する。

元の関数と、制約条件に未定常数 λ を掛けたものを x, y, λ の関数と見なし、

$$g(x, y, \lambda) = f(x, y) + \lambda(x - y - 1) \quad (44)$$

その制約無しでの最小値を求める。

g の各変数に関する偏微分を 0 とおいた方程式は

$$\frac{\partial g}{\partial x} = 2x + \lambda - 3 = 0 \quad (45)$$

$$\frac{\partial g}{\partial y} = 2y - \lambda - 4 = 0 \quad (46)$$

$$\frac{\partial g}{\partial \lambda} = x - y - 1 = 0 \quad (47)$$

となる。（最小とすべき関数 g のラグランジュの乗数に関する偏微分は、必ず制約条件式となる。）

上のふたつから、

$$x = (3 - \lambda)/2 \quad \text{and} \quad y = (4 + \lambda)/2 \quad (48)$$

を得るが、これを 3 番目、すなわち制約条件そのものに代入すると

$$-1 - \frac{1}{2}(\lambda + 4) - \frac{\lambda - 3}{2} = 0 \quad (49)$$

を得る。これから

$$\lambda = -\frac{3}{2} \quad (50)$$

となり、これを上式に代入すると

$$x = \frac{9}{4} \quad \text{and} \quad y = \frac{5}{4} \quad (51)$$

を得る。また、そのときの最小値は $f = -\frac{41}{8}$ となる。

例 2

例 1 と同じ関数

$$f(x, y) = x^2 + y^2 - 3x - 4y \quad (52)$$

を、異なる制約条件

$$x^2 + y^2 = 25 \quad (53)$$

の下で最小化する。

元の関数と、制約条件に未定常数 λ を掛けたものを x, y, λ の関数と見なし、

$$g(x, y, \lambda) = f(x, y) + \lambda(x^2 + y^2 - 25) \quad (54)$$

その制約無しでの最小値を求める。

g の各変数に関する偏微分を 0 とおいた方程式は

$$\frac{\partial g}{\partial x} = 2x + 2\lambda x - 3 = 0 \quad (55)$$

$$\frac{\partial g}{\partial y} = 2y + 2\lambda y - 4 = 0 \quad (56)$$

$$\frac{\partial g}{\partial \lambda} = x^2 + y^2 - 25 = 0 \quad (57)$$

となる。(最小とすべき関数 g のラグランジュの乗数に関する偏微分は、必ず制約条件式となる。)

上のふたつから、

$$x = \frac{3}{2(\lambda + 1)} \quad \text{and} \quad y = \frac{2}{\lambda + 1} \quad (58)$$

を得るが、これを 3 番目、すなわち制約条件そのものに代入すると

$$\frac{25}{4(\lambda + 1)^2} - 25 = 0 \quad (59)$$

を得る。これから

$$\lambda = -\frac{3}{2} \quad \text{or} \quad \lambda = -\frac{1}{2} \quad (60)$$

となり、これを上式に代入すると

$$x = -3, y = -4 \quad \text{or} \quad x = 3, y = 4 \quad (61)$$

を得る。また、そのときの f の最小値は、 $(x=3, y=4)$ において 0 となる。(最大値は 50 となる。)

例 3

線形回帰の残差の 2 乗和

$$\text{RSS} = \sum_{i=1}^n \left(v_i - \sum_{j=1}^q u_{ij} \xi_j \right)^2 = (\mathbf{v} - \mathbf{U}\boldsymbol{\xi})' (\mathbf{v} - \mathbf{U}\boldsymbol{\xi}) \quad (62)$$

を

$$\|\mathbf{U}\boldsymbol{\xi}\|^2 = d$$

という制約の下で RSS を最小とする解 $\boldsymbol{\xi}^{**}$ を求める。ただし、RSS の制約無しでの最小値を与える解を

$$\hat{\boldsymbol{\xi}} = (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{v} \quad (63)$$

とする。ラグランジュの乗数を λ とすれば、制約条件下での最小化基準は

$$Q = \text{RSS} + \lambda(d - \boldsymbol{\xi}'\mathbf{U}'\mathbf{U}\boldsymbol{\xi})$$

となり、その偏微分は以下ようになる。

$$\begin{aligned} \frac{\partial Q}{\partial \boldsymbol{\xi}} &= -2\mathbf{U}'\mathbf{v} + 2\mathbf{U}'\mathbf{U}\boldsymbol{\xi} - 2\mathbf{U}'\mathbf{U}\boldsymbol{\xi} \\ &= 2(-\mathbf{U}'\mathbf{v} + (1 - \lambda)\mathbf{U}'\mathbf{U}\boldsymbol{\xi}) \end{aligned}$$

したがって、

$$\begin{aligned} \boldsymbol{\xi}^{**} &= \frac{1}{(1 - \lambda)}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{v} \\ &= \frac{1}{(1 - \lambda)}\hat{\boldsymbol{\xi}} \end{aligned}$$

ここで、 Q のラグランジュの乗数に関する偏微分を 0 と置いて得られる制約条件式に上記の $\boldsymbol{\xi}^{**}$ を代入して得られる

$$\begin{aligned} \|\mathbf{U}\boldsymbol{\xi}^{**}\|^2 &= \left(\frac{1}{(1 - \lambda)}\right)^2 \hat{\boldsymbol{\xi}}'\mathbf{U}'\mathbf{U}\hat{\boldsymbol{\xi}} \\ &= \left(\frac{1}{(1 - \lambda)}\right)^2 \hat{\mathbf{v}}'\hat{\mathbf{v}} \\ &= d \end{aligned}$$

の関係を使うと

$$\frac{1}{(1 - \lambda)} = \sqrt{\frac{d}{\hat{\mathbf{v}}'\hat{\mathbf{v}}}}$$

すなわち、

$$\boldsymbol{\xi}^{**} = \frac{\sqrt{d}}{\|\hat{\mathbf{v}}\|}\hat{\boldsymbol{\xi}}$$

となる。

例 4

p カテゴリーの多項分布の確率を π_j , $j = 1, 2, \dots, p$, で表すと、

$$\sum_{j=1}^p \pi_j = 1 \tag{64}$$

である。第 j 番目のカテゴリーの度数を f_j とした時のに、Eq.(64) の条件の下で対数尤度

$$\ln L = \sum_{j=1}^p f_j \ln \pi_j \tag{65}$$

を最大とする解を求める。ラグランジュの乗数を λ とし

$$Q = \ln L + \lambda \left(1 - \sum_{j=1}^p \pi_j\right) \tag{66}$$

を考える。これをパラメタで偏微分すると

$$\frac{\partial Q}{\partial \pi_k} = \frac{f_k}{\pi_k} - \lambda \quad (67)$$

を得るが、これを 0 と置くことにより、

$$\pi_k = \frac{1}{\lambda} f_k \quad (68)$$

となる。ここで、上式を全て加えると

$$\sum_{j=1}^p \pi_k = \sum_{j=1}^p \frac{1}{\lambda} f_k = \frac{1}{\lambda} \sum_{j=1}^p f_k \quad (69)$$

となるが、Eq.(64) より（これは Q のラグランジュの乗数に関する偏微分を 0 と置いた結果である。）これは 1 である。したがって

$$\frac{1}{\lambda} \sum_{k=1}^p f_k = 1 \quad (70)$$

より、

$$\lambda = \sum_{k=1}^p f_k \quad (71)$$

を得る。また、

$$\pi_k = \frac{f_k}{\sum_{k=1}^p f_k} \quad (72)$$

である。

例 4

先の例にあらわれた対数尤度

$$\ln L = \sum_{j=1}^p f_j \ln \pi_j \quad (73)$$

を $r2 \times p$ の行列 \mathbf{B} と $r2 \times 1$ のベクトル \mathbf{c} を用いて

$$\mathbf{B}\boldsymbol{\pi} = \mathbf{c} \quad \text{or} \quad \sum_{j=1}^p b_{lj}\pi_j = c_l, \quad l = 1, 2, \dots, r2 \quad (74)$$

という制約条件の下で最大化することを考える。ただし、全成分の和が 1 となるという制約は上記の制約の中に含まれているものとする。（たとえば、 \mathbf{B} 行列の第 1 行が全て 1 で、 $c_1 = 1$ とする。）

Let us define the constrained log likelihood as

$$Q = \ln L - \sum_{l=1}^{r2} \lambda_l \left(\sum_{j=1}^p b_{lj}\pi_j - c_l \right), \quad (75)$$

where $\lambda_l, l = 1, 2, \dots, r_2$, is the Lagrangean multipliers. By differentiating Q with respect to π_j and setting the result equal to zero, we have,

$$\pi_j = \frac{f_j}{\sum_{l=1}^{r_2} \lambda_l b_{lj}}, \quad j = 1, 2, \dots, p. \quad (76)$$

Therefore, the original form B constraints can be written as

$$\sum_{j=1}^p b_{kj} \pi_j = \sum_{j=1}^p \frac{b_{kj} f_j}{\sum_{l=1}^{r_2} \lambda_l b_{lj}} = c_k \quad (77)$$

or,

$$\sum_{j=1}^p \frac{b_{kj} f_j}{\sum_{l=1}^{r_2} \lambda_l b_{lj}} - c_k = 0. \quad (78)$$

for $k = 1, 2, \dots, r_2$. This is the system of r_2 equations for unknown λ_l 's, and the numerical solution may be obtained by Newton's method. The necessary derivative matrix of the above equations is

$$\frac{\partial g_k}{\partial \lambda_m} = - \sum_{j=1}^p \frac{f_j b_{kj} b_{mj}}{(\sum_{l=1}^{r_2} \lambda_l b_{lj})^2}, \quad (79)$$

for $k, m = 1, 2, \dots, r_2$ where g_k stands for the left hand side of the k -th equation in Eq.(78). After obtaining the value of the Lagrangean multipliers, the maximum likelihood estimates of π can be calculated by substituting those values into equation Eq.(76).

Note that the same can be accomplished more easily if we use the equivalent form A constraints.

2.2 双対系

$p \times 1$ のパラメタ β にかかる線形の制約は、以下の二つの形に書くことができる。

Form A 制約 (A,d)

$$\beta = A\gamma + d, \quad (80)$$

ただし、 A は階数が r_1^* である $p \times r_1$ の行列、 d は $p \times 1$ のベクトル、 γ は $r_1 \times 1$ の任意のベクトルである。

Form B 制約 (B,c)

$$B\beta = c, \quad (81)$$

ただし、 B は、階数が r_2^* である $r_2 \times p$ の行列、 c は $r_2 \times 1$ のベクトルであるが、この制約が意味を持つためには、 c が B の列空間の中になければならない。

この二つの線形制約の形は互いに双対系をなしており、与えられた線形制約をどちらの形でも表現することが出来る。これらの二つの制約のどちらかが与えられた場合、他の一方への変換は以下のようにして行うことができる。

Form A から Form B への変換

階数が r_1^* の form A の制約 (A, d) が与えられたときに、それと同値の階数 $p - r_1^*$ の form B の制約 (B_A, c_A) は

$$B_A = A_R, \quad c_A = A_R d, \quad (82)$$

で与えられる。ただし、 A_R は $p - r_1^* \times p$ の行列であり、

$$Q_A = I_p - A(A'A)^{-1}A' = A_L A_R. \quad (83)$$

を満たすものである。(一意には定まらない。) すなわち、 B_A 行列として、 A 行列の列空間に直交するような、階数が r_1^* のフルランクの行列を用いればよい。

したがって

$$B_A \beta = A_R(A\gamma + d) \quad (84)$$

$$= A_R A\gamma + A_R d = \mathbf{0} + c_A \quad (85)$$

となる。

Form B から Form A への変換

階数が r_2^* の Form B の制約 (B, c) が与えられた場合、それと同値の階数 $p - r_2^*$ の form A の制約 (A_B, d_B) は

$$A_B = B_L, \quad d_B = (B'B)^{-1}B'c, \quad (86)$$

となる。ただし、 B_L は $p \times p - r_2^*$ の行列で

$$Q_{B'} = I_p - B'(BB')^{-1}B = B_L B_R. \quad (87)$$

を満たすものである。(一意には定まらない。) すなわち、 A_B 行列として、 B 行列の行空間に直交するような、階数が $p - r_2^*$ のフルランクの行列を用いればよい。

したがって

$$B\beta = B(A_B\gamma + d_B) = B(B_L\gamma + (B'B)^{-1}B'c) \quad (88)$$

$$= BB_L\gamma + B(B'B)^{-1}B'c = \mathbf{0} + c \quad (89)$$

となる。(整合性のある制約の場合)

一般に、Form B の形の制約は、その意味することを理解しやすい形であるが、前節に示したように、ラグランジュの方法を用いたその取り扱いが面倒である。片や Form A は、その意味するところの理解は難しいが、解がきわめて簡単となり、特に線形モデルの場合は、求める $\hat{\beta}$ は

$$\hat{\beta} = (A'X'XA)^{-1}A'X'y + d \quad (90)$$

となる。

先の例に示した (L, c) に対応する (A, d) は以下ようになる。

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad d = \mathbf{0} \quad (91)$$