

端点制約付き単調スプライン

前川眞一

東京工業大学大学院 社会理工学研究科 人間行動システム専攻

031102,03,30,1201,02,03,04 2003. 12. 4

1 モデル

昇冪の順に並べられたデータ、 $x_i, y_i, i = 1, 2, \dots, n$ が与えられているとする。これに、 K 個の節点を持つ M 次の単調増加スプライン関数

$$\hat{y} = \mathbf{H}^{(K,M)}(\mathbf{x})\boldsymbol{\beta} \quad (1)$$

を最小2乗法を用いてフィットすることを考える。ただし、 $\mathbf{H}^{(K,M)}(\mathbf{x})$ は $n \times (M + K + 1)$ の B スプライン行列であり、

$$\text{if } x_i \leq x_j \text{ then } \hat{y}_i \leq \hat{y}_j \quad (2)$$

である。単調増加の条件は、 H が B スプライン行列の場合、回帰係数に

$$\beta_1 \leq \beta_2 \leq \dots \leq \beta_q, q = M + K + 1 \quad (3)$$

の制約を課すことによって得られるが、これは、その要素として 1 のみを持つ下三角行列（対角を含む下三角の部分全て 1、残りは全て 0） \mathbf{G} を用いたモデル

$$\hat{y} = \mathbf{H}^{(K,M)}(\mathbf{x})\mathbf{G}\mathbf{b}, b_j > 0, j = 2, 3, \dots, q \quad (4)$$

と同値であり、

$$b_j = c_j^2, j = 2, 3, \dots, q \quad (5)$$

と再度書き直すことにより、パラメタに制約を課さない形に書き換えることが出来る。その解法に関しては、前川 (1997, section 3.4.3, 4.3.2.1, 4.4.2, 4.6.1) を参照。

2 端点に関する制約

本稿では、上記の単調スプラインが、その両端で

$$(x_1, \min y) \quad \text{and} \quad (x_n, \max y), \min y \leq \max y \quad (6)$$

を通るという制約の下に、回帰係数を求める方法について述べる。

記述を簡略にするため、回帰モデルを

$$\hat{y} = \mathbf{Z}\mathbf{b}, b_j > 0, j = 2, 3, \dots, q \quad (7)$$

ただし、

$$\mathbf{Z} = \mathbf{H}^{(K,M)}(\mathbf{x})\mathbf{G} \quad (8)$$

と書き換える。 \mathbf{H} 行列の性質から、その第 1 行目は第 1 列目に一つの 1 が入り、残りは 0 である。同様に、その第 n 行目は第 n 列目に一つの 1 が入り、残りは 0 である。したがって、 \mathbf{Z} 行列の第 1 行目は第 1 列目に一つの 1 が入り、残りは 0、また第 n 行目は全て 1 となるため、条件 Eq.(6) は

$$\hat{y}_1 = b_1 = \text{miny}, \quad \text{and} \quad \hat{y}_n = \sum_{j=1}^q b_j = \text{maxy} \quad (9)$$

となる。すなわち、制約条件 Eq.(6) の元でデータに単調スプラインをフィットすることは、Eq.(7) のモデルのパラメタを

$$b_1 = \text{miny}, \quad \sum_{j=1}^q b_j = \text{maxy}, \quad b_j > 0, \quad j = 2, 3, \dots, q \quad (10)$$

という制約条件の下で求めることに等しい。したがって、制約条件無しの場合 c を用いた以下のモデル

$$b_1 = \text{miny} \quad (11)$$

$$b_j = (\text{maxy} - \text{miny}) \times \frac{c_j^2}{\sum_{\ell=2}^q c_\ell^2}, \quad j = 2, \dots, q \quad (12)$$

と同値である。

2.1 解法

Gauss-Newton 法を用いるとして、必要となる Jacobian 行列は以下の通りである。

$$\frac{\partial \hat{y}_{ij}}{\partial c_k} = 2(\text{maxy} - \text{miny}) z_{ik} \left(\frac{c_k}{\sum_{\ell=2}^q c_\ell^2} - \frac{c_k^3}{(\sum_{\ell=2}^q c_\ell^2)^2} \right) \quad (13)$$

$$-2(\text{maxy} - \text{miny}) \frac{c_k}{\sum_{\ell=2}^q c_\ell^2} \sum_{j \neq 1, k} z_{ij} c_j^2, \quad k = 2, 3, \dots, q \quad (14)$$

なぜなら、

$$\hat{y}_{ij} = z_{i1} b_1 + (\text{maxy} - \text{miny}) \sum_{j=2}^q z_{ij} b_j \quad (15)$$

$$= z_{i1} b_1 + (\text{maxy} - \text{miny}) z_{ik} \frac{c_k^2}{\sum_{\ell=2}^q c_\ell^2} + (\text{maxy} - \text{miny}) \sum_{j \neq 1, k} z_{ij} \frac{c_j^2}{\sum_{\ell=2}^q c_\ell^2} \quad (16)$$

であり、

$$\frac{\partial \frac{c_k^2}{\sum_{\ell=2}^q c_\ell^2}}{\partial c_k} = 2 \left(\frac{c_k}{\sum_{\ell=2}^q c_\ell^2} - \frac{c_k^3}{(\sum_{\ell=2}^q c_\ell^2)^2} \right) \quad (17)$$

$$\frac{\partial \frac{c_j^2}{\sum_{\ell=2}^q c_\ell^2}}{\partial c_k} = -2 \frac{c_j^2 c_k}{(\sum_{\ell=2}^q c_\ell^2)^2} \quad (j \neq 1, k) \quad (18)$$

となるからである。なお、 c はその単位に関して不定性があるため、繰り返し計算のなかで、その 2 乗和を 1 に固定しておくことと便利である。

なお、 $miny$ が指定されていない場合は、上記のモデルにおいて

$$miny = b_1 \quad (19)$$

と置けば

$$\hat{y}_i = z_{i1}b_1 + (maxy - b_1) \sum_{j=2}^q z_{ij}b_j \quad (20)$$

であるから、

$$\frac{\partial \hat{y}_i}{\partial b_1} = z_{i1} - \sum_{j=1}^q z_{ij}b_j \quad (21)$$

を利用して b_1 を同時に求めればよい。

また、 $miny$ は指定されているが、 $maxy$ が指定されていない場合は、通常の単調スプラインのパラメタを求める際に、 b_1 を $miny$ に固定したものが解となる。したがって、わざわざ Eq.(12) の様なリパラメトライゼーションをする必要はない。しかし、強いてこのモデルの下で $maxy$ をも推定しようとするならば、

$$\hat{y}_i = z_{i1}b_1 + (maxy - b_1) \sum_{j=2}^q z_{ij} \frac{c_j^2}{\sum_{\ell=2}^q c_\ell^2} \quad (22)$$

であるから、 b_1 と c が求まった後に、 $maxy$ は線形の回帰分析を用いて

$$maxy = \frac{\sum_{i=1}^n (y_i - (z_{i1} - \sum_{j=2}^q z_{ij} \frac{c_j^2}{\sum_{\ell=2}^q c_\ell^2})b_1) \sum_{j=2}^q z_{ij} \frac{c_j^2}{\sum_{\ell=2}^q c_\ell^2}}{\sum_{i=1}^n (\sum_{j=2}^q z_{ij} \frac{c_j^2}{\sum_{\ell=2}^q c_\ell^2})^2} \quad (23)$$

として推定をすることが出来る。なお、繰り返し計算に際しては、この推定の直後に Eq.(12) を用いて b を再定義する必要がある。