Bスプラインと多項式回帰モデルにおける優越性の検定

総合研究大学院大学　複合科学研究科　統計科学専攻
博士課程　　加藤　直広

1　回帰モデルにおける優越性の検定

2群それぞれに対して目的変数 $y_i(i=0,1)$ が $y_i = f_i(t)(t \in T)$ と表されているとする．ここで，T は説明変数の定義域である．

上の検定に対する帰無仮説と対立仮説は

帰無仮説 H_0：　任意の $t \in T$ に対して $f_0(t) = f_1(t)$

対立仮説 H_1：　任意の $t \in T$ に対して $f_0(t) \le f_1(t)$,　　ある $t \in T$ に対して $f_0(t) < f_1(t)$

と表される．群間の差を $f(t:c) = f_1(t) - f_0(t) = c'\psi(t)$ と定義すると，上の検定問題は

$$H_0 : c = 0$$
$$H_1 : c \in K - \{0\}, \quad K = \{c \mid \text{任意の } t \in T \text{ に対して } f(t:c) \ge 0\}$$

と書き直せる．

c の推定量 $\hat{c}$ が平均 c，共分散行列 Σ を持つ正規分布に従っていると仮定する．

この場合，尤度比統計量は $\lambda = ||\hat{c}_K||^2_{\Sigma^{-1}}$ となる．ここで，

$$||x||^2_{\Sigma^{-1}} = x'\Sigma^{-1}x, \quad \hat{c}_K = \underset{x \in K}{\operatorname{argmin}} ||\hat{c} - x||^2_{\Sigma^{-1}} \tag{1.1}$$

である．

本研究では，f_i がBスプライン関数の場合と多項式の場合に λ の帰無仮説の下での分布を求める．

2　Bスプラインと多項式による回帰

定理. n を c の次元とするとき次が成り立つ．

(1)　c が K の内点 $\leftrightarrow f(x:c)$ は T で零点を持たない．

(2)　c が K の境界 $\leftrightarrow f(x:c)$ は T で零点を持つ．

(3)　c が K の extreme ray $\leftrightarrow f(x:c)$ は T で $n-1$ 個の零点を持つ．

f が多項式の場合，チェビシェフ系の一般論によりこの定理は証明されている．f がBスプライン関数の場合，別に証明を与えた．

始めに，多項式の場合を考える．本発表では特に，$f_i(t)$ が高々3次式の場合を扱う．このとき，$\psi(t) = (t^3, t^2, t, 1)'$，$T = [0, \infty)$ となる．上の定理より K の境界は $f(x:c)$ が零点を持つときである．Markov-Lukacs の定理（詳しくは [3] を参照.）により，K の境界を5個に分類しパラメーター表記する．

境界の例として以下が考えられる．

$$B = \{c_1 = (q_1^2, -2q_1q_0, q_0^2, 0)' \mid q_1, q_0 > 0\}.$$

このとき，$f(t : c_1)$ のグラフは図 1 のようになる．

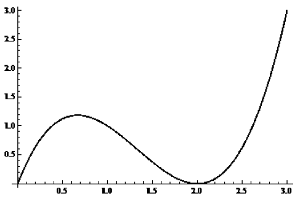

図 1 K の境界の例

ここで，c_1 での法錐は以下のように書ける．

$$N = \{v_1 = -n_1(q_0^2, q_1q_0, q_1^2, 0)' - n_2(0, 0, 0, 1)' \mid n_1, n_2 \geq 0\}.$$

これらを用いることにより，c での第 2 基本形式が HG^{-1} と書けることが解る．ここで，

$$H(c_1, v_1) = \begin{pmatrix} 2q_0^2n_1 & -2q_1q_0n_1 \\ -2q_1q_0n_1 & 2q_1^2n_1 \end{pmatrix}, \quad G(c_1) = \begin{pmatrix} 4q_1^2 + 4q_0^2 & 4q_1q_0 \\ 4q_1q_0 & 4q_1^2 + 4q_0^2 \end{pmatrix}$$

である．

次に，B スプライン関数の場合を考える．[1] より，$\psi(t) = (M_{31}(t), \ldots, M_{34}(t))'$ となる．本発表では各節点 ξ_i に対して $\xi_i - \xi_{i-1} = a$ を仮定した 2 次の B スプライン関数を考える．$T = [\xi_0, \xi_2]$ とすると，境界の例として以下が挙げられる．

$$B = \left\{c_2 = r\cos t(\tan t - 2a^2 - y_0^2 + 3ay_0, 2a^2 + y_0^2 - 3ay_0, y_0^2 - ay_0, y_0^2 + ay_0)' \,\middle|\, r > 0, t \in \left(0, \frac{\pi}{2}\right), y_0 \in [0, a)\right\}$$

ここで，c_1 での法錐は以下のように書ける．

$$N = \{v_2 = -n(0, 2y_0^2, 2a^2 + 4ay_0 - 4y_0^2, 2a^2 - 4ay_0 + 2y_0^2)' \mid n \geq 0\}.$$

これらを用いて，3 次式の場合と同様に第 2 基本形式を求めることができる．

[2] より，上で求めた境界のパラメーター表記と第 2 基本形式を用いて χ^2 分布の重み w_i は

$$w_{4-e} = \frac{1}{\Omega_{4-e}\Omega_e} \sum_B \sum_N \int_B du \int_{N \cap S^3} dv \, \mathrm{tr}_{\,d-3+e}(H(u, v)G^{-1}(v)), \quad e = \dim(B), \tag{2.1}$$

となる．ここで Ω_i は i 次元単位球面 S^i の体積である．

参考文献

[1] 市田 浩三 , 吉本 富士市 (1979).「スプライン関数とその応用 」, 教育出版.

[2] A. Takemura and S. Kuriki (2002). On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Ann. Appl. Probab.* , **12**, 768–796.

[3] H. Frenk, K. Roos, T. Terlaky and S. Zhang (2000). *High Performance Optimization.* Kluwer Academic Publishers.