分解可能部分モデルに基づく
比例反復法の改良

46205　遠藤　祐司
指導教員　竹村彰通 教授

平成 18 年 2 月 6 日

概 要

分割表における階層モデルのセル確率の最尤推定量は一般には明示的に得られない．このような場合には反復計算を用いる必要がある．最も一般的な方法は比例反復法であり，これは生成集合族の要素に一つづつ順に着目してセル確率を更新していく方法である．今回は，分解可能モデルに基づいて，より大きなまとまりに着目して更新を行う比例反復法の改良を提案する．提案する更新アルゴリズムが最尤推定量に収束することを示し，さらに収束先に近い場合には提案する方法が通常の比例反復法より良い更新法となっていることを示す．

1　階層モデルと分解可能モデル

今回の話題となる分割表におけるモデルを以下に記す．紹介するモデルには，対数線形モデル $\supset$ 階層モデル $\supset$ 分解可能モデル という包含関係がある．

定義．　階層モデル
対数線形モデルにおいて高次の交互作用を考える時，その交互作用の変数によって構成された低次の交互作用項や主効果項を全て残すようなモデルを階層モデルと呼ぶ．階層モデルの項のうち，ある主効果項を含む最も高次の項を極大項という．また，極大項の集合を生成集合という．

定義．　分解可能モデル
生成集合 $\Lambda = \{C_1, C_2, \cdots, C_k\}$ の要素が，添え字を適当に並べ替えることで

$$C_t \cap (C_1 \cup \cdots \cup C_{t-1}) = C_t \cap C_r$$
$$\text{ここで } r \text{ は}, 1, 2, \cdots, t-1 \text{ のいずれかの値}$$

を，$t = 2, \cdots, k$ の全てで満たすとき，そのグラフィカルモデルを分解可能モデル (decomposable model) という．対応する弦グラフ G のクリークの集合を C，minimal vertex separator の集合を S，その重複度を $\nu(S)$，相対頻度を $m(i)$ としたとき，セル確率の最尤推定量は次のように求められる．

$$p(i) = \frac{\prod_{C \in \mathcal{C}} m(i_C)}{\prod_{S \in \mathcal{S}} m(i_S)^{\nu(S)}}$$

2　比例反復法

分解可能モデルにおいてはセル確率の最尤推定量が明示的に求められるが，一般に階層モデルにおいては反復計算が必要となる．代表的な方法が次に記す比例反復法であり，一度の更新で生成集合族の一つの要素が更新される．

アルゴリズム 1（通常の比例反復法）
$p^{(0)}(i) = \mathbf{1}$ とする．階層モデルの生成集合を Λ とし，それぞれの要素 $\lambda \in \Lambda$ について，

$$p^{(t+1)}(i) = p^{(t)}(i) \times \frac{m(i_\lambda)}{p^{(t)}(i_\lambda)}$$

の更新を行う．ただし t はステップ数である．

3　比例反復法の改良

本研究で提案する具体的なアルゴリズムは次のようなものである．

アルゴリズム 2.
階層モデルの生成集合を $\mathcal{C}$ とする．この時，クリークの集合としてそれぞれ $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m$ を持つ m 個の弦グラフ $G_1(V_1, E_1)$，$G_2(V_2, E_2), \ldots, G_m(V_m, E_m)$ が存在し，

$$\mathcal{C} = \bigcup_{j=1}^{m} \mathcal{C}_j$$

を満たすとする．また，$G_1, G_2, \ldots, G_m$ のセパレーターの集合をそれぞれ $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_m$ とする．この時，$j = 1, 2, \ldots m$ について反復を，

$$q_i^{(t+1)} = q_i^{(t)} \times \frac{\prod_{C \in \mathcal{C}_j} m_{i_C}}{\prod_{S \in \mathcal{S}_j} m_{i_S}} \times \frac{\prod_{S \in \mathcal{S}_j} q_{i_S}^{(t)}}{\prod_{C \in \mathcal{C}_j} q_{i_C}^{(t)}}$$

のように定義する．ただし t はステップ数である．

このように定義した**アルゴリズム** 2 は生成集合族の複数の要素について一度に更新を行う．ただし，**アルゴリズム** 2 では更新後の確率の和が 1 を越えてしまうため収束性の証明が困難である．そこで，次の**アルゴリズム** 3，**アルゴリズム** 4 を考える．

アルゴリズム 3.
アルゴリズム 2 と同様の条件の下で反復を，

$$p_i^{(t+1)} = p_i^{(t)} \times \left(\frac{\prod_{C \in \mathcal{C}_j} m_{i_C}}{\prod_{S \in \mathcal{S}_j} m_{i_S}} \times \frac{\prod_{S \in \mathcal{S}_j} p_{i_S}^{(t)}}{\prod_{C \in \mathcal{C}_j} p_{i_C}^{(t)}} \right)^{\alpha}$$

のように定義する．ただし t はステップ数である．

補題 1．アルゴリズム 3 において，更新後の和が 1 となるような α_0 が唯一存在する．

このような α_0 を用いて次の**アルゴリズム** 4 を定義する．

アルゴリズム 4.
アルゴリズム 2 と同様の条件の下で反復を，

$$p_i^{(t+1)} = p_i^{(t)} \times \left(\frac{\prod_{C \in \mathcal{C}_i} m_{i_C}}{\prod_{S \in \mathcal{S}_i} m_{i_S}} \times \frac{\prod_{S \in \mathcal{S}_i} p_{i_S}^{(t)}}{\prod_{C \in \mathcal{C}_i} p_{i_C}^{(t)}} \right)^{\alpha_0}$$

のように定義する．ただし t はステップ数である．

このようにして定義した**アルゴリズム** 4 と，ある条件下での**アルゴリズム** 3 は収束性を証明することができる．

定理 1．アルゴリズム 4 は最尤推定量に収束する．
証明の概要

$$\begin{aligned} I(p^*; p^{(t+1)}) &= I(p^*; p^{(t)}) \\ &- \alpha_0 \sum_i p_i^* \log \frac{\prod_{C \in \mathcal{C}_1} p_{i_C}}{\prod_{S \in \mathcal{S}_1} p_{i_S}} \times \frac{\prod_{S \in \mathcal{S}_1} p_{i_S}^{(t)}}{\prod_{C \in \mathcal{C}_1} p_{i_C}^{(t)}} \end{aligned}$$

であるが，右辺第二項が非負となることを示すことが出来る．

系 1． $\alpha \leq \alpha_0$ であるような α を用いた場合，**アルゴリズム** 3 は最尤推定量に収束する．

また，改良された比例反復法が次の定理の意味で通常の比例反復法よりも良い事が分かる．

定理 2．
t ステップ目において更新中のセルが収束先に十分近いとし，次のステップにおいて**アルゴリズム** 1，**アルゴリズム** 4 を実行した後の収束先からの KL 情報量を I_1, I_4 とする．この時次の式が成り立つ．

$$I_1(p^* : p^{(t+1)}) \geq I_4(p^* : p^{(t+1)})$$

証明の概要
収束先に十分近い場合には α_0 が十分 1 に近く，

$$I_4(p^{(*)}, p^{(t+1)}) = I_1(p^{(*)}, p^{(t+1)}) - \sum_i p_i^* \log \frac{m_{i_{C_j}} \times p_{i_{S_{j-1}}}^{(t)}}{m_{i_{S_{j-1}}} \times p_{C_j}^{(t)}}$$

と書くことができ，右辺第 2 項は非負であることを示すことが出来る．

また上記の定理より，分解可能部分モデルはより大きいものを考えた方が良いことが分かる．

4 今後の課題

- **アルゴリズム** 2 の収束性の証明
- 条件を満たす分解可能部分モデルの集合を得るアルゴリズム
- 効率的な実装方法

参考文献

[1] Kullback, S. (1968), Probability densities with given marginals. *Ann. Math. Statist.*, Vol.39, pp.1236–1243.

[2] Darroch, J. N. and Ratcliff, D. (1972), Generalized iterative scaling for log-linear models. *Ann. Math. Statist.*, Vol.43, pp.1470–1480.

[3] Ruschendorf, L. (1995), Convergence of the iterative proportional fitting procedure. *Ann. Stat.*, Vol.23, pp.1160–1174.