Bayes 法による潜在クラスモデルのパラメータ推定

黒田 正博
倉敷芸術科学大学産業科学技術学部

(1998 年 9 月 30 日 受理)

1　はじめに

潜在構造分析は，直接観測することのできない潜在的な要因を観測可能なデータから探っていく手法であり，社会学，心理学あるいは医学等の分野においてよく用いられる．このように観測可能な変数 (顕在変数) のほかに，直接観測できない変数 (潜在変数) を組み込んだモデルが潜在変数モデルであり，特に，顕在変数および潜在変数が共にカテゴリカル変数であるモデルを潜在クラスモデルと呼ぶ．

潜在クラスモデルのパラメータを推定する方法として，EM アルゴリズム [3] による最尤推定法がよく適用される [6]．このアルゴリズムにもとづく推定法では，初期値の選択が適切であれば尤度関数を最大にすることが知られている．しかし，一般にはそれが一意に定まることは保証されていないので [7]，幾つかの初期値の組み合わせを試し，尤度関数を最大にするものを推定値として採用する場合がある [5]．このとき，それは事前知識をもとに求められるのであろうが，それでも尤度を最大にするような初期値を見つけるのは，困難であることが予想される．このように対象としている現象に関して，通常何らかの事前知識をもって解析にあたることが多い．その際，彼らが確信度の高い事前知識をもっていれば，それを推定に積極的に活用しようとするであろう．このような知識を推定に反映させる方法の 1 つとして，潜在クラスモデルのパラメータに事前分布を仮定する Bayes 的アプローチによる推定法の適用が考えられる．また，事前分布を設定することにより，潜在クラスモデルの解析の際に考慮しなければならないモデルおよびパラメータの同定可能性に関する問題も解消できる．

本論文では，Bayes の逐次学習による潜在クラスモデルのパラメータ推定法を提案する．また，数値シュミレーションにより，事前分布のパラメータの選択が推定値に及ぼす影響を調べる．そして，推定値と真の確率を比較をすることで推定法の有効性を検証する．

2　Bayes の遂次学習

2.1　潜在クラスモデルとそのパラメータ

顕在変数 $X_1, \cdots, X_N$，潜在変数 Y からなる潜在クラスモデルを考える．ここで，変数の要素をそれぞれ $x_1, \cdots, x_N, y$ であらわし，$C_1, \cdots, C_N, J$ 個のカテゴリーをそれぞれもつとする．変数の集合を $Z = (X_1, \cdots, X_N, Y) = (X, Y)$，その要素の集合を $z = (x_1, \cdots, x_N, y) = (x, y)$ としたとき，$Z = z$ における同時確率を $p_Z(z) = \Pr\{Z = z\}$ であらわす．このとき

局所独立性の仮定により，その同時確率 $p_Z(z)$ は，

$$p_Z(z) = p_Y(y) \prod_{k=1}^{N} p_{X_k}(x_k|y)$$

と分解することができる．ここで，

$$\begin{aligned} p_{X_k}(x_k|y) &= \Pr\{X_k = x_k | Y = y\}, & (1) \\ p_Y(y) &= \Pr\{Y = y\} = \sum_{k=1}^{N} \sum_{x_k=1}^{C_k} p_Z(z). & (2) \end{aligned}$$

2.2 Bayes の逐次学習による事後分布の推定

モデルパラメータ $\{p_Z(z)\}$ の事前分布として hyper Dirichlet 分布 [2] を仮定する．このとき，事前分布 $\pi(\{p_Z(z)\} \mid \{\alpha_Z(z)\})$ は以下のように分解することができる．

$$\begin{aligned} \pi(\{p_Z(z)\} \mid \{\alpha_Z(z)\}) &= \pi(\{p_Y(y)\} \mid \{\alpha_Y(y)\}) \prod_{k=1}^{N} \pi(\{p_{X_k}(x_k|y)\} \mid \{\alpha_{Y_k}(y_k)\}) \\ &= C \prod_{y=1}^{J} p_Y(y)^{\alpha_Y(y)-1} \prod_{k=1}^{N} \prod_{x_k=1}^{C_k} p_{X_k}(x_k|y)^{\alpha_{Y_k}(y_k)-1}. \end{aligned}$$

ここで，$Y_k = (X_k, Y)$ $(k = 1, \cdots, N)$ であり，

$$\begin{aligned} \alpha_Y(y) &= \sum_{k=1}^{N} \sum_{x_k=1}^{C_k} \alpha_Z(z) \\ \alpha_{Y_k}(y_k) &= \sum_{l \neq k}^{N} \sum_{x_l=1}^{C_l} \alpha_Z(z) \\ C^{-1} &= \prod_{y=1}^{J} \frac{\Gamma[\alpha_Y(y)]}{\Gamma[\alpha(+)]} \prod_{k=1}^{N} \prod_{x_k=1}^{C_k} \frac{\Gamma[\alpha_{Y_k}(y_k)]}{\Gamma[\alpha_Y(y)]}. \end{aligned}$$

また，事前知識の確信度は事前分布のハイパーパラメータに

$$\alpha(+) = \sum_{y=1}^{J} \sum_{k=1}^{N} \sum_{x_k=1}^{C_k} \alpha_Z(z)$$

によって与えられ，精度（precision）と呼ばれる．

このとき，ある要素 $Z = \tilde{z} = (\tilde{x}_1, \cdots, \tilde{x}_N, \tilde{y})$ における事前分布の期待値は，

$$\begin{aligned} \mathrm{E}[p_Z(\tilde{z})] &= \mathrm{E}\left[p_Y(\tilde{y}) \prod_{k=1}^{N} p_{X_k}(\tilde{x}_k|\tilde{y}) \right] \\ &= C \prod_{y=1}^{J} \frac{\Gamma[\alpha_Y(y) + \delta_{y\tilde{y}}]}{\Gamma[\alpha(+) + 1]} \prod_{k=1}^{N} \frac{\Gamma[\alpha_{Y_k}(y_k) + \delta_{y_k \tilde{y}_k}]}{\Gamma[\alpha_Y(y) + \delta_{y\tilde{y}}]} \\ &= \frac{\alpha_Y(\tilde{y})}{\alpha(+)} \prod_{k=1}^{N} \frac{\alpha_{Y_k}(\tilde{y}_k)}{\alpha_Y(\tilde{y})}. \end{aligned}$$

ここで，δ はクロネッカーデルタである．

潜在クラスモデルでは，直接観測可能な変数は顕在変数のみであり，潜在変数に関する観測値は常に欠測している．

観測値として

$$X = \tilde{x} = (\tilde{x}_1, \cdots, \tilde{x}_N)$$

が得られたとき，その尤度 L は

$$\begin{aligned} L &= p_X(\tilde{x}) = \Pr\{X = (\tilde{x}_1, \cdots, \tilde{x}_N)\} \\ &= \sum_{\tilde{y}=1}^{J} p_Y(\tilde{y}) \prod_{k=1}^{N} p_{X_k}(\tilde{x}_k|\tilde{y}) \end{aligned} \tag{3}$$

であり，事後分布 $\pi(\{p_Z(z)\} \mid X = \tilde{x})$ は

$$\begin{aligned} \pi(\{p_Z(z)\} \mid X = \tilde{x}) &= C_X \sum_{y=1}^{J} p_Y(\tilde{y}) \prod_{k=1}^{N} p_{X_k}(\tilde{x}_k|\tilde{y}) \prod_{y=1}^{J} p_Y(y)^{\alpha_Y(y)-1} \\ &\quad \times \prod_{k=1}^{N} \prod_{x_k=1}^{C_k} p_{X_k}(x_k|y)^{\alpha_{Y_k}(y_k)-1} \end{aligned} \tag{4}$$

である．ここで，

$$C_X^{-1} = \sum_{y=1}^{J} \prod_{y=1}^{J} \frac{\Gamma[\alpha_Y(y) + \delta_{y\tilde{y}}]}{\Gamma[\alpha(+) + 1]} \prod_{k=1}^{N} \prod_{x_k=1}^{C_k} \frac{\Gamma[\alpha_{Y_k}(y_k) + \delta_{y_k\tilde{y}_k}]}{\Gamma[\alpha_Y(y) + \delta_{y\tilde{y}}]}.$$

この Bayes の逐次学習では観測値に欠測値が含まれるため，事前分布に自然共役な分布を選んでも分布の再生性は保たれず，事後分布 (4) は混合 hyper Dirichlet 分布

$$\pi(\{p_Z(z)\} \mid X = \tilde{x}) = \sum_{\tilde{y}=1}^{J} w_X(\tilde{y}) \pi(\{p_Z(z)\} \mid \{\alpha_Z(z) + \delta_{z\tilde{z}}\}) \tag{5}$$

で表現される．ここで，

$$\begin{aligned} w_X(\tilde{y}) &= C_X \prod_{y=1}^{J} \frac{\Gamma[\alpha_Y(y) + \delta_{y\tilde{y}}]}{\Gamma[\alpha(+) + 1]} \prod_{k=1}^{N} \prod_{x_k=1}^{C_k} \frac{\Gamma[\alpha_{Y_k}(y_k) + \delta_{y_k\tilde{y}_k}]}{\Gamma[\alpha_Y(y) + \delta_{y\tilde{y}}]} \\ &= \mathrm{E}[p_Y(\tilde{y}|\tilde{x})]. \end{aligned}$$

このとき，事後分布の期待値は

$$\mathrm{E}\left[p_Z(z) \mid X = \tilde{x}\right] = \sum_{\tilde{y}=1}^{J} w_X(\tilde{y}) \mathrm{E}[p_Z(z) \mid Z = (\tilde{x}, \tilde{y})]$$

であり，2 次モーメントは

$$\mathrm{E}\left[p_Z^2(z) \mid X = \tilde{x}\right] = \sum_{\tilde{y}=1}^{J} w_X(\tilde{y}) \mathrm{E}[p_Z^2(z) \mid Z = (\tilde{x}, \tilde{y})]$$

である．

また，$\{p_{Y_k}(y_k)\}$ $(k = 1, \cdots, N)$ の事後分布 $\pi(\{p_{Y_k}(y_k)\} \mid X = \tilde{x})$ は，

$$\pi(\{p_{Y_k}(y_k)\} \mid X = \tilde{x}) = \sum_{\tilde{y}=1}^{J} w_X(\tilde{y}) \pi(\{p_{Y_k}(y_k)\} \mid \{\alpha_Y(y_k) + \delta_{y_k\tilde{y}_k}\}) \tag{6}$$

であり，その期待値および 2 次モーメントはそれぞれ

$$\mathrm{E}[p_{Y_k}(y_k) \mid X = \tilde{x}] = \frac{\alpha_{Y_k}(y_k) + \delta_{y_k \tilde{y}_k} w_X(\tilde{y}_k)}{\alpha(+) + 1}, \qquad (7)$$

$$\mathrm{E}[p^2_{Y_k}(y_k) \mid X = \tilde{x}] = \frac{[\alpha_{Y_k}(y_k) + 1][\alpha_{Y_k}(y_k) + 2\delta_{y_k \tilde{y}_k} w_X(\tilde{y}_k)]}{[\alpha(+) + 2][\alpha(+) + 1]}.$$

これより周辺事後分散は

$$\mathrm{Var}[p_{Y_k}(y_k) \mid X = \tilde{x}] = \mathrm{E}[p^2_{Y_k}(y_k) \mid X = \tilde{x}] - E[p_{Y_k}(y_k) \mid X = \tilde{x}]^2$$

によって求めることができる．

3 混合事後分布の近似

本論文で仮定する Bayes の逐次学習では，観測値が得られる度に事後分布を順次もとめる．しかし，これを混合分布として表現した場合，データの増大に伴い非常に複雑な密度関数をもつ分布になり，モデルパラメータを推定することが困難になることが考えられる．これを回避するための方法の 1 つとして，近似分布による置き換えが考えられる．混合分布を近似する方法としては，deterministic あるいは stochastic な手法が提案され [10]，数値シュレーションによってこれらの手法の性能比較もおこなわれている [1].

ここでは, 混合事後分布 (5) を single hyper Dirichlet 分布 $\pi(\{p_Z(z)\} \mid \{\alpha^*_Z(z)\})$ によって近似する [9].

まず混合事後分布の平均分散和を以下のように定義する．

$$\mathrm{EV}_{Mixture} = \sum_{y=1}^{J} \sum_{k=1}^{N} \sum_{x_k=1}^{C_k} \mathrm{E}[p_{Y_k}(y_k) \mid X = \tilde{x}] \mathrm{Var}[p_{Y_k}(y_k) \mid X = \tilde{x}]. \qquad (8)$$

これより，近似分布 $\pi(\{p_Z(z)\} \mid \{\alpha^*_Z(z)\})$ の平均分散和 EV_{Single} は，

$$\mathrm{EV}_{Single} = \sum_{y=1}^{J} \sum_{k=1}^{N} \sum_{x_k=1}^{C_k} \mathrm{E}[p_{Y_k}(y_k) \mid \{\alpha^*_{Y_k}(y_k)\}] \mathrm{Var}[p_{Y_k}(y_k) \mid \{\alpha_{Y_k}(y_k)\})]. \qquad (9)$$

ここで，

$$\mathrm{E}[p_{Y_k}(y_k) \mid \{\alpha^*_{Y_k}(y_k)\}] = \frac{\alpha^*_{Y_k}(y_k)}{\alpha^*(+)}$$

$$\mathrm{Var}[p_{Y_k}(y_k) \mid \{\alpha^*_{Y_k}(y_k)\}] = \frac{[\alpha^*_{Y_k}(y_k)][\alpha^*(+) - \alpha^*_{Y_k}(y_k)]}{[\alpha^*(+)]^2[\alpha^*(+) + 1]}.$$

このとき，近似分布のパラメータ $\{\alpha^*_Z(z)\}$ は，

$$\mathrm{E}[p_{Y_k}(y_k) \mid X = \tilde{x}] = \mathrm{E}[p_{Y_k}(y_k) \mid \{\alpha^*(y_k)\}]$$

および

$$\mathrm{EV}_{Mixture} = \mathrm{EV}_{Single}$$

を満足するように決定されるので，近似分布の平均分散和 (9) は

$$\mathrm{EV}_{Single} = \sum_{y=1}^{J}\sum_{k=1}^{N}\sum_{x_k=1}^{C_k} \frac{\mathrm{E}[p_{Y_k}(y_k) \mid X=\tilde{x}]}{\alpha^*(+)+1}[1-\mathrm{E}[p_{Y_k}(y_k) \mid X=\tilde{x}]] \tag{10}$$

によって求めることができる．これにより，近似分布のパラメータの精度 $\alpha^*(+)$ は

$$\alpha^*(+) = \sum_{y=1}^{J}\sum_{k=1}^{N}\sum_{x_k=1}^{C_k} \frac{\mathrm{E}[p_{Y_k}(y_k) \mid X=\tilde{x}]^2}{\mathrm{EV}_{Mixture}}[1-\mathrm{E}[p_{Y_k}(y_k) \mid X=\tilde{x}]] - 1. \tag{11}$$

また周辺パラメータ $\{\alpha^*_{Y_k}(y_k)\}$ $(k=1,\cdots,N)$ は

$$\alpha^*_{Y_k}(y_k) = \alpha^*(+)\,\mathrm{E}[p_{Y_k}(y_k) \mid X=\tilde{x}]. \tag{12}$$

これによって，事後分布 $\pi(\{p_Z(z)\} \mid \{\alpha^*_Z(z)\})$ が決定される．

4 パラメータ推定

観測値が逐次得られる Bayes の学習過程のもとでのモデルパラメータの推定は，以下のステップによっておこなわれる．

1. 事前分布の期待値 $\{\mathrm{E}[p_Y(y)]\}$，$\{\mathrm{E}[p_{X_k}(x_k|y)]\}$ および精度 $\alpha(+)$ を事前知識の確信度によって決定し，事前分布のハイパーパラメータ $\{\alpha_Z(z)\}$ を計算する．

$$\alpha_Z(z) = \alpha(+)\mathrm{E}[p_Y(y)]\prod_{k=1}^{N}\mathrm{E}[p_{X_k}(x_k|y)].$$

2. 観測値が得られたとき，これの周辺事後分布 (6) の期待値 (7) および平均分散和 (8) を計算する．

3. 精度 $\alpha^*(+)$ および周辺パラメータ $\{\alpha^*_{Y_k}(y_k)\}$ を式 (11)，(12) より計算し，事後分布 $\pi(\{p_Z(z)\} \mid \{\alpha^*_Z(z)\})$ を求める．

Step 2，3 を観測値が得られるたびに繰り返すことで，事後分布のパラメータ $\{\alpha^*_Z(z)\}$ は逐次更新されていく．観測値がすべて得られたとき，その事後分布の期待値 $\{\mathrm{E}[p_Y(y) \mid \{\alpha^*_Y(y)\}]\}$，$\{\mathrm{E}[p_{X_k}(x_k|y) \mid \{\alpha^*_{Y_k}(y_k)\}]\}$ がモデルパラメータの推定値である．

5 数値シュミレーション

Bayes の逐次学習によるパラメータ推定法の有効性を検証するために，数値シュミレーションをおこなう．

顕在変数 $X=(X_1,X_2,X_3,X_4)$，潜在変数 Y からなり，それらがすべて 2 値変数であるような潜在クラスモデルを仮定する．また，この数値シュミレーションで用いる事前分布のハイパーパラメータ $\{\alpha_Z(z)\}$ は次のような事前知識のもとで決定されたものであり，それらを Case 1，Case2，Case 3 と呼ぶことにする．

Case 1: モデルパラメータ $\{p_{X_k}(x_k|y)\}$ のみ十分な事前知識をもつ

Case 2: モデルパラメータすべてに十分な事前知識をもつ場合

Case 3: モデルパラメータにすべてに全く事前知識を持たない

ここでは，事前期待値を 0.5 にすることで，モデルパラメータについて全く事前知識がないということを表現する．また，事前分布のパラメータの精度 $\alpha(+)$ を変えて数値シュミレーションをおこない，それが推定値に及ぼす影響を調べる．その際，推定値と真の確率の近さ測る尺度として Kullback-Leibler 情報量を用いる．

ここでは，観測データは多項分布 $\mathrm{Mu}(1,\{q_X(x)\})$ にしたがう多項分布モデルを考える．ここで，$\{q_X(x)\}$ は，表 2 に示された真の確率から

$$q_X(x)=\sum_{y=1}^{2} q_Y(y)q_{X_k}(x_k|y)$$

によって計算される (表 3)．このシュミレーションでは，この多項分布にしたがう乱数を 1000 個発生させる．

乱数より発生させた観測値をもとにモデルパラメータを推定値を求めるシュミレーションを 100 回行い，その平均値 $\{\bar{p}_Y(y)\}$，$\{\bar{p}_{X_k}(x_k|y)\}$ と標準偏差 (表中では括弧内の数値) を求める．それをまとめたものが 表 4，5，6 である．

表 4 は，Case 1 の推定値の平均値およびその標準偏差である．この表より，$\{p_{X_k}(x_k|y)\}$ の推定値は，真の確率にほぼ収束していることがわかる．また，$\{p_Y(y)\}$ の推定値 に関しては，事前期待値を 0.5 に設定したにもかかわらず，真の確率に近い値に収束している．このことは，$\{p_Y(y)\}$ について事前知識がなくても，$\{p(i_{X_k|Y})\}$ の事前分布のハイパーパラメータをうまく選べば，この推定法は適用可能であることを示唆している．

表 5 より，Case 2 における推定値は，真の確率に収束していることがわかる．また，$I_{1.0}(\bar{p};q)=0.0057$ に対して $I_{10.0}(\bar{p};q)=0.0008$ になっていることからわかるように，$\alpha(+)=10.0$ の推定値がより真の確率に近いところで収束している．これは，精度が推定に大きく反映されることを意味し，事前知識の確信度の強さを事前分布にそれとして与えることができる Bayes 的アプローチによる推定法の有効性を示している．

Case 3 の推定に関しては，表 6 からわかるように，Bayes の逐次学習がうまくおこなわれていない．特に，$\{p_Y(y)\}$ の推定値は事前期待値から全く更新されていない．このことは，Case 3 の状況において提案した推定法は適用できないことを示している．しかしながら，表 7 より $\{p_X(x)\}$ の推定値については，真の確率 $\{q_X(x)\}$ にむかって収束していることから，$\{p_X(x)\}$ に関する Bayes 学習がおこなわれているといえるかもしれない．

6 おわりに

モデルパラメータに関する事前知識がある場合，それを事前分布のパラメータで表現できる Bayes の逐次学習によるパラメータ推定法は大変有効であることが数値シュミレーションの結果より確かめられた．また，標準偏差の数値より，得られる推定が安定していることを確かめることができる．

仮定した潜在クラスモデルが同定可能でない場合，モデルパラメータに何らかの制約を与えて推定をおこなうが，Bayes 的アプローチでは，事前分布を設定するので，この問題は解消できる．また，顕在変数 $X=(X_1,X_2)$，潜在変数 Y で，そのカテゴリー数が $C_1=C_2=3$，$J=2$ という通常では同定できないモデルについも，Bayes 的アプローチによる方法を用い

ればパラメータの推定をおこなうことができる [4], [8]. このように，提案した Bayes の逐次学習による推定法は，従来では同定できなかったモデルのパラメータ推定に適用でき，かつ研究者の知識を積極的に活用できる点でも有効な手法といえる.

参考文献

[1] Cowell, R.G., Dawid, A. P. and Sebastiani, P. (1996). A comparision of sequential learning methods for incomplete data. *BAYESAIN STATISTICS 5* (J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds), Oxford University Press, 533–541.

[2] Dawid, A. P. and Lauritzen, S. L.(1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21**, 1272–1317.

[3] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc.* **B 39**, 1–38.

[4] Evans, M. J., Gilula, Z. and Guttman, I. (1989). Latent class analysis of two-way contingency tables by Bayesian methods. *Biometrika* **76**, 557–563.

[5] Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–231.

[6] Goodman, L. A. (1979). On the estimation of parameters in latent structure analysis. *Psychometrika* **44**, 123–128.

[7] Harberman, S. J. (1974). Log-linear models for frequency tables by indirect observation: Maximum likelihood equations. *Ann. Statist.* **2**, 1124–1147.

[8] Kuroda, M., Niki, N. and Geng, Z. (1998). Parameter Estimation of Latent Class Models by Bayesian Learning. *in preparation.*

[9] Spiegelhalter, D. J. and Lauritzen, S. L. (1990). Sequential updating of conditional probabilities on directed graphical structures. *Networks* **20**, 579–605.

[10] Tanner, M.A. (1996). *Tools for statistical inference*, (3rd edition) Springer Series in Statistics, Springer-Verlag.

表 1: 事前分布の期待値 $\{\mathrm{E}[p_Y(y)]\}$, $\{\mathrm{E}[p_{X_k}(x_k|y)]\}$

		Case 1		Case 2		Case 3	
		$y=1$	$y=2$	$y=1$	$y=2$	$y=1$	$y=2$
$\mathrm{E}[p_{X_1}(x_1|y)]$	$x_1=1$	0.100	0.600	0.100	0.600	0.500	0.500
$\mathrm{E}[p_{X_2}(x_2|y)]$	$x_2=1$	0.500	0.050	0.700	0.250	0.500	0.500
$\mathrm{E}[p_{X_3}(x_3|y)]$	$x_3=1$	0.020	0.650	0.020	0.650	0.500	0.500
$\mathrm{E}[p_{X_4}(x_4|y)]$	$x_4=1$	0.250	0.800	0.450	0.950	0.500	0.500
$\mathrm{E}[p_Y(y)]$		0.500	0.500	0.300	0.700	0.500	0.500

表 2: 真の確率 $\{q_Y(y)\}$, $\{q_{X_k}(x_k|y)\}$

		$y=1$	$y=2$
$q_{X_1}(x_1|y)$	$x_1=1$	0.200	0.700
$q_{X_2}(x_2|y)$	$x_2=1$	0.600	0.150
$q_{X_3}(x_3|y)$	$x_3=1$	0.050	0.750
$q_{X_4}(x_4|y)$	$x_4=1$	0.350	0.900
$q_Y(y)$		0.200	0.800

表 3: 真の確率の周辺確率 $\{q_X(x)\}$

		$x_4=1$		$x_4=2$	
		$x_3=1$	$x_3=2$	$x_3=1$	$x_3=2$
$x_1=1$	$x_2=1$	0.057	0.027	0.007	0.017
	$x_2=2$	0.026	0.040	0.006	0.060
$x_1=2$	$x_2=1$	0.322	0.112	0.036	0.022
	$x_2=2$	0.139	0.067	0.017	0.045

表 4: Case 1 の推定値の平均および標準偏差

		$\alpha(+) = 1.0$		$\alpha(+) = 10.0$	
		$y=1$	$y=2$	$y=1$	$y=2$
$\bar{p}_X(x_1\|y)$	$x_1=1$	0.223	0.704	0.221	0.705
		(0.0056)	(0.0005)	(0.0024)	(0.0003)
$\bar{p}_X(x_2\|y)$	$x_2=1$	0.581	0.142	0.573	0.143
		(0.0043)	(0.0003)	(0.0020)	(0.0002)
$\bar{p}_X(x_3\|y)$	$x_3=1$	0.086	0.758	0.065	0.763
		(0.0031)	(0.0010)	(0.0011)	(0.0003)
$\bar{p}_X(x_4\|y)$	$x_4=1$	0.371	0.905	0.378	0.904
		(0.0060)	(0.0003)	(0.0020)	(0.0005)
$\bar{p}_Y(y)$		0.221	0.779	0.222	0.778
		(0.0021)	(0.0021)	(0.0005)	(0.0005)
$I_{\alpha(+)}(\bar{p};q)$		0.0078		0.0064	

表 5: Case 2 の推定値の平均および標準偏差

		$\alpha(+) = 1.0$		$\alpha(+) = 10.0$	
		$y=1$	$y=2$	$y=1$	$y=2$
$\bar{p}_X(x_1\|y)$	$x_1=1$	0.178	0.691	0.192	0.696
		(0.0049)	(0.0004)	(0.0026)	(0.0003)
$\bar{p}_X(x_2\|y)$	$x_2=1$	0.626	0.154	0.608	0.151
		(0.0052)	(0.0002)	(0.0025)	(0.0002)
$\bar{p}_X(x_3\|y)$	$x_3=1$	0.043	0.735	0.041	0.745
		(0.0014)	(0.0007)	(0.0007)	(0.0003)
$\bar{p}_X(x_4\|y)$	$x_4=1$	0.313	0.893	0.339	0.897
		(0.0053)	(0.0002)	(0.0020)	(0.0001)
$\bar{p}_Y(y)$		0.181	0.819	0.193	0.807
		(0.0012)	(0.0012)	(0.0003)	(0.0003)
$I_{\alpha(+)}(\bar{p};q)$		0.0057		0.0008	

表 6: Case 3 の推定値の平均および標準偏差

		$\alpha(+)=1.0$		$\alpha(+)=10.0$	
		$y=1$	$y=2$	$y=1$	$y=2$
$\bar{p}_X(x_1\|y)$	$x_1=1$	0.601	0.601	0.600	0.600
		(0.0005)	(0.0005)	(0.0005)	(0.0005)
$\bar{p}_X(x_2\|y)$	$x_2=1$	0.236	0.236	0.236	0.236
		(0.0004)	(0.0004)	(0.0004)	(0.0004)
$\bar{p}_X(x_3\|y)$	$x_3=1$	0.608	0.608	0.608	0.608
		(0.0005)	(0.0005)	(0.0004)	(0.0004)
$\bar{p}_X(x_4\|y)$	$x_4=1$	0.791	0.791	0.791	0.791
		(0.0004)	(0.0004)	(0.0003)	(0.0003)
$\bar{p}_Y(y)$		0.500	0.500	0.500	0.500
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
$I_{\alpha(+)}(\bar{p};q)$		1.8614		1.8614	

表 7: Case 3 の周辺パラメータ $\{p_X(x)\}$ の推定値 ($\alpha(+)=1.0$)

		$x_4=1$		$x_4=2$	
		$x_3=1$	$x_3=2$	$x_3=1$	$x_3=2$
$x_1=1$	$x_2=1$	0.068	0.044	0.018	0.012
		(0.057)	(0.027)	(0.007)	(0.017)
	$x_2=2$	0.045	0.029	0.012	0.008
		(0.026)	(0.040)	(0.006)	(0.060)
$x_1=2$	$x_2=1$	0.220	0.142	0.058	0.038
		(0.322)	(0.112)	(0.036)	(0.022)
	$x_2=2$	0.147	0.095	0.039	0.025
		(0.139)	(0.067)	(0.017)	(0.045)

括弧内の数値は，真の確率 $\{q_X(x)\}$

Bayesian Parameter Estimation for Latent Class Models

Masahiro Kuroda

Dept. of Computer Science and Mathematics,
Kurashiki University of Science and the Arts,
2640 Nishinoura, Tsurajima-cho, Kurashiki-shi,
Okayama, 712-8505, Japan
(Received September 30,1998)

Abstract

Latent class analysis usually suffers from unidentifiability problems. These problem can be overcome by taking the Bayesian approach where a prior distribution of the latent parameters is given. We assume that the prior distribution is a Dirichlet distribution and is strong Markov which is sometimes called the hyper Dirichlet law. Given one observation then the prior distribution should be modified to be the posterior distribution which is given as a mixture of hyper Dirichlet distributions, since the values of the latent variables are always missing. Mixture representations might be mathematically exact but too complicated to express the current status of a Bayesian learning process and also to be used in numerical calculations. We find the approximation of a single hyper Dirichlet to the true mixture posterior distribution works in a practical sense if their means and variances match each other. The posterior latent parameters are estimated from this approximate distribution after an iterative sequence of learning actions.