情報論的学習理論テクニカルレポート 2009
Technical Report on Information-Based Induction Sciences 2009 (IBIS2009)

変分ベイズ法を用いた混合ベルヌーイ分布学習の相図について

Phase Diagram Study on Variational Bayes Learning of Bernoulli Mixture

梶 大介*† 渡辺 澄夫‡
Daisuke Kaji Sumio Watanabe

Abstract: Variational Bayes learning is widely used in statistical models that contain hidden variables, for example, normal mixtures, binomial mixtures, and hidden Markov models. Although it is reported that variational Bayes learning of mixtute model has a phase transition structure which depends on hyperparameters of mixture ratio, the detail behavior and relation of hyperparameters concerning the phase transition have not yet known. In the present paper, we experimentally investigate the phase diagram concerning the hyperparameters by using the Bernoulli mixture and show the guidance to set the hyperparameters.

Keywords: variational Bayes, phase transition, phase diagram, hyperparameter, Bernoulli mixture

1 まえがき

平均場近似を用いることで，事後分布の計算を EM アルゴリズムと同程度にすることが可能になる変分ベイズ法は，混合分布モデルや隠れマルコフモデルなどの隠れ変数をもつ確率モデルに適用され，音声認識や画像処理，遺伝子解析など様々な分野でその有効性が示されている [3, 8]．変分ベイズ法を用いた指数型分布族の混合分布の学習ではハイパーパラメータにより，事後分布が確率モデルのすべてのコンポーネントを使って表現する場合と冗長なコンポーネントを用いず，より少ないコンポーネントで表現する場合に分かれる ”相転移 ”が起こることが分かっている [1, 12].

本稿では，混合ベルヌーイ分布を用いて混合比とベルヌーイ分布のハイパーパラメータを変化させた場合の変分ベイズ法により得られる予測分布の変化を実験的に調べ，その相図を示す．相図から，相転移点と混合比/ベルヌーイ分布の両ハイパーパラメータの依存関係やハイパーパラメータに関する設定指針などが得られる．

*東京工業大学大学院総合理工学研究科, 226-8503 神奈川県横浜市緑区長津田 4259, tel. 045-924-5018, e-mail kaji@cs.pi.titech.ac.jp
Tokyo Inistitute of Technology, 4259 Nagatsuda, Midori-ku,Yokohama, 226-8503 Japan
†コニカミノルタエムジー (株), 192-8505 東京都八王子市石川町 2970, tel. 042-660-8157, e-mail daisuke.kaji@konicaminolta.jp
Konicaminolta Medical & Graphic,INC, 2970 Ishikawa-machi, Hachioji-shi,Tokyo, 192-8505, Japan
‡東京工業大学 精密工学研究所, 226-8503 神奈川県横浜市緑区長津田 4259, e-mail swatanab@cs.pi.titech.ac.jp,
PI Lab., Tokyo Institute of Technology, 4259 Nagatsuta Midori-ku, Yokohama, 226-8503, Japan

2 混合ベルヌーイ分布

混合ベルヌーイ分布は潜在クラス解析に用いられる分布として知られており，2 値データのクラスタリングやレコメンデーションシステムなどの応用に広く用いられている [3, 10]．さらにベルヌーイ分布の事前分布が後述するように 1 つのハイパーパラメータであるため，実験・解析が容易である．以上のような背景から本稿では，変分ベイズ法の相転移構造の検討に混合ベルヌーイ分布を用いる．ベルヌーイ分布の確率密度関数は以下で与えられる．

$$B(\boldsymbol{x}|\boldsymbol{\mu}) = \prod_{i=1}^{M} \mu^{x_i}(1-\mu)^{(1-x_i)},$$

ここで $\boldsymbol{x} = (x_1, \cdots, x_M)^T$ はデータ，$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_M)^T$，$M$ はそれぞれパラメータとデータの次元である．このとき，混合ベルヌーイ分布は

$$p(\boldsymbol{x}|\boldsymbol{\pi}, \boldsymbol{\mu}) = \sum_{k=1}^{K} \pi_k B(\boldsymbol{x}|\boldsymbol{\mu}_k),$$

で定義される．ただし，$\boldsymbol{\pi}$ は $B(\boldsymbol{x}|\boldsymbol{\mu}_k)$ の混合比を表し，K はコンポーネント数を表す．次にデータ $\boldsymbol{x}$ に対する隠

れ変数を導入する．隠れ変数 $\boldsymbol{z}$ はデータ $\boldsymbol{x}$ がどの分布から発生したかを示す競合的ベクトル $\boldsymbol{z} = (0, \cdots, 1, \cdots, 0)$ として表される．このとき，隠れ変数 $\boldsymbol{z}$ およびデータ $\boldsymbol{x}$ の分布の事前分布として共役事前分布であるディリクレ分布とベータ分布を用いると $\boldsymbol{Z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N), \boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N), \boldsymbol{\pi}$ および $\boldsymbol{\theta}$ の分布はそれぞれ

$$p(\boldsymbol{Z}|\boldsymbol{\pi}) = \prod_{n=1}^{N}\prod_{k=1}^{K}\pi_k^{z_{nk}},$$

$$p(\boldsymbol{X}|\boldsymbol{Z},\boldsymbol{\theta}) = \prod_{n=1}^{N}\prod_{k=1}^{K}\left(\prod_{m}^{M}\theta_{km}^{x_{nm}}(1-\theta_{km})^{(1-x_{nm})}\right)^{z_{nk}},$$

$$p(\boldsymbol{\pi}) = \frac{\Gamma(Ka)}{\Gamma(a)^K}\prod_{k}^{K}\pi_k^{a-1},$$

$$p(\boldsymbol{\theta}) = \prod_{k=1}^{K}\prod_{m=1}^{M}\left(\frac{\Gamma(2b)}{\Gamma(b)^2}\theta_{km}^{b-1}(1-\theta_{km})^{b-1}\right)$$

で与えられる．(a, b) は事前分布 $p(\boldsymbol{\pi})$，$p(\boldsymbol{\theta})$ のパラメータでハイパーパラメータと呼ばれる．

3 変分ベイズ法

3.1 変分ベイズ法の一般式の導出

本節では，$\boldsymbol{Y}$ によりパラメータを含むすべての隠れ変数，$\boldsymbol{X}$ ですべての観察可能な変数を表す．このとき，任意の確率分布 $q(\boldsymbol{Y})$ と事後分布 $p(\boldsymbol{Y}|\boldsymbol{X})$ に対して次式が成り立つ．

$$F(\boldsymbol{X}) = \bar{F}[q(\boldsymbol{Y})] + KL(q(\boldsymbol{Y})\|p(\boldsymbol{Y}|\boldsymbol{X})),$$

ここで自由エネルギー F, 変分自由エネルギー $\bar{F}$ および Kullback-Leibler 距離 KL は以下で表される．

$$F(\boldsymbol{X}) = -\log\int p(\boldsymbol{X},\boldsymbol{Y})d\boldsymbol{Y} = -\log p(\boldsymbol{X}),$$

$$\bar{F}[q(\boldsymbol{Y})] = \int q(\boldsymbol{Y})\log\frac{q(\boldsymbol{Y})}{p(\boldsymbol{X},\boldsymbol{Y})}d\boldsymbol{Y},$$

$$KL(q(\boldsymbol{Y})\|p(\boldsymbol{Y}|\boldsymbol{X})) = \int q(\boldsymbol{Y})\log\frac{q(\boldsymbol{Y})}{p(\boldsymbol{Y}|\boldsymbol{X}))}d\boldsymbol{Y}.$$

変分事後分布 $q(\boldsymbol{Y})$ は $\bar{F}[q(\boldsymbol{Y})]$ の最小化によって与えられ，これは $q(\boldsymbol{Y})$ と真の事後分布 $p(\boldsymbol{Y}|\boldsymbol{X})$ の Kullback-Leibler 距離の最小化と同値である．ここで変分ベイズ法は事後分布の計算困難性を回避するため，パラメータと隠れ変数が条件付独立であることを仮定する．したがって $\boldsymbol{w}$ をパラメータとするとき，$q(\boldsymbol{Y})$ は

$$q(\boldsymbol{Y}|\boldsymbol{X}) = q(\boldsymbol{Z},\boldsymbol{w}|\boldsymbol{X}) = q_1(\boldsymbol{Z}|\boldsymbol{X})q_2(\boldsymbol{w}|\boldsymbol{X})$$

と表せる．

上記の汎関数 $\bar{F}[q(\boldsymbol{Y}|\boldsymbol{X})]$ の q_1 ，q_2 に関する最小化は変分法により実行される．最小化問題を $\sum_{\boldsymbol{Z}} q_1(\boldsymbol{Z}|\boldsymbol{X}) = 1$, $\int q_2(\boldsymbol{w}|\boldsymbol{X})d\boldsymbol{w} = 1$ の制限のもと解くことで次式を得る．

$$\log q_1(\boldsymbol{Z}|\boldsymbol{X}) = E_{q_2}[\log P(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{w})] + C_1, \quad (1)$$

$$\log q_2(\boldsymbol{w}|\boldsymbol{X}) = E_{q_1}[\log P(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{w})] + C_2, \quad (2)$$

ここで C_1, C_2 は正規化定数である．上式は互いの分布による平均値の計算を含んでいる．したがって，変分ベイズ法による学習は (1) および (2) の逐次的繰り返し演算により実行される．

3.2 混合ベルヌーイ分布の変分ベイズ法

前節で導いた変分ベイズ法の一般更新式 (1)，(2 を 2) 節の設定のもとで計算することで以下の混合ベルヌーイ分布の変分ベイズ学習アルゴリズムを得る．

VB e-step

$$\log\rho_{nk} = \psi(\alpha_k) - \psi\left(\sum_{k}^{K}\alpha_k\right) + \sum_{m=1}^{M}G(\eta_{km},\eta'_{km})$$

$$r_{nk} = \frac{\rho_{nk}}{\sum_{k=1}^{K}\rho_{nk}}$$

VB m-step

$$N_k = \sum_{n=1}^{N}r_{nk}, \quad a_k = a + N_k$$

$$\eta_{km} = b + \sum_{n=1}^{N}r_{nk}x_{nm}, \quad \eta'_{km} = b + \sum_{n=1}^{N}r_{nk}(1-x_{nm})$$

ここで

$$G(\eta_{km},\eta'_{km}) = x_{nm}\psi(\eta_{km}) - x_{nm}\psi(\eta'_{km}) + \psi(\eta'_{km}) - \psi(\eta_{km}+\eta'_{km})$$

とした．ただし，ψ はディガンマ関数と呼ばれ $\psi(a) \equiv \frac{d}{da}\log\Gamma(a) = \frac{\Gamma'(a)}{\Gamma(a)}$ である．上記のアルゴリズムを実行することで事後分布

$$q_1(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}),$$

$$q_2(\boldsymbol{\theta}) = \prod_{k=1}^{K}\prod_{m=1}^{M}\mathrm{Beta}(\theta_{km}|\eta_{km},\eta'_{km})$$

が得られる．

4　変分ベイズ学習の相転移

変分ベイズ学習の変分自由エネルギーについては，次の定理によりその上界と下界が与えられている．

[**Theorem1:K.Watanabe,S.Watanabe**]

真の分布の混合数を K_0，予測モデルの混合数を K とするとき，変分自由エネルギー $\hat{F}_n$ は次の不等式を満たす．

$$\lambda_1 \log n + nK_n(\hat{w}) + c_1 < \hat{F}_n - S_n < \lambda_2 \log n + c_2$$

ここで S_n は真の分布の経験エントロピー，$K_n(\hat{w})$ は変分ベイズ法が定めるパラメータ $\hat{w}$ に対する経験カルバック距離，c_1, c_2 は定数である．また，定数 λ_1 と λ_2 は次で与えられる．ここで $M^* = \frac{M+1}{2}$ とおく．

$$\lambda_1 = \begin{cases} (K-1)a + \frac{M}{2}, & (a \leq M^*) \\ \frac{MK+K-1}{2}, & (a > M^*) \end{cases}$$

$$\lambda_2 = \begin{cases} (K-K_0)a + \frac{MK_0+K_0-1}{2}, & (a \leq M^*) \\ \frac{MK+K-1}{2}, & (a > M^*) \end{cases}$$

ここで a は前述の混合比側のハイパーパラメータである．

上記の定理は変分自由エネルギーの挙動が $a = M^*$ を境に大きく変化することを示唆している．具体的には，上述の変分自由エネルギーの下界を実現するためには，$a \leq \quad M^* \quad$ では，”事後分布はサンプルが十分に得ることができないような混合分布に対する混合比パラメータを 0 に近づける”必要があるのに対して，$a > M^*$ では”冗長なパラメータを均等に使い混合分布を実現する”必要がある．したがって，$a = M^*$ 周辺で事後分布の挙動が大きく変化，すなわち相転移が起こることを示していると考えられる [1].

また，汎化誤差については，混合正規分布の場合，相転移点の前後で大きく変化するが相転移点ではその変動は小さく，安定することが示されている [12].

以下では混合ベルヌーイ分布において，真の分布から発生するサンプルをその平均値で代用することで，変分ベイズアルゴリズムを力学系と見なし，混合比および混合分布両方のハイパーパラメータが事後分布に与える影響について調査した．

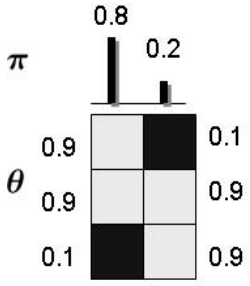
5　実験

5.1　変分ベイズ学習の力学系

ベルヌーイ分布側の次元が $M = 3$ である真の分布を以下で与える．

$$p^*(\boldsymbol{x}) = 0.8 \cdot \left(0.9^{\boldsymbol{x}_1} \cdot 0.1^{1-\boldsymbol{x}_1}\right) + 0.2 \cdot \left(0.1^{\boldsymbol{x}_2} \cdot 0.9^{1-\boldsymbol{x}_2}\right)$$

この分布をすべてのサンプルの発生確率とともに示したのが図 1 である．



Type	Image			Prob.
S_1 :(1,1,1)				0.081
S_2 :(1,1,0)			■	0.585
S_3 :(1,0,0)		■	■	0.065
S_4 :(1,0,1)		■		0.009
S_5 :(0,1,1)	■			0.153
S_6 :(0,1,0)	■		■	0.081
S_7 :(0,0,1)	■	■		0.017
S_8 :(0,0,0)	■	■	■	0.009

図 1: 真の分布 (左) と各サンプルの発生確率．真の分布は白いほど高い確率を表す．また，サンプルについては白が 1，黒が 0 を表す．

ここで真の分布について，上部の棒グラフがその混合比を表し，その下にベルヌーイ分布のパラメータ，すなわち各分布における $\boldsymbol{x}_i (i = 1, 2)$ の発生頻度をグレースケールで表した．ここで白いほど発生確率が高いものとしている．上述の定理から，この分布は $a = \frac{3+1}{2} = 2$ に相転移点をもつと考えられる．また，真の分布から発生するデータは，図 1 右の表にある 8 つのパターンで，それぞれの確率を表の最右列に示した．

以下の実験では，真の分布からサンプルを直接発生させるのではなく，総サンプル数 N に対して各サンプルの確率の比でそれぞれのデータが発生すると考える．この場合，サンプルによる揺らぎを考慮する必要がなくなるため，変分ベイズ学習のアルゴリズムはハイパーパラメータをもつ力学系と見なすことができる．なお，以下の実験では $N = 10000$ としている．

前述の 8 種類のサンプル $S_i (i = 1, \cdots, 8)$ の第 t 成分 $S_i^{(t)} (t = 1, 2, 3)$，S_1〜S_8 のそれぞれの発生確率を P_1〜P_8 として変分ベイズアルゴリズムを書き換えた力学系は以下にようになる．

VB e-step

$$
\begin{aligned}
\log \rho_{S_i k} &= \Psi_\alpha(k) \\
&+ S_i^{(1)}\Psi_1(k) + (1-S_i^{(1)})\Psi_1'(k) + \Psi_2'(k) \\
&+ S_i^{(2)}\Psi_1(k) + (1-S_i^{(2)})\Psi_1'(k) + \Psi_2'(k) \\
&+ S_i^{(3)}\Psi_1(k) + (1-S_i^{(3)})\Psi_1'(k) + \Psi_2'(k) \\
r_{S_i k} &= \frac{\rho_{S_i k}}{\sum_{k=1}^{4}\rho_{S_i k}}
\end{aligned}
$$

VB m-step

$$
\begin{aligned}
N_k &= \sum_{i=1}^{4} NP_i r_{S_i k}, \quad a_k = a + N_k \\
\eta_{1k} &= b + r_{S_1 k}NP_1 + r_{S_2 k}NP_2 + r_{S_3 k}NP_3 + r_{S_4 k}NP_4 \\
\eta_{2k} &= b + r_{S_1 k}NP_1 + r_{S_2 k}NP_2 + r_{S_5 k}NP_5 + r_{S_6 k}NP_6 \\
\eta_{3k} &= b + r_{S_1 k}NP_1 + r_{S_4 k}NP_4 + r_{S_5 k}NP_5 + r_{S_7 k}NP_7 \\
\eta'_{1k} &= b + r_{S_5 k}NP_5 + r_{S_6 k}NP_6 + r_{S_7 k}NP_7 + r_{S_8 k}NP_8 \\
\eta'_{2k} &= b + r_{S_3 k}NP_3 + r_{S_4 k}NP_4 + r_{S_7 k}NP_7 + r_{S_8 k}NP_8 \\
\eta'_{3k} &= b + r_{S_2 k}NP_2 + r_{S_3 k}NP_3 + r_{S_6 k}NP_6 + r_{S_8 k}NP_8
\end{aligned}
$$

ここで

$$
\begin{aligned}
\Psi_\alpha(k) &= \psi(\alpha_k) - \psi\left(\sum_k^K \alpha_k\right), \\
\Psi_1(k) &= \psi(\eta_{k1}) - \psi(\eta'_{k1}) + \psi(\eta'_{k1}) - \psi(\eta_{k1}+\eta'_{k1}), \\
\Psi_2(k) &= \psi(\eta_{k2}) - \psi(\eta'_{k2}) + \psi(\eta'_{k2}) - \psi(\eta_{k2}+\eta'_{k2}), \\
\Psi'_1(k) &= \psi(\eta'_{k1}) - \psi(\eta_{k1}+\eta'_{k1}), \\
\Psi'_2(k) &= \psi(\eta'_{k2}) - \psi(\eta_{k2}+\eta'_{k2})
\end{aligned}
$$

とした.

5.2 実験結果

学習モデルの混合分布数を $K=4$ として，上記のアルゴリズムにより学習を行った結果が図 2 である．ここで横軸，縦軸はそれぞれハイパーパラメータ a, b であり，いずれも 0.001〜10 まで変化させている (log スケールで表示)．また，図のグレースケールは学習結果の混合比 (平均パラメータ) を大きい順に並び替えた $\pi_1, \cdots, \pi_4$ に対して $z = |\pi_1 - 0.8| + |\pi_2 - 0.2|$ を算出したものであり，z が 0 に近い (黒い) ほど冗長な分布を含まず，混合比も含めて真の分布に近い学習結果と考えることができる．この結果から上述の定理が示唆するように，冗長な表現への切り替え（相転移）は $a = M^* = \frac{3+1}{2} = 2$ の前後で発生しているが，その値は b に依存していることがわかる．

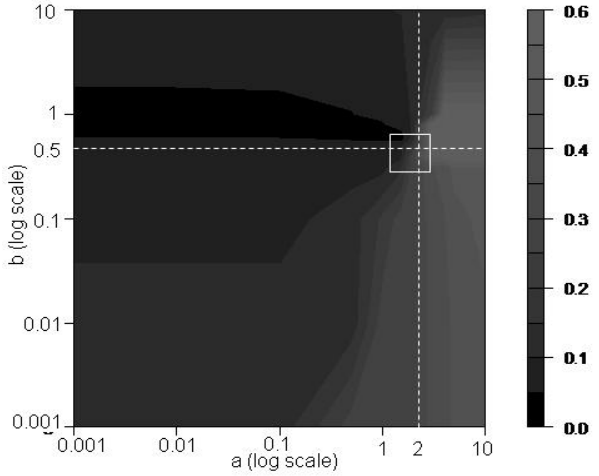


図 2: ハイパーパラメータと混合比の関係．横軸は a，縦軸は b であり，濃淡は $z = |\pi_1 - 0.8| + |\pi_2 - 0.2|$ の値を表す.

この様子をさらに図 2 中の四角で囲った領域で拡大したものが図 3 上段右の図である．この学習結果は大きく以下の 3 つの種類に分類することががができる．

- 領域 A:コンポーネント数を絞り込み，2 つの混合分布で表現する..
- 領域 B:A から C への移行過程.
- 領域 C:すべてのコンポーネントを用いて分布を表現する.

この分類にしたがって領域を分け，相図を作成したものが上段左の図である．また，それぞれの領域での学習結果に対する予測分布 (平均パラメータによる分布) を下段に示した．この図から，冗長なコンポーネントを除き，より少ない混合分布数で学習結果を表現する場合には a を小さくし，b を 0.5 より大きくとると良いことがわかる．特に $b = 0.5$ から $b = 1$ のときにより冗長な項の混合比が一番低くなっている．また，a が小さい場合でも b を小さく設定すると，コンポーネント数が増える傾向にある．これはベルヌーイ分布側の確率が 1 または 0 近づくような事前分布を与えるハイパーパラメータを設定することで，小さなカテゴリを検出しやすくなるためと考えられる．このようなハイパーパラメータの設定はアンケートやマーケティング解析などの ”少数意見の抽出 ”に応用することができる [11].
相図からは，さらにハイパーパラメータを変えた際の予測分布の変化の様子の違いを読み取ることもできる．すなわち，$a > 2.0$ の領域で b を大きくした場合，B のような移行過程領域から A の冗長性のない分布の領域に

向かう途中で，C の冗長な表現をする領域を $b = 0.5$ 付近で通過することになる．一方，$a < 2.0$ のような領域では領域 C を跨らずに，直接的に領域 A に向かうことになる．

図 4 はベルヌーイ分布の次元を $M = 2$ とした場合の真の分布 (左) と実験結果である．

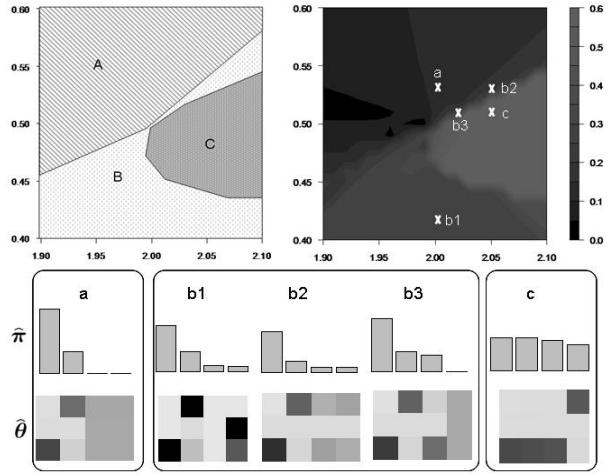


図 3: 相転移の領域 (上段:左), ハイパーパラメータと混合比の関係拡大図 (上段:右), 各領域での平均パラメータによる学習結果 (下段)

この場合も相転移点や前述の領域 A,B,C の位置関係は大きくは変わらず，相図としてはほぼ同じものが得られる．

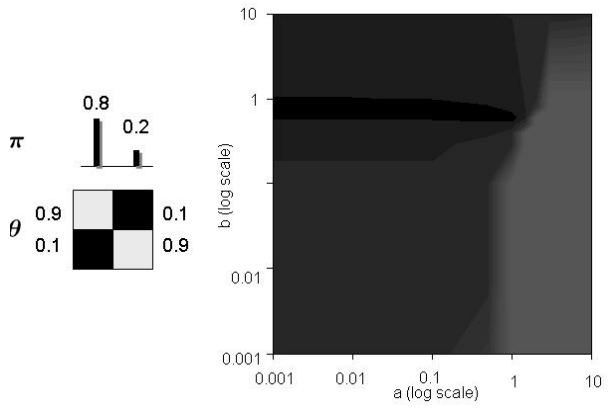


図 4: $M = 2$ での真の分布 (左) とハイパーパラメータと混合比の関係 (右)

これらの結果から設定したハイパーパラメータ a, b を変更することで抽出するクラスタの粒度やコンポーネント使い方，すなわち，すべてのコンポーネントを使用するか/コンポーネントの絞込みを行うかを調整できることがわかった．応用の観点では，これらの相図は混合ベルヌーイ分布をクラスタリングのツールとして用いる場合，目的とする分類粒度に応じて，どのようなハイパーパラメータを設定すべきかの方針を与える図になっていると考えられる．

6 おわりに

変分ベイズ法を用いた混合ベルヌーイ分布の学習におけるハイパーパラメータと学習結果の関係を調べ，$M = 2, M = 3$ の場合の相図を示した．相図は相転移点での挙動に関する多くの情報を与えるだけでなく，応用の立場からもクラスタリングへ利用する際のハイパーパラメータ設定に関する指針を提供する．一方，相転移と変分自由エネルギー，汎化誤差の関係についてはまだ多くのことは分かっておらず，理論的な解明を含め今後の課題である．

参考文献

[1] K. Watanabe and S. Watanabe. Stochastic complexities of general mixture models in Variational Bayesian Approximation. *Neural Computation*, Vol. 18, No. 5, pp.1007-1065, 2006.

[2] S. Nakajima and S. Watanabe. Variational Bayes Solution of Linear Neural Networks and its Generalization Performance. *Neural Computation*, Vol. 19, No. 4, pp. 1112-1153, 2007.

[3] C. M. Bishop. Pattern Recognition and Machine Learning. *Springer*, 2006.

[4] S. Watanabe. Algebraic analysis for singular statistical estimation. *Proc. of International Journal of AlgorithmicLearningTheory Lecture Notes on Computer Sciences*,1720, pp.39-50, 1999.

[5] S. Watanabe. Algebraic Analysis for Nonidentifiable LearningMachines. *Neural Computation*, Vol.13, No.4, pp.899-933, 2001

[6] S. Watanabe. Learning efficiency of redundant neural networksin Bayesian estimation. *IEEE Transactions on NeuralNetworks* , Vol.12, No.6, pp.1475-1486, 2001.

[7] H. Attias. Inferring parameters and structure of latent variable models by variational Bayes, In *Proc. of Uncertainty in Artificial Intelligence(UAI' 99)*,1999.

[8] M. J. Beal. Variational Algorithms for approximate Bayesian inference. *PhD thesis, University College London*, 2003.

[9] Z. Ghahramani and M. J. Beal. Graphical Models and Variational Methods. In *Advanced Mean Field. Methods*. MIT Press, 2000

[10] P. F. Lazarsfeld and N. W. Henry. Latent structure analysis. *Houghton Mifflin*, 1968

[11] D. Kaji and S. Watanabe. Optimal Hyperparameters for Generalized Learning and Knowledge Discovery in Variational Bayes. *To appear in Proc. of ICONIP*, 2009

[12] 大山 慎史, 渡辺　澄夫. 変分ベイズ学習におけるハイパーパラメータの汎化誤差への影響について. 信学技報（NC 研究会）, January 2009.