

# 混合 von Mises-Fisher 分布に対する変分法的ベイズ推定 Variational Bayes Inference of Mixed von Mises-Fisher Distribution

田辺昭博\*  
Akihiro Tanabe

福水健次†  
Kenji Fukumizu

大羽成征‡  
Shigeyuki Oba

石井信§  
Shin Ishii

**Abstract:** High dimensional directional data is getting important in several applications such as analysis of text and gene expression data. In this report, we proposed variational Bayes algorithm of mixed von Mises-Fisher distribution, which is the probability density function defined on unit hypersphere. We applied this algorithm to artificial data and computed free energy, and showed that free energy can be used as a criterion to determine the number of units.

**Keywords:** variational Bayes, directional data, von Mises-Fisher distribution, free energy

## 1 はじめに

文書データや遺伝子発現データなどの高次元ベクトルデータを解析する際、ベクトル間の類似度の指標にユークリッド距離を用いると良い解析が行えない場合がある。そこで、その代わりとして相関係数を用いた解析を行うことで成功をおさめたという報告がなされている。また、遺伝子発現データの階層化クラスタリングでは、相関係数距離が経験的によいとされる。相関係数は、各ベクトルをそのノルムで正規化したもとの内積であるため、相関係数を用いるということはベクトルのノルムを無視することに相当する。このようにデータベクトルを正規化したものを指向性データ (directional data) とよぶ [3]。

指向性データの自然な生成モデルとして、単位超球面上における多変量 Gauss 分布に対応する、von Mises-Fisher 分布が考えられる。これは  $n+1$  次元ユークリッド空間上の Gauss 分布を  $n$  次元超球面上に射影した分布である。例えば Dhillon らは、von Mises-Fisher 分布の混合モデルに対する最尤推定法および EM アルゴリズムによる解法を導出し、遺伝子発現データに対するクラスタ

解析などに適用した [1],[2],[6]。

混合分布モデルのパラメトリック推定に基づくクラスタ解析は、上記の例のみならず広く行われている。ベイズ推定を用いることで、周辺化尤度がクラスタ数の決定に使える可能性がある。しかし、ベイズ推定は一般に積分計算が困難であることが多い。近似手法として計算機の能力に頼った MCMC 法が用いられることもあるが、近年、より効率的な手法としての変分ベイズ推定法が注目されている。

本研究では、混合 von Mises-Fisher 分布に対する変分ベイズ推定法を提案する。はじめに von Mises-Fisher 分布を導入し、変分ベイズ推定の基本原理を概観した後、それを混合 von Mises-Fisher 分布推定問題に適用した。また、導出されたアルゴリズムを人工データに適用し、混合数の決定について検討を行った。

## 2 von Mises-Fisher 分布

### 2.1 von Mises-Fisher 分布

$p$  次元の単位確率ベクトル  $\mathbf{x}$  ( $\|\mathbf{x}\|=1$ ) が確率密度関数

$$p(\mathbf{x}) = C_p(\boldsymbol{\kappa}) e^{\boldsymbol{\kappa}^T \mathbf{x}} \quad (1)$$

に従って分布するとき、 $\mathbf{x}$  は  $p$  変数の von Mises-Fisher 分布  $M_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\kappa})$  に従うという [3]。ただし  $T$  はベクトルの転置を意味する。ここで、パラメータ  $\boldsymbol{\mu}$  はこの分布の中心方向を定める単位ベクトル、 $\boldsymbol{\kappa}$  はその方向への分布の集中度を示すパラメータである。正規化定数  $C_p(\boldsymbol{\kappa})$  は

$$C_p(\boldsymbol{\kappa}) = \frac{\boldsymbol{\kappa}^{p/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\boldsymbol{\kappa})} \quad (2)$$

\*奈良先端科学技術大学院大学情報科学研究科, 〒630-0192 奈良県生駒市高山町 8916-5 奈良先端科学技術大学院大学 論理生命学分野, TEL 0743-72-5986, FAX 0743-72-5989, e-mail akihi-ta@is.naist.jp  
Nara Institute of Science and technology Graduate School of Information Science

†統計数理科学研究所,  
The Institute of Statistical Mathematics

‡奈良先端科学技術大学院大学情報科学研究科,  
Nara Institute of Science and technology Graduate School of Information Science

§奈良先端科学技術大学院大学情報科学研究科,  
Nara Institute of Science and technology Graduate School of Information Science

で表される。  $I_p(x)$  は次数  $p$  の第 1 種変形 Bessel 関数であり、

$$I_p(x) = \frac{1}{\pi} \int_0^\pi e^{x \cos t} \cos(pt) dt \quad (3)$$

で定義される。

## 2.2 混合 von Mises-Fisher 分布

$m$  個のユニットからなる混合 von Mises-Fisher 分布の確率密度関数は

$$p(\mathbf{x}; \theta) = \sum_{i=1}^m \alpha_i M_p(\mathbf{x}; \boldsymbol{\mu}_i, \kappa_i) \quad (4)$$

で表すことができる。ただし、 $\mathbf{x} \in S^{p-1}$ 、 $0 < \alpha_i < 1$ 、 $\sum_{i=1}^m \alpha_i = 1$  とする。  $\theta \equiv \{\alpha_i, \boldsymbol{\mu}_i, \kappa_i\}_{i=1}^m$  はパラメータ集合である。以下では  $D = \{\mathbf{x}_n\}_{n=1}^N$  で観測データの集合を表すものとし、各データ点  $\mathbf{x}_n$  は互いに独立であるとする。

(4) 式に隠れ変数  $Z$  を導入する。  $Z = \{z_i^n\}_{n=1, i=1}^m$  を、サンプル  $\mathbf{x}_n$  が第  $i$  ユニットモデルから生成されていれば  $z_i^n = 1$ 、そうでなければ  $z_i^n = 0$  と定義した隠れ変数の集合と定義すれば、完全データ集合の結合分布は

$$p(D, Z | \theta) = \prod_{i=1}^m \prod_{n=1}^N \{\alpha_i C_p(\kappa_i) e^{\kappa_i \boldsymbol{\mu}_i^T \mathbf{x}_n}\}^{z_i^n} \quad (5)$$

と書ける。両辺対数を取ると

$$\begin{aligned} \log p(D, Z | \theta) \\ = \sum_{i=1}^m \sum_{n=1}^N z_i^n \{\log \alpha_i + \log C_p(\kappa_i) + \kappa_i \boldsymbol{\mu}_i^T \mathbf{x}_n\} \end{aligned} \quad (6)$$

となる。

## 3 変分ベイズ法に関して

ベイズ推定は学習データが少ない場合に最尤推定と比較して予測分布の精度が高く、周辺化尤度基準によるモデル選択に利用できる一方で、尤度周辺化に関する積分計算が困難であるという実用上の問題がある。変分ベイズ法では、この積分計算を変分法を用いることで近似する [4]。

一般論として、観測データ集合を  $D$ 、観測データのそれぞれに対する隠れ変数からなる集合を  $Z$ 、パラメータの集合を  $\theta$  とする。

変分ベイズ法では、すべての未知量を周辺化した周辺

尤度の対数の下限を定義する。

$$\begin{aligned} \mathcal{L}(D) &= \log p(D) = \log \sum_Z \int p(D, Z, \theta) d\theta \\ &= \log \sum_Z \int q(Z, \theta) \frac{p(D, Z, \theta)}{q(Z, \theta)} d\theta \\ &\geq \sum_Z \int q(Z, \theta) \log \frac{p(D, Z, \theta)}{q(Z, \theta)} d\theta \\ &\stackrel{\text{def}}{=} \mathcal{F}[q] \end{aligned} \quad (7)$$

$\mathcal{F}[q]$  は自由エネルギーとよばれ、テスト分布とよぶ新たな分布  $q(Z, \theta)$  の汎関数である。  $q(Z, \theta)$  が真の事後分布  $p(Z, \theta | D)$  と一致するときに、 $\mathcal{F}[q]$  は対数周辺尤度  $\mathcal{L}(D)$  と一致する。変分ベイズ法ではテスト分布に独立分解近似  $q(Z, \theta) = q(Z)q(\theta)$  を仮定し、この近似のもとで  $\mathcal{F}[q]$  を  $q$  に関して最大化する。このとき  $q(Z)q(\theta)$  は独立分解近似のもとで事後分布に (KL divergence の意味で) 最も近づき、 $\mathcal{F}[q]$  は  $\mathcal{L}(D)$  に最も近づく。また、周辺尤度最大化によるモデル選択を、自由エネルギー最大化基準で近似することが可能である。これは、モデルに対する一様分布を仮定したもとのモデル事後分布最大化基準と等価である。

今、パラメータ集合  $\theta$  が  $I$  個の独立なパラメータ  $\{\theta_i\}_{i=1}^I$  からなるとすると、(7) 式は、

$$\begin{aligned} \mathcal{F}[q] &= \left\langle \log \frac{p(D, Z | \theta)}{q(Z)} \right\rangle_{q(Z), q(\theta)} \\ &\quad + \sum_{i=1}^I \left\langle \log \frac{p(\theta_i)}{q(\theta_i)} \right\rangle_{q(\theta_i)} \\ &= \langle \log p(D, Z | \theta) \rangle_{q(Z), q(\theta)} - \langle \log q(Z) \rangle_{q(Z)} \\ &\quad + \sum_{i=1}^I \left\{ \langle \log p(\theta_i) \rangle_{q(\theta_i)} - \langle \log q(\theta_i) \rangle_{q(\theta_i)} \right\} \end{aligned} \quad (8)$$

と書ける。ただし、 $p(\theta)$  はパラメータの事前分布であり、 $\langle f(x) \rangle_{q(x)}$  は  $f(x)$  の  $q(x)$  に関する期待値を表す。

(8) 式を  $q(Z)$  に関して最大化すると、 $Z$  の最適テスト分布が得られる。

$$q(Z) = C \exp \langle \log p(D, Z | \theta) \rangle_{q(\theta)} \quad (9)$$

$C$  は  $\sum_Z q(Z) = 1$  となるための正規化定数である。

$q(\theta_i)$  についても同様に最大化すると、

$$q(\theta_i) = C' p(\theta_i) \exp \langle \log p(D, Z | \theta) \rangle_{q(Z), q(\theta_{-i})} \quad (10)$$

を得る。  $C'$  は  $\int q(\theta_i) d\theta_i = 1$  となるための正規化定数であり、 $\theta_{-i}$  は  $\theta$  から  $\theta_i$  を除いた集合を表す。(9)(10) 式の繰返しアルゴリズムは VB-EM アルゴリズムとよばれ、局所解への収束が保証されている。

## 4 混合 von Mises-Fisher 分布に対する VB-EM アルゴリズムの導出

### 4.1 準備

独立分解近似にしたがい、テスト分布  $q(Z, \theta)$  を

$$q(Z, \theta) = q(Z)q(\boldsymbol{\alpha}) \prod_{i=1}^m q(\boldsymbol{\mu}_i) \quad (11)$$

と表すものとする。期待値計算を行うために、まず次の計算を行う。

$$\begin{aligned} & \langle \log p(D, Z | \theta) \rangle_{q(Z)} \\ &= \sum_{i=1}^m \sum_{n=1}^N \sum_Z q(Z) z_i^n \{ \log \alpha_i + \log C_p(\boldsymbol{\kappa}_i) + \boldsymbol{\kappa}_i \boldsymbol{\mu}_i^T \mathbf{x}_n \} \\ &= \sum_{i=1}^m [ \bar{N}_i \{ \log(\alpha_i) + \log(C_p(\boldsymbol{\kappa}_i)) + \boldsymbol{\kappa}_i \boldsymbol{\mu}_i^T \bar{\mathbf{x}}_i \} ] \end{aligned} \quad (12)$$

ここで

$$\bar{z}_i^n = \sum_Z q(Z) z_i^n, \quad (13)$$

$$\bar{N}_i = \sum_{n=1}^N \bar{z}_i^n, \quad (14)$$

$$\bar{\mathbf{x}}_i = \frac{1}{\bar{N}_i} \sum_{n=1}^N \bar{z}_i^n \mathbf{x}_n \quad (15)$$

とした。

### 4.2 テスト事後分布の計算

前章の (9)(10) 式と (12) 式を用いて、以下を計算する。ここでは簡単のため、 $\boldsymbol{\kappa}_i$  は既知とした。

$$q(\boldsymbol{\alpha}) \propto p(\boldsymbol{\alpha}) \exp \langle \log p(D, Z | \theta) \rangle_{q(Z), q(\boldsymbol{\mu})}, \quad (16)$$

$$q(\boldsymbol{\mu}_i) \propto p(\boldsymbol{\mu}_i) \exp \langle \log p(D, Z | \theta) \rangle_{q(Z), q(\boldsymbol{\alpha})}, \quad (17)$$

$$q(Z) \propto \exp \langle \log p(D, Z | \theta) \rangle_{q(\boldsymbol{\mu}), q(\boldsymbol{\alpha})} \quad (18)$$

ここで、パラメータの事前分布として

$$p(\boldsymbol{\alpha}) = \text{Dirichlet}(\{\alpha_i\}_{i=1}^m; \boldsymbol{\phi}_0), \quad (19)$$

$$p(\boldsymbol{\mu}_i) = M_p(\boldsymbol{\mu}_i; \boldsymbol{\mu}_0, \boldsymbol{\kappa}_0) \quad (20)$$

であるとする。以下で計算されるようにテスト分布  $q(\boldsymbol{\alpha})$ ,  $q(\boldsymbol{\mu}_i)$  はそれぞれ、事前分布  $p(\boldsymbol{\alpha})$ ,  $p(\boldsymbol{\mu}_i)$  と同じ分布族になる。このような性質をもつ事前分布を自然共役事前分布とよび、ベイズ推定において取り扱いが容易であるためよく用いられる。 $\boldsymbol{\phi}_0$ ,  $\boldsymbol{\mu}_0$ ,  $\boldsymbol{\kappa}_0$  はそれぞれ自然共役事前分布のハイパーパラメータである。ただし、 $\text{Dirichlet}(\mathbf{p}; \mathbf{u})$  は、 $\mathbf{p} = (p_1, \dots, p_n)$ ,  $\mathbf{u} = (u_1, \dots, u_n)$

に対し、

$$\text{Dirichlet}(\mathbf{p}; \mathbf{u}) = \frac{1}{\zeta(\mathbf{u})} \prod_{i=1}^n p_i^{u_i-1}, \quad (21)$$

$$\zeta(\mathbf{u}) = \frac{\prod_{i=1}^n \Gamma(u_i)}{\Gamma(\sum_{i=1}^n u_i)} \quad (22)$$

で定義される Dirichlet 分布の確率密度関数である。(16) 式に (12)(19) 式を代入して計算すると、

$$\begin{aligned} q(\boldsymbol{\alpha}) &\propto \prod_{i=1}^m \alpha_i^{\phi_0-1} \exp(\bar{N}_i \log \alpha_i) \\ &= \prod_{i=1}^m \exp(\log \alpha_i^{\phi_0-1} + \log \alpha_i^{\bar{N}_i}) \\ &= \prod_{i=1}^m \alpha_i^{\phi_0 + \bar{N}_i - 1} \end{aligned} \quad (23)$$

つまり

$$q(\boldsymbol{\alpha}) = \text{Dirichlet}(\{\alpha_i\}_{i=1}^m; \{\phi_i\}_{i=1}^m) \quad (24)$$

となる。ここで  $\phi_i = \phi_0 + \bar{N}_i$  とした。

同様に、(17) 式に (12)(20) 式を代入して計算すると

$$\begin{aligned} q(\boldsymbol{\mu}_i) &\propto e^{\boldsymbol{\kappa}_0 \boldsymbol{\mu}_0^T \boldsymbol{\mu}_i} \cdot e^{\bar{N}_i \boldsymbol{\kappa}_i \boldsymbol{\mu}_i^T \bar{\mathbf{x}}_i} \\ &= e^{r_i \mathbf{e}_i^T \boldsymbol{\mu}_i} \end{aligned} \quad (25)$$

となる。ただし、 $r_i \mathbf{e}_i = \bar{N}_i \boldsymbol{\kappa}_i \bar{\mathbf{x}}_i + \boldsymbol{\kappa}_0 \boldsymbol{\mu}_0$ ,  $\|\mathbf{e}_i\| = 1$ ,  $r_i = \|\bar{N}_i \boldsymbol{\kappa}_i \bar{\mathbf{x}}_i + \boldsymbol{\kappa}_0 \boldsymbol{\mu}_0\|$  である。したがって

$$q(\boldsymbol{\mu}_i) = M_p(\boldsymbol{\mu}_i; \mathbf{e}_i, r_i) \quad (26)$$

となる。

また、(18) 式より

$$\begin{aligned} & \log q(Z) \\ &\propto \left\langle \sum_{i=1}^m \sum_{n=1}^N z_i^n \{ \log \alpha_i + \log C_p(\boldsymbol{\kappa}_i) + \boldsymbol{\kappa}_i \boldsymbol{\mu}_i^T \mathbf{x}_n \} \right\rangle_{q(\boldsymbol{\alpha}), q(\boldsymbol{\mu})} \\ &= \sum_{i=1}^m \sum_{n=1}^N z_i^n [ \langle \log \alpha_i \rangle_{q(\boldsymbol{\alpha})} + \log C_p(\boldsymbol{\kappa}_i) + \boldsymbol{\kappa}_i \langle \boldsymbol{\mu}_i \rangle_{q(\boldsymbol{\mu}_i)}^T \mathbf{x}_n ] \end{aligned} \quad (27)$$

ここで

$$\langle \log \alpha_i \rangle_{q(\boldsymbol{\alpha})} = \Psi(\phi_0 + \bar{N}_i) - \Psi(m\phi_0 + \sum_{i=1}^m \bar{N}_i), \quad (28)$$

$$\langle \boldsymbol{\mu}_i \rangle_{q(\boldsymbol{\mu}_i)} = \mathbf{e}_i \quad (29)$$

を用いる。ただし  $\Psi(\mathbf{x}) = \frac{\partial \log \Gamma(\mathbf{x})}{\partial \mathbf{x}}$  である。すると、(27) 式は

$$\log q(Z) = \sum_{i=1}^m \sum_{n=1}^N z_i^n \gamma_i^n \quad (30)$$

となる。ただし、

$$\gamma_i^n = \left\{ \Psi(\phi_0 + \bar{N}_i) - \Psi\left(m\phi_0 + \sum_{i=1}^m \bar{N}_i\right) \right\} + \log C_p(\kappa_i) + \kappa_i \mathbf{x}_n^T \mathbf{e}_i \quad (31)$$

である。したがって、

$$\bar{z}_i^n = q(z_i^n = 1) = \frac{\exp(\gamma_i^n)}{\sum_{j=1}^m \exp(\gamma_j^n)} \quad (32)$$

となる。

### 4.3 自由エネルギーの計算

前節で求めた各パラメータの事後分布を用いて、(8)式の自由エネルギーを計算する。

$$\mathcal{F}[q(Z), q(\theta)] = H_{\alpha} + \sum_{i=1}^m H_{\mu_i} + H_Z \quad (33)$$

$$H_Z = \langle \log p(D, Z | \theta) \rangle_{q(Z), q(\theta)} - \langle \log q(Z) \rangle_{q(Z)} \quad (34)$$

$$H_{\alpha} = \langle \log p(\alpha) \rangle_{q(\alpha)} - \langle \log q(\alpha) \rangle_{q(\alpha)} \quad (35)$$

$$H_{\mu_i} = \langle \log p(\mu_i) \rangle_{q(\mu_i)} - \langle \log q(\mu_i) \rangle_{q(\mu_i)} \quad (36)$$

とおくと、 $H_{\alpha}$  は (19)(24) 式を用いて、

$$\begin{aligned} H_{\alpha} &= \left\langle \log \left\{ \frac{1}{\zeta(\phi_0)} \prod_{i=1}^m \alpha_i^{\phi_0 - 1} \right\} \right\rangle_{q(\alpha)} \\ &\quad - \left\langle \log \left\{ \frac{1}{\zeta(\phi_i)} \prod_{i=1}^m \alpha_i^{\phi_i - 1} \right\} \right\rangle_{q(\alpha)} \\ &= \left\{ -\log \zeta(\phi_0) + (\phi_0 - 1) \sum_{i=1}^m \langle \log \alpha_i \rangle_{q(\alpha)} \right\} \\ &\quad - \left\{ -\log \zeta(\phi_i) + \sum_{i=1}^m (\phi_i - 1) \langle \log \alpha_i \rangle_{q(\alpha)} \right\} \\ &= \log \frac{\zeta(\phi_i)}{\zeta(\phi_0)} \\ &\quad - \sum_{i=1}^m \bar{N}_i \left\{ \Psi(\phi_0 + \bar{N}_i) - \Psi\left(m\phi_0 + \sum_{i=1}^m \bar{N}_i\right) \right\} \quad (37) \end{aligned}$$

$H_{\mu_i}$  は (20)(26) 式を用いて、

$$\begin{aligned} H_{\mu_i} &= \langle \log \{ C_p(\kappa_i) e^{\kappa_0 \mu_0^T \mu_i} \} \rangle_{q(\mu_i)} \\ &\quad - \langle \log \{ C_p(r_i) e^{r_i \mathbf{e}_i^T \mu_i} \} \rangle_{q(\mu_i)} \\ &= \left\{ \log C_p(\kappa_0) + \kappa_0 \mu_0^T \langle \mu_i \rangle_{q(\mu_i)} \right\} \\ &\quad - \left\{ \log C_p(r_i) + r_i \mathbf{e}_i^T \langle \mu_i \rangle_{q(\mu_i)} \right\} \\ &= \log \frac{C_p(\kappa_0)}{C_p(r_i)} + \kappa_0 \mu_0^T \mathbf{e}_i - r_i \quad (38) \end{aligned}$$

と計算される。以下では、 $\mu_i$  の事前分布 (20) 式において  $\kappa_0 \rightarrow 0$  とした。これにより  $p(\mu_i) \rightarrow \text{const.}$  となり、 $\mu_i$  についての無情報事前分布となる。実際に計算するときは、

$$H_{\mu_i} = -\log C_p(r_i) - r_i \quad (39)$$

を行えばよい。

また、 $H_Z$  は (30)(32) 式より、

$$\begin{aligned} H_Z &= \left\langle \sum_{i=1}^m \sum_{n=1}^N z_i^n \gamma_i^n \right\rangle_{q(Z)} - \sum_{n=1}^N \langle \log q(z^n) \rangle_{q(z^n)} \\ &= \sum_{i=1}^m \sum_{n=1}^N \bar{z}_i^n \gamma_i^n - \sum_{n=1}^N \sum_{i=1}^m q(z_i^n = 1) \log q(z_i^n = 1) \\ &= \sum_{i=1}^m \sum_{n=1}^N [\bar{z}_i^n (\gamma_i^n - \log \bar{z}_i^n)] \\ &= \sum_{i=1}^m \sum_{n=1}^N \bar{z}_i^n \left( \log \sum_{j=1}^m \exp(\gamma_j^n) \right) \\ &= \sum_{n=1}^N \log \sum_{i=1}^m \exp(\gamma_i^n) \quad (40) \end{aligned}$$

となる。

以上の (37)(39)(40) 式を用いて自由エネルギー  $\mathcal{F}[q(Z), q(\theta)]$  を計算する。

### 4.4 VB-EM アルゴリズム

以上の議論により、混合 von Mises-Fisher 分布に対する VB-EM アルゴリズムは以下のようにまとめられる。

#### VB-EM on vMF mixture

**初期化** 事前分布のハイパーパラメータ  $\phi_0$  を設定する。

$\mathbf{e}_i$  の初期値はデータからランダムに  $m$  個サンプリングして用いる。

$$\bar{N}_i^{(0)} \leftarrow \frac{N}{m}$$

$t \leftarrow 0$  とする。

以下を収束するまで繰返し計算する。

#### VB-E ステップ

$i = 1, 2, \dots, m, n = 1, 2, \dots, N$  に対し、

#### **モデルパラメータの更新**

$z_i^n = \frac{1}{\sum_{j=1}^m \exp(\gamma_j^n)} \exp(\gamma_i^n)$  を計算する。

#### **期待値計算**

$H_Z$  を計算する。

$\mathcal{F}[q(Z), q(\theta)] = H_{\alpha} + \sum_{i=1}^m H_{\mu_i} + H_Z$  を計算する。

## VB-M ステップ

$i = 1, 2, \dots, m, n = 1, 2, \dots, N$  に対し,

## モデルパラメータの更新

$$\bar{N}_i^{(t)} \leftarrow \sum_{n=1}^N \bar{z}_i^n$$

$$\mathbf{e}_i^{(t)} \leftarrow \frac{\kappa_i \sum_{n=1}^N \bar{z}_i^n}{\|\kappa_i \sum_{n=1}^N \bar{z}_i^n\|}$$

$\boldsymbol{\mu}_i^{(t)} \leftarrow \mathbf{e}_i^{(t)}$  とする.

## 期待値計算

$H_{\alpha}, H_{\mu_i}$  をそれぞれ計算する.

$t \leftarrow t+1$  として VB-E ステップに戻る.

## 5 実験

人工データを用いて実験を行った.

### [サンプリング]

$S^2$  上と  $S^9$  上にそれぞれランダムに 200 点サンプリングしたデータを用いた.  $S^2$  上のデータについては, ユニット数  $m = 4$ , 中心ベクトル  $\boldsymbol{\mu}_1 = (1, 0, 0), \boldsymbol{\mu}_2 = (0, 0, 1), \boldsymbol{\mu}_3 = (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0), \boldsymbol{\mu}_4 = (0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ , 各ユニットの concentration パラメータをそれぞれ  $\kappa = 20$  とした.

$S^9$  上のデータについては, ユニット数  $m = 4$ , 中心ベクトル  $\boldsymbol{\mu}_1 = (1, 0, 0, 0, 0, 0, 0, 0, 0), \boldsymbol{\mu}_2 = (0, 0, 0, 0, 0, 0, 0, 0, 0), \boldsymbol{\mu}_3 = (0, 0, 0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, 0, 0, 0), \boldsymbol{\mu}_4 = (0, 0, 0, 0, 0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, 0)$ , 各ユニットの concentration パラメータをそれぞれ  $\kappa = 7$  とした.

### [実験方法]

このようにして得られたデータに対して,  $m = 1, 2, \dots, 10$  において, それぞれ 50 エポックずつ VB-EM アルゴリズムを反復して学習を行うという過程を, 初期条件を変えながら 10 回繰り返し, 自由エネルギーを計算した. この実験では, 簡単のため,  $\kappa_i$  を既知としたもとで  $\boldsymbol{\mu}_i$  の推定を行った. ハイパーパラメータとして  $\phi_0 = 0.01$  とした.

また数値計算を行う際に, 第 1 種変形 Bessel 関数  $I_n(x)$  を

$$I_n(x) \approx \frac{1}{\sqrt{2\pi x}} \exp(x) \quad (\text{for } n \ll x) \quad (41)$$

とする近似式を利用した [5],[7].

### [実験結果]

図 1 は  $S^2$ , 図 2 は  $S^9$  においてそれぞれ  $m = 4$  と固定して EM アルゴリズムの反復回数に対する自由エネルギーの推移を見たものである. 横軸はエポック数, 縦軸

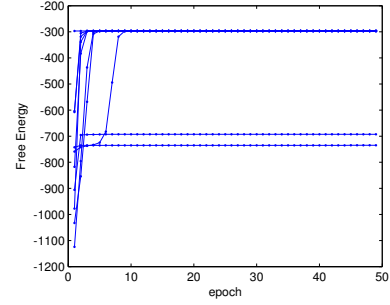


図 1: 自由エネルギーの推移 ( $S^2$ )

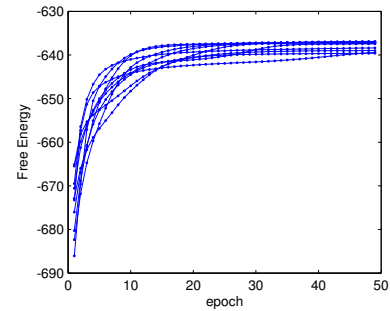


図 2: 自由エネルギーの推移 ( $S^9$ )

は自由エネルギー値を表し, 初期値を変えながら 10 回の試行を行った際の推移を重ねてプロットしてある. 自由エネルギーは単調増加的に収束し, そのスピードも早い. ただし, 図 1 において局所最適解に収束してしまっている場合が 2 例あった.

図 3 は  $m = 4$  としたときの各  $\boldsymbol{\mu}_i$  の推定位置を大きい点で表示したものである. ここで, 小さい点はサンプル点, 中心から外へ向かう線分は真の  $\boldsymbol{\mu}_i$  の位置をそれぞれ示している. 真の値とのユークリッド距離を計算するとその差は  $10^{-7}$  のオーダーとなりよい推定値が得られていることがわかる.

図 4 は  $S^2$ , 図 5 は  $S^9$  においてそれぞれユニット数を調べるために, ユニット数を 1 から 10 まで変化させそれぞれ 10 回試行を行ったときの, 各ユニット数ごとの

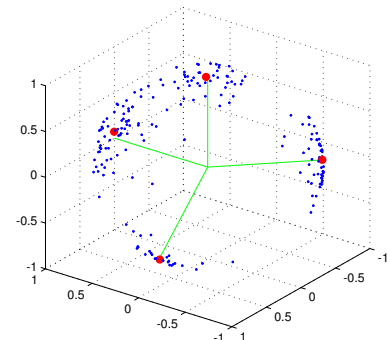


図 3:  $m = 4$  としたときの  $\boldsymbol{\mu}_i$  の推定位置

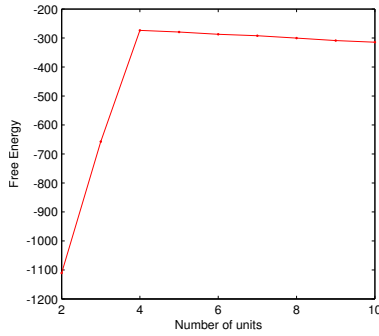


図 4: 各ユニット数に対する自由エネルギーの値 ( $S^2$ )

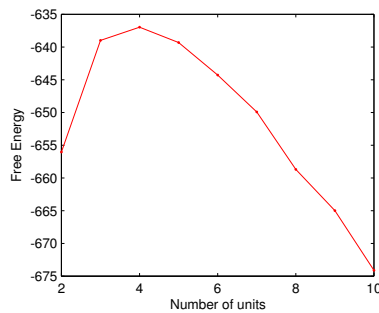


図 5: 各ユニット数に対する自由エネルギーの値 ( $S^9$ )

自由エネルギーの最大値を結んだグラフである。横軸はユニット数、縦軸は自由エネルギーの値を示している。自由エネルギーはユニット数  $m=4$  において最大となり、 $m$  がそれより増えるにしたがって次第に減少して行くことがわかる。また、サンプル 100 セットに対して同様の試行を行った結果、ユニット数  $m=4$  のとき、自由エネルギーが最大とならなかったのは  $S^2$  においては 100 サンプル中 3 サンプル、 $S^9$  においては 100 サンプル中 18 サンプルであった。このことから、自由エネルギー最大化基準によってユニット数を決定すると、真のデータ生成過程における値とよく一致することが分かる。

以上の結果より、本アルゴリズムによって、データ数が少ない場合でも精度の高いパラメータ推定を行うことができることが示された。またユニット数の決定に関して、自由エネルギー値が良い評価基準となることがわかった。

## 6 結論

本研究では、単位超球面上で定義される混合 von Mises-Fisher 分布に関して、変分法的ベイズ推定を行いパラメータを推定するアルゴリズムを提案した。人工データに適用した結果、データ数が少ない場合でも真の値に近い推定がなされていることが確認された。また、自由エネルギーがユニット数の決定に関して、良い評価基準となることがわかった。今後は、より高次元の実データに対して本アルゴリズムを適用することでその有効性を確かめ、また超球面上でのクラスタリングなどに利用していく予定である。

## 参考文献

- [1] I. S. Dhillon and S. Sra. Modeling Data using Directional Distributions. Technical Report TR-03-06, Department of Computer Sciences, University of Texas, January 2003.
- [2] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra. Clustering on hyperspheres using Expectation Maximization. Technical Report TR-03-07, Department of Computer Sciences, University of Texas, February 2003.
- [3] K.V.Mardia and P. Jupp. *Directional Statistics*. John Wiley and Sons Ltd., 2nd edition, 2000.
- [4] 樺島祥介, 上田修功 『計算統計 1 確率計算の新しい手法』統計化学のフロンティア 11, 岩波書店, 2003.
- [5] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Dover Publ. inc., New York, 1974.
- [6] A. Banerjee, I.Dhillon, J.Gosh, and S.sura *Generative Model-based Clustering of Directional Data* Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003.
- [7] G. N. Watson. *A Treatise on the Theory of Bessel Functions*. Cambridge University Press, 2nd edition, 1952.