

# 混合射影正規分布によるクラスタリングについて

小坪琢人

平成 30 年 1 月 11 日

## 目 次

1	はじめに	1
2	はじめに	1
3	混合 Projected Normal 分布	1
3.1	Projected Normal 分布 . . . . .	1
4	混合射影正規分布	3
4.1	射影正規分布 . . . . .	3
4.2	混合射影正規分布 . . . . .	4
5	解析手法	6
5.1	t-SNE . . . . .	6
5.2	マルコフ連鎖モンテカルロ法 . . . . .	6
6	データ解析	7
6.1	評価指標 . . . . .	7
6.2	解析データ . . . . .	7
6.3	解析結果 . . . . .	8
7	まとめと今後の課題	9
8	謝辞	9
	参考文献	10

## 図 目 次

3.1	算術平均による平均の定義 (左), 円周上における平均方向の定義 (右) . . . . .	2
3.2	Text excluding the matrix . . . . .	3
4.1	Text excluding the matrix . . . . .	4
6.1	混合データ (クラスター数:4) . . . . .	8
6.2	Mixture Projected Normal 分布により推定した混合分布 (左), Mixture von Mises 分布により推定した混合分布 (右) . . . . .	8

## 表 目 次

6.1	PN 分布によるクラスター推定 (左), VM 分布によるクラスター推定 (右) . . . . .	9
-----	--	---

## 1 はじめに

## 2 はじめに

円周上や球面上のデータを扱う統計手法を方向統計学といい、近年多様体上の統計分析手法として、注目を集めている。方向統計学とは風向データ、渡り鳥の移動方向データなどの角度観測値を含むデータを対称とする統計学である。単位超球面上 ( $S^{p-1}$ ) に分布するようなデータをユークリッド空間上のデータとして扱うと不適当なことがある。データを多様体上の分布として扱うことで、数学的モデルを作る際に低次元で考えられるなどの利点がある。円周あるいは超球面上の分布を生成するいくつかの方法が知られている。代表的なものは、巻き込み法、射影法、条件付法、投影法である。本研究では射影法を用いた射影分布によるクラスタリングについて議論する。

まず代表的なクラスタリング手法として、 $k$  平均法が挙げられる。 $k$  平均法は、各データと各クラスタの中心のユークリッド距離を最小化することで、各データをクラスタリングする。しかし確率変数が円周上や球面上に値をとるようなデータに対して、ユークリッド空間のクラスタリングを前提とした  $K$  平均法はしばしば誤った分類結果をもたらす。Dhillon and Modha (2001) は、ユークリッド距離に基づく非類似度の尺度を単位球面上に射影したコサイン非類似度の最小化に基づく超球面上の  $k$  平均法を提案した。データの単位方向ベクトルと各クラスタにおける重心ベクトルとのコサイン非類似度を最小化することで、各データをクラスタリングする。

超球面上の  $k$  平均法は確率モデルを仮定しないノンパラメトリックな手法であるのに対し、パラメトリックな超球面上のクラスタリング手法として、Gopal and Yang (2014) による von Mises Fisher 分布の混合分布を用いた手法がある。単位超球面上の分布として知られる、von Mises Fisher 分布を用いて、MCMC サンプリングにより高次元データのクラスタリングを行う。MCMC を用いることで求める事後分布のパラメータの平均値、標準偏差を求めることができる。本研究では方向データの分布として知られる、射影正規分布の混合分布によるクラスタリングの性能評価を行う。射影正規分布はパラメータによって、単峰性・二峰性の形状を取ることができるので、多峰性のデータをクラスタリングする際に混合 von Mises Fisher 分布によるクラスタリングとどのような差異が現れるか比較する。

## 3 混合 Projected Normal 分布

### 3.1 Projected Normal 分布

射影分布は平面または空間上の放射状の射影によって得られる。一般的には、多変量正規ベクトルをノルムで割ることで、単位超球面上への射影分布が得られる。多変量正規ベクトル ( $k$  次元) を  $X$  として、 $k \geq 2$  の場合には、単位超球面上への単位ベクトル  $U$  は  $U = X/\|X\|$  で表される。このとき  $U$  は  $k$  次元の一般化射影正規分布に従い、 $U \sim \mathcal{PN}_k(\mu, \Sigma)$  と表せる。一般化射影正規分

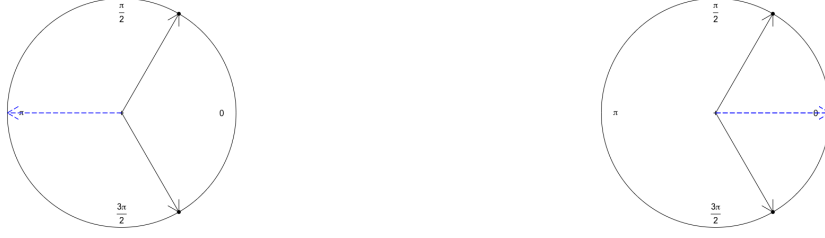


図 3.1: 算術平均による平均の定義 (左), 円周上における平均方向の定義 (右)

布は, パラメータ  $\boldsymbol{\mu}, \Sigma$  をもち, Presnell (1998) らによる,  $\Sigma = \mathcal{I}$  と定義されていた, 射影正規分布を一般化したものである.  $\mathcal{PN}_k(\boldsymbol{\mu}, \mathcal{I})$  は平均方向  $\boldsymbol{\mu}$  に対して, 単峰性かつ対称の分布となるが,  $\mathcal{PN}_k(\boldsymbol{\mu}, \Sigma)$  では非対称分布もしくは二峰性分布となる. ここで平均方向について説明する. 図 3.1 に示した,  $\pi/3$  (実線) ラジアンと  $5\pi/3$  ラジアン (実線) の平均方向について考える. 算術平均では  $(\pi/3 + 5\pi/3)/2 = \pi$  (点線) となるが, 2 つの平均方向としては不適切なことが分かる. よってこれらの角度を円周上の点  $(\cos \frac{\pi}{3}, \sin \frac{\pi}{3}) = (\frac{1}{2}, \frac{\sqrt{3}}{2})$ ,  $(\cos \frac{5\pi}{3}, \sin \frac{5\pi}{3}) = (\frac{1}{2}, -\frac{\sqrt{3}}{2})$  と表し, 2 つのベクトルの平均を取ることで  $\frac{1}{2}(\cos 0, \sin 0)$  となり, 平均方向は 0 ラジアン (点線) となる.

Wang and Gelfand (2013) によると,  $\mathcal{PN}_2(\boldsymbol{\mu}, \Sigma)$  の円形データの場合, 単位円上の方向を表す  $U = (\cos \Theta, \sin \Theta)^T$  における  $\theta$  の確率密度は以下で表す.

$$p(\theta; \boldsymbol{\mu}, \Sigma) = \frac{1}{2\pi A(\theta)} |\Sigma|^{-\frac{1}{2}} \exp(C) \left\{ 1 + \frac{B(\theta)}{\sqrt{A(\theta)}} \frac{\Phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)}{\phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)} \right\} I_{[0, 2\pi)}(\theta)$$

ここで,  $\boldsymbol{u}^T = (\cos \theta, \sin \theta)$ ,  $A(\theta) = \boldsymbol{u}^T \Sigma^{-1} \boldsymbol{u}$ ,  $B(\theta) = \boldsymbol{u}^T \Sigma^{-1} \boldsymbol{\mu}$ ,  $C = -\frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$  であり,  $I_{[0, 2\pi)}(\cdot)$  は指示関数,  $\Phi(\cdot)$ ,  $\phi(\cdot)$  は標準正規分布の確率密度関数と累積密度関数である.

射影正規分布において対称分布, 非対称分布, 二峰性分布となる例を図 4.1 に示す.

Daniel Hernandez-Stumpfhauser et al (2016) によると,  $\mathcal{PN}_3(\boldsymbol{\mu}, \Sigma)$  の球形データの場合, 単位球面上の方向を表す  $U = (\cos \Theta_1 \sin \Theta_2, \sin \Theta_1 \sin \Theta_2, \cos \Theta_2)^T$  における  $\theta_1, \theta_2$  の確率密度は以下で表す.

$$p(\boldsymbol{\theta}; \boldsymbol{\mu}, \Sigma) = \left( \frac{1}{2\pi A(\boldsymbol{\theta})} \right)^{\frac{3}{2}} |\Sigma|^{-\frac{1}{2}} \exp(C) \left( \left[ 1 + D(\boldsymbol{\theta}) \frac{\Phi\{D(\boldsymbol{\theta})\}}{\phi\{D(\boldsymbol{\theta})\}} \right] D(\boldsymbol{\theta}) + \frac{\Phi\{D(\boldsymbol{\theta})\}}{\phi\{D(\boldsymbol{\theta})\}} \right) I_{[0, 2\pi)}(\theta_1) I_{[0, \pi)}(\theta_2)$$

ここで,  $\boldsymbol{u}^T = (\cos \theta_1 \sin \theta_2, \sin \theta_1 \sin \theta_2, \cos \theta_2)$ ,  $D(\boldsymbol{\theta}) = B(\boldsymbol{\theta}) A^{-\frac{1}{2}}(\boldsymbol{\theta})$ ,  $A(\boldsymbol{\theta}) = \boldsymbol{u}^T \Sigma^{-1} \boldsymbol{u}$ ,  $B(\boldsymbol{\theta}) = \boldsymbol{u}^T \Sigma^{-1} \boldsymbol{\mu}$ ,  $C = -\frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$  であり,  $I_{[0, 2\pi)}(\cdot)$ ,  $I_{[0, \pi)}(\cdot)$  は指示関数,  $\Phi(\cdot)$ ,  $\phi(\cdot)$  は標準正規分布の確率密度関数と累積密度関数である.

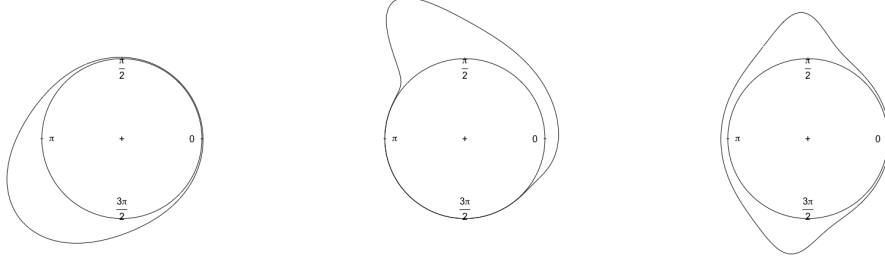


図 3.2:  $\boldsymbol{\mu} = \begin{pmatrix} -1.2 \\ -0.95 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  とした場合の対称な射影正規分布 (左),  $\boldsymbol{\mu} = \begin{pmatrix} -0.19 \\ 2.09 \end{pmatrix}, \Sigma = \begin{pmatrix} 2.49 & -1.85 \\ -1.85 & 1.96 \end{pmatrix}$  とした場合の非対称な射影正規分布 (中央),  $\boldsymbol{\mu} = \begin{pmatrix} -0.24 \\ 0.15 \end{pmatrix}, \Sigma = \begin{pmatrix} 0.209 & 0.068 \\ 0.068 & 2.25 \end{pmatrix}$  とした場合の二峰性となる射影正規分布 (右)

## 4 混合射影正規分布

### 4.1 射影正規分布

射影分布は平面または空間上の放射状の射影によって得られる。一般的には、多変量正規ベクトルをノルムで割ることで、単位超球面上への射影分布が得られる。多変量正規ベクトル ( $k$  次元) を  $X$  として、 $k \geq 2$  の場合には、単位超球面上への単位ベクトル  $U$  は  $U = X/\|X\|$  で表される。このとき  $U$  は  $k$  次元の一般化射影正規分布に従い、 $U \sim \mathcal{PN}_k(\boldsymbol{\mu}, \Sigma)$  と表せる。一般化射影正規分布は、パラメータ  $\boldsymbol{\mu}, \Sigma$  をもち、Presnell (1998) らによる、 $\Sigma = \mathcal{I}$  と定義されていた、射影正規分布を一般化したものである。 $\mathcal{PN}_k(\boldsymbol{\mu}, \mathcal{I})$  は平均方向  $\boldsymbol{\mu}$  に対して、単峰性かつ対称の分布となるが、 $\mathcal{PN}_k(\boldsymbol{\mu}, \Sigma)$  では非対称分布もしくは二峰性分布となる。ここで平均方向について説明する。 $\pi/3$  ラジアンと  $5\pi/3$  ラジアン の平均方向について考える。算術平均では  $(\pi/3 + 5\pi/3)/2 = \pi$  となるが、2つの平均方向としては不適切なことが分かる。よってこれらの角度を円周上の点  $(\cos \pi/3, \sin \pi/3) = (\frac{1}{2}, \frac{\sqrt{3}}{2})$ ,  $(\cos 5\pi/3, \sin 5\pi/3) = (\frac{1}{2}, -\frac{\sqrt{3}}{2})$  と表し、2つのベクトルの平均を取ることで  $\frac{1}{2}(\cos 0, \sin 0)$  となり、平均方向は 0 ラジアンとなる。

Wang and Gelfand (2013) によると、 $\mathcal{PN}_2(\boldsymbol{\mu}, \Sigma)$  の円形データの場合、単位円上の方向を表す  $U = (\cos \Theta, \sin \Theta)^T$  における  $\theta$  の確率密度は以下で表す。

$$p(\theta; \boldsymbol{\mu}, \Sigma) = \frac{1}{2\pi A(\theta)} |\Sigma|^{-\frac{1}{2}} \exp(C) \left\{ 1 + \frac{B(\theta)}{\sqrt{A(\theta)}} \frac{\Phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)}{\phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)} \right\} I_{[0, 2\pi)}(\theta)$$

ここで、 $\mathbf{u}^T = (\cos \theta, \sin \theta)$ ,  $A(\theta) = \mathbf{u}^T \Sigma^{-1} \mathbf{u}$ ,  $B(\theta) = \mathbf{u}^T \Sigma^{-1} \boldsymbol{\mu}$ ,  $C = -\frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$  であり、 $I_{[0, 2\pi)}(\cdot)$  は指示関数、 $\Phi(\cdot)$ ,  $\phi(\cdot)$  は標準正規分布の確率密度関数と累積密度関数である。

射影正規分布において対称分布, 非対称分布, 二峰性分布となる例を図 4.1 に示す.

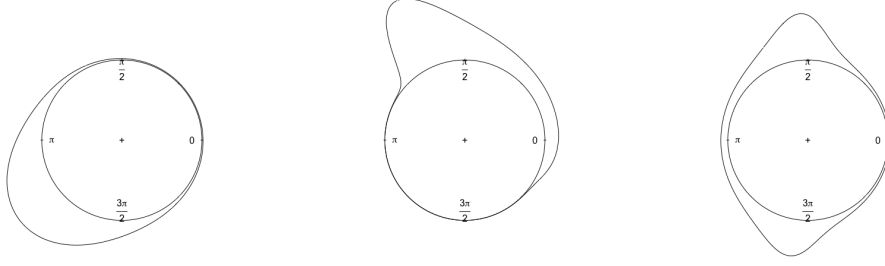


図 4.1:  $\boldsymbol{\mu} = \begin{pmatrix} -1.2 \\ -0.95 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  とした場合の対称な射影正規分布 (左),  $\boldsymbol{\mu} = \begin{pmatrix} -0.19 \\ 2.09 \end{pmatrix}, \Sigma = \begin{pmatrix} 2.49 & -1.85 \\ -1.85 & 1.96 \end{pmatrix}$  とした場合の非対称な射影正規分布 (中央),  $\boldsymbol{\mu} = \begin{pmatrix} -0.24 \\ 0.15 \end{pmatrix}, \Sigma = \begin{pmatrix} 0.209 & 0.068 \\ 0.068 & 2.25 \end{pmatrix}$  とした場合の二峰性となる射影正規分布 (右)

Daniel Hernandez-Stumpfhauser et al (2016) によると,  $\mathcal{PN}_3(\boldsymbol{\mu}, \Sigma)$  の球形データの場合, 単位球面上の方向を表す  $U = (\cos \Theta_1 \sin \Theta_2, \sin \Theta_1 \sin \Theta_2, \cos \Theta_2)^T$  における  $\theta_1, \theta_2$  の確率密度は以下で表す.

$$p(\boldsymbol{\theta}; \boldsymbol{\mu}, \Sigma) = \left( \frac{1}{2\pi A(\boldsymbol{\theta})} \right)^{\frac{3}{2}} |\Sigma|^{-\frac{1}{2}} \exp(C) \left( \left[ 1 + D(\boldsymbol{\theta}) \frac{\Phi\{D(\boldsymbol{\theta})\}}{\phi\{D(\boldsymbol{\theta})\}} \right] D(\boldsymbol{\theta}) + \frac{\Phi\{D(\boldsymbol{\theta})\}}{\phi\{D(\boldsymbol{\theta})\}} \right) I_{[0, 2\pi)}(\theta_1) I_{[0, \pi)}(\theta_2)$$

ここで,  $\mathbf{u}^T = (\cos \theta_1 \sin \theta_2, \sin \theta_1 \sin \theta_2, \cos \theta_2)$ ,  $D(\boldsymbol{\theta}) = B(\boldsymbol{\theta})A^{-\frac{1}{2}}(\boldsymbol{\theta})$ ,  $A(\boldsymbol{\theta}) = \mathbf{u}^T \Sigma^{-1} \mathbf{u}$ ,  $B(\boldsymbol{\theta}) = \mathbf{u}^T \Sigma^{-1} \boldsymbol{\mu}$ ,  $C = -\frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$  であり,  $I_{[0, 2\pi)}(\cdot)$ ,  $I_{[0, \pi)}(\cdot)$  は指示関数,  $\Phi(\cdot)$ ,  $\phi(\cdot)$  は標準正規分布の確率密度関数と累積密度関数である.

## 4.2 混合射影正規分布

$m$  個のユニットからなる円周上の射影正規分布の混合分布は以下のように定式化できる.

$$p(\boldsymbol{\theta}; \mathbf{w}, \boldsymbol{\mu}, \Sigma) = \sum_{j=1}^m w_j \mathcal{PN}_2(\boldsymbol{\theta}; \boldsymbol{\mu}_j, \Sigma_j)$$

ただし,  $w_j$  は混合比率であり,  $0 < w_j < 1$ ,  $\sum_{j=1}^m w_j = 1$  を満たす. 混合射影正規分布におけるパラメータは,  $w, \boldsymbol{\mu}_j, \Sigma_j$  であるが, パラメータに識別可能性を保持するという制約を加えると共分散行列  $\Sigma$  は

$$\Sigma_j = \begin{pmatrix} \tau_j^2 & \rho_j \tau_j \\ \rho_j \tau_j & 1 \end{pmatrix}$$

と定義されるので, パラメータベクトルは  $\boldsymbol{\eta} = (w_1, \dots, w_m, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \tau_1, \dots, \tau_m, \rho_1, \dots, \rho_m)^T$ , パラメータベクトルの次元は  $d = 4m$  となる. 各パラメータの事前分布を  $\boldsymbol{\mu}_j \sim N(\mathbf{0}, 10^5 \mathbf{I}_2)$ ,

$\tau_j \sim \text{halfCauchy}(0, 5)$ ,  $\rho_j \sim U(-1, 1)$ ,  $\mathbf{w} \sim \text{Dirichlet}(2, 2, \dots, 2)$  と設定する. ここで混合比率  $\mathbf{w}$  は  $m$  次の *Dirichlet* 分布に従うものとする.  $\boldsymbol{\mu}_j$  は 0 付近に存在するという背景を元に, 正規分布による弱情報事前分布とし,  $\tau_j$  は非負かつ値が大きくなりすぎないように, 裾の重いコーシー分布を 0 で切断した, 半コーシー分布を弱情報事前分布とする.  $\rho_j$  は  $-1 \sim 1$  に存在することが分かっているので, 一様分布を弱情報事前分布とし,  $\mathbf{w}$  は非負かつ和が 1 となる性質をもつディリクレ分布を弱情報事前分布とする.

角度データ  $\theta$  が得られたときの, パラメータベクトル  $\boldsymbol{\eta}$  の事後分布を  $p(\boldsymbol{\eta}|\theta)$ , パラメータベクトルの事前分布を  $p(\boldsymbol{\eta})$  とすると事後分布は以下で定義できる.

$$p(\boldsymbol{\eta}|\theta) = \frac{p(\theta|\boldsymbol{\eta})p(\boldsymbol{\eta})}{p(\theta)} \propto p(\theta|\boldsymbol{\eta})p(\boldsymbol{\eta})$$

すなわち, 事後分布  $p(\boldsymbol{\eta}|\theta)$  は尤度  $p(\theta|\boldsymbol{\eta})$  と事前分布  $p(\boldsymbol{\eta})$  の積に比例する. ここで,  $p(\theta)$  はすでに得られた, データ  $\theta$  にのみ依存する定数値であるので,  $p(\theta|\boldsymbol{\eta})p(\boldsymbol{\eta})$  が  $\boldsymbol{\eta}$  の分布の形を作り,  $p(\theta)$  は正規化定数とみなすことができる. 尤度と事前分布の計算は簡単だが, 周辺分布  $p(\theta)$  の計算は一般に簡単ではないので, 事後分布に比例する分布  $p(\theta|\boldsymbol{\eta})p(\boldsymbol{\eta})$  から乱数サンプルを発生させて事後分布の代わりとする. この方法で得られたサンプルを MCMC サンプルと呼ぶ. 尤度関数を式 4.1 に示す.

$$\begin{aligned} \log p(\theta|\boldsymbol{\eta}) &= \sum_{j=1}^m \{\log w_j + \log \mathcal{PN}_2(\theta; \boldsymbol{\mu}_j, \Sigma_j)\} \\ &= \sum_{j=1}^m \left[ \log w_j - \log 2\pi - \log A - \frac{1}{2} \log |\Sigma_j| + C + \log \left\{ 1 + \frac{B(\theta)}{\sqrt{A(\theta)}} \frac{\Phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)}{\phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)} \right\} \right] \\ &\propto \sum_{j=1}^m \left[ \log w_j - \log A - \frac{1}{2} \log |\Sigma_j| + C + \log \left\{ 1 + \frac{B(\theta)}{\sqrt{A(\theta)}} \frac{\Phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)}{\phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)} \right\} \right] \quad (4.1) \end{aligned}$$

$m$  個のユニットからなる 球面上の射影正規分布の混合分布は以下のように定式化できる.

$$p(\boldsymbol{\theta}; \mathbf{w}, \boldsymbol{\mu}, \Sigma) = \sum_{j=1}^m w_j \mathcal{PN}_3(\boldsymbol{\theta}; \boldsymbol{\mu}_j, \Sigma_j)$$

ただし,  $w_j$  は混合比率であり,  $0 < w_j < 1$ ,  $\sum_{j=1}^m w_j = 1$  を満たす. 混合射影正規分布におけるパラメータは,  $w, \boldsymbol{\mu}_j, \Sigma_j$  であるが, パラメータに識別可能性を保持するという制約を加えると共分散行列  $\Sigma$  は

$$\Sigma_j = \begin{pmatrix} \Sigma_j^* + \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T & \boldsymbol{\gamma}_j \\ \boldsymbol{\gamma}_j^T & 1 \end{pmatrix}$$

と定義されるので, パラメータベクトルは  $\boldsymbol{\eta} = (w_1, \dots, w_m, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \Sigma_1^*, \dots, \Sigma_m^*, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_m)^T$ , パラメータベクトルの次元は  $d = 4m$  となる. 各パラメータの事前分布を  $\boldsymbol{\mu}_j \sim N(\mathbf{0}, 10^5 \mathbf{I}_3)$ ,

$\gamma_j \sim N(\mathbf{0}, 10^5 \mathbf{I}_2)$ ,  $\Sigma_j^* \sim \text{inverseWishart}(4, \mathbf{I}_2)$ ,  $\mathbf{w} \sim \text{Dirichlet}(2, 2, \dots, 2)$  と設定する. ここで混合比率  $\mathbf{w}$  は  $m$  次の *Dirichlet* 分布に従うものとする.  $\mu_j, \gamma_j$  は 0 付近に存在するという背景を元に, 正規分布による弱情報事前分布とし,  $\Sigma_j^*$  は正定値対象行列となるように, 多変量正規分布の共分散行列の共役事前分布として扱われる逆ウィシャート分布を用いる.  $\mathbf{w}$  は非負かつ和が 1 となる性質をもつディリクレ分布を弱情報事前分布とする.

角度データ  $\theta$  が得られたときの, パラメータベクトル  $\eta$  の事後分布を  $p(\eta|\theta)$ , パラメータベクトルの事前分布を  $p(\eta)$  とすると事後分布は以下で定義できる.

$$p(\eta|\theta) = \frac{p(\theta|\eta)p(\eta)}{p(\theta)} \propto p(\theta|\eta)p(\eta)$$

すなわち, 事後分布  $p(\eta|\theta)$  は尤度  $p(\theta|\eta)$  と事前分布  $p(\eta)$  の積に比例する. ここで,  $p(\theta)$  はすでに得られた, データ  $\theta$  にのみ依存する定数値であるので,  $p(\theta|\eta)p(\eta)$  が  $\eta$  の分布の形を作り,  $p(\theta)$  は正規化定数とみなすことができる. 尤度と事前分布の計算は簡単だが, 周辺分布  $p(\theta)$  の計算は一般に簡単ではないので, 事後分布に比例する分布  $p(\theta|\eta)p(\eta)$  から乱数サンプルを発生させて事後分布の代わりとする. この方法で得られたサンプルを MCMC サンプルと呼ぶ. 尤度関数を式 4.2 に示す.

$$\begin{aligned} \log p(\theta|\eta) &= \sum_{j=1}^m \{\log w_j + \log \mathcal{PN}_3(\theta; \mu_j, \Sigma_j)\} \\ &= \sum_{j=1}^m \left[ \log w_j - \frac{3}{2} \log 2\pi - \frac{3}{2} \log A - \frac{1}{2} \log |\Sigma_j| + C + \log \left( \left[ 1 + D(\theta) \frac{\Phi\{D(\theta)\}}{\phi\{D(\theta)\}} \right] D(\theta) + \frac{\Phi\{D(\theta)\}}{\phi\{D(\theta)\}} \right) \right] \\ &\propto \sum_{j=1}^m \left[ \log w_j - \frac{3}{2} \log A - \frac{1}{2} \log |\Sigma_j| + C + \log \left( \left[ 1 + D(\theta) \frac{\Phi\{D(\theta)\}}{\phi\{D(\theta)\}} \right] D(\theta) + \frac{\Phi\{D(\theta)\}}{\phi\{D(\theta)\}} \right) \right] \end{aligned} \quad (4.2)$$

## 5 解析手法

### 5.1 t-SNE

t-SNE は高次元データの次元を圧縮するアルゴリズムであり, 特に高次元データを可視化する際に有用である. 詳しいアルゴリズムは L. Maaten and G. Hinton (2008) を参照されたい.

### 5.2 マルコフ連鎖モンテカルロ法

本研究では MCMC アルゴリズムの一つである, ハミルトニアン・モンテカルロ法 (HMC) を用いて分布の推定を行う. ハミルトニアンとはポテンシャルエネルギー  $U(\eta)$  と運動エネルギー  $V(\mathbf{q})$  の和で定義される物理量  $H(\eta, \mathbf{q}) = U(\eta) + V(\mathbf{q})$  のことであり. ここで,  $\mathbf{q} = (q_1, \dots, q_d)^T$  は  $\eta$



と同じ  $d$  次元のベクトルである。  $t$  番目のステップにおける、ハミルトニアン方程式は以下で定義される。

$$\frac{d\eta_j(t)}{dt} = \frac{\partial V(\mathbf{q})}{\partial q_j}, \frac{dq_j(t)}{dt} = -\frac{\partial U(\boldsymbol{\eta})}{\partial \eta_j}$$

ここで、 $\boldsymbol{\eta}, \mathbf{q}$  の同時確率密度関数  $p(\boldsymbol{\eta}, \mathbf{q})$  をハミルトニアン  $H(\boldsymbol{\eta}, \mathbf{q})$  の関数で定義したとき、上記のハミルトン方程式に従って、 $\boldsymbol{\eta}, \mathbf{q}$  の値を変化させると、確率密度関数  $p(\boldsymbol{\eta}, \mathbf{q}) = f(H(\boldsymbol{\eta}, \mathbf{q}))$  からのサンプリングとみなすことができる。また  $\boldsymbol{\eta}, \mathbf{q}$  が独立であることを用いて、リープ・フロッグ法で  $\boldsymbol{\eta}$  の分布を求める。リープ・フロッグ法について以下に簡単に示すが、詳しい解説は Neal (2011) や Hoffman and Gelman (2013) を参照されたい。

1.  $\eta_j(0), q_j(0)$  の初期値を設定する。  $\eta_j(0)$  は各パラメータの事前分布からの設定し、  $q_j(0)$  は  $N(0, M_j)$  からランダムサンプリングする。  $M_j$  は任意の分散パラメータである。

2. 以下の式に従い、 $\boldsymbol{\eta}, \mathbf{q}$  を更新する。  $\epsilon$  は状態変化のステップ幅を表す。

$$q_j(t + \frac{\epsilon}{2}) = q_j(t) - \frac{\epsilon}{2} \frac{\partial U}{\partial \eta_j}(\boldsymbol{\eta}(t)), \eta_j(t + \epsilon) = \eta_j(t) + \epsilon \frac{q_j(t + \frac{\epsilon}{2})}{M_j}, q_j(t + \epsilon) = q_j(t + \frac{\epsilon}{2}) - \frac{\epsilon}{2} \frac{\partial U}{\partial \eta_j}(\boldsymbol{\eta}(t + \epsilon))$$

3. 得られたパラメータの採択、棄却を決定する。上記のステップで得られたパラメータを  $\boldsymbol{\eta}^*, \mathbf{q}^*$  として、現在が  $t$  回目の反復とすると、サンプルの棄却率は

$$r = \frac{p(\boldsymbol{\eta}^*|\boldsymbol{\theta})p(\mathbf{q}^*)}{p(\boldsymbol{\eta}(t-1)|\boldsymbol{\theta})p(\mathbf{q}(t-1))}$$

と定義され、0 から 1 の間でランダムに選んだ数値より、  $r$  が高ければ  $\boldsymbol{\eta}^*$  を採択する。サンプル数が定めた値に達したら終了する。

## 6 データ解析

### 6.1 評価指標

情報量基準 (WAIC) を用いて、混合分布のクラスター推定を行う。AIC や BIC などの情報量基準には、事後分布が正規分布で近似されている必要があるなど、様々な制約が存在するが、WAIC は真の分布、確率モデル、事前分布がどのような場合でも用いることができる。

### 6.2 解析データ

$\mu(0 < \mu < 2\pi), \kappa(\kappa \geq 0)$  をパラメータ、  $\theta(0 \leq \theta < 2\pi)$  を確率変数とする、Von Mises 分布の確率密度関数は以下の式で表せる。また式中の第一種変形ベッセル関数の定義式を示す。

$$f(\theta; \mu, \kappa) = \frac{\exp[\kappa \cos(\theta - \mu)]}{2\pi I_0(\kappa)}, I_j(\kappa) = \left(\frac{\kappa}{2}\right)^j \sum_{i=0}^{\infty} \frac{\left(\frac{\kappa^2}{4}\right)^i}{i\Gamma(j+i+1)}$$

ユニット数  $m = 4$ , 角度  $\theta = \left(\frac{7\pi}{4}, \frac{2\pi}{3}, \frac{5\pi}{4}, \frac{\pi}{6}\right)$ , パラメータ  $\kappa = (2, 10, 5, 15)$ , 混合比率  $(0.4, 0.2, 0.3, 0.1)$  として, 乱数を生成する.

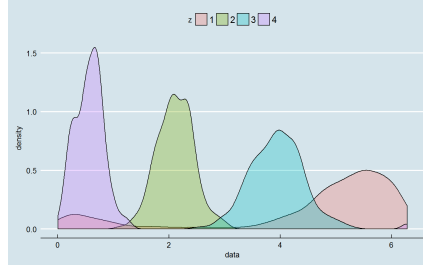


図 6.1: 混合データ (クラスター数:4)

### 6.3 解析結果

Mixture von Mises 分布では元のデータと同じく 4 つの元となる分布を推定したが, Mixture Projected Normal 分布においては 3 つの元となる分布により以下の結果が得られた.

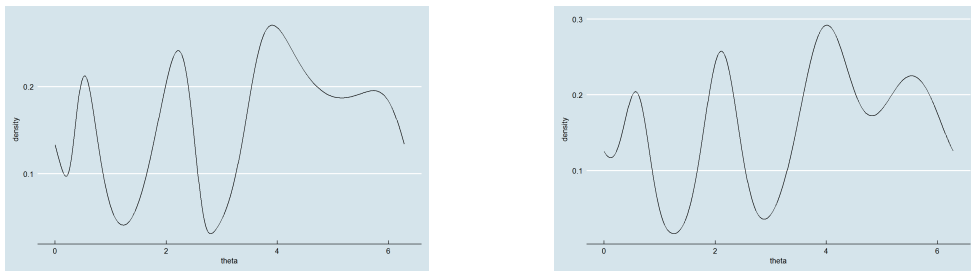


図 6.2: Mixture Projected Normal 分布により推定した混合分布 (左), Mixture von Mises 分布により推定した混合分布 (右)

得られた混合分布から, 混合データの分類を行う. 混合分布を構成する, 分布の中で出現確率が最も高いものをそのデータの予測クラスタとする. 横軸を真のクラスタ, 縦軸を予測クラスタとして集計表を示す.

表 6.1: PN 分布によるクラスター推定 (左), VM 分布によるクラスター推定 (右)

		Predict		
		1	2	3
True	1	38	10	352
	2	1	192	7
	3	0	2	298
	4	93	0	7

		Predict			
		1	2	3	4
True	1	333	10	39	18
	2	7	192	1	0
	3	298	2	0	0
	4	0	1	94	5

## 7 まとめと今後の課題

## 8 謝辞

本研究を行うにあたり, ご指導を頂いた塩濱教授に感謝しています. またお忙しい中, 学部生の質問に対応して頂いた院生の皆様にも感謝しています.

## 参考文献

- [1] Presnell. B, Morrison. S. P, and Littell. R. C. Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, Vol. 443, pp. 1068–1077, 1998.
- [2] D.Hernandez-Stumpfhauser, F. Jay Breidt, and Mark J. van der Woerd. The general projected normal distribution of arbitrary dimension: Modeling and bayesian inference. *Journal of Bayesian Analysis*, Vol. 12, No. 1, pp. 113–133, 2017.
- [3] S. Dhillon and S. Modha. Concept decompositions for large sparse text data using clustering. *Machine Learning*, Vol. 42, pp. 143–175, 2001.
- [4] Siddharth Gopal and Yiming Yang. Von mises-fisher clustering models. In *ICML*, 2014.
- [5] Daniel Hernandez-Stumpfhauser, F. Jay Breidt, and Mark J. van der Woerd. Supplementary material of the general projected normal distribution of arbitrary dimension: Modeling and bayesian inference. *Bayesian Analysis*, 2016.
- [6] Hoffman. M and Gelman. A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 2014.
- [7] Neal. R. M. Mcmc using hamiltonian dynamics. In *Handbook of Markov Chain MonteCarlo*, pp. 113–162, 2011.
- [8] L. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning*, Vol. 9, pp. 2579–2605, 2008.
- [9] F. Wang and A. E. Gelfand. Directional data analysis under the general projected normal distribution. *Statistical Methodology*, Vol. 10, No. 1, pp. 113–127, 2013.

関係ないページです！

Presnell(1998) [1]

Wang and Gelfand (2013) [9]

D. Hemandez(2017) [2]

Dhillon and Modha(2001) [3]

Gopal and Yang(2014) [4]

Neal (2011) [7]

Hoffman and Gelman (2013) [6]

Daniel Hernandez-Stumpfhauser (2016) [5]

L. Maaten and G. Hinton (2008) [8]