

# 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 8 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
3.2	混合 Projected Normal 分布 . . . . .	3
4	解析手法	4
4.1	マルコフ連鎖モンテカルロ法 . . . . .	4
5	シミュレーション	5
5.1	解析データ . . . . .	5
5.2	解析結果 . . . . .	6
6	まとめと今後の課題	7
7	謝辞	7
	参考文献	8

## 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) . . . . .	1
3.2	Text excluding the matrix . . . . .	2
5.1	混合データ (クラスター数:4) . . . . .	6

5.2 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) . . . . .	6
--	---

## 表 目 次

5.1 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) . . . . .	6
--	---

## 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

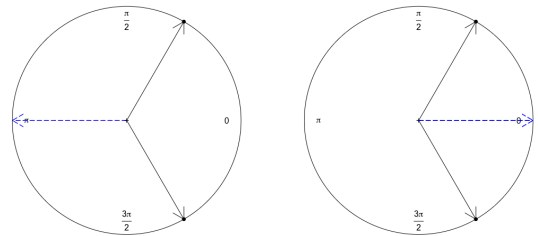


図 3.1: 算術平均による平均の定義 (左), 円周上  
1 における平均方向の定義 (右)

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．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,  $\mathbf{u}^T = (\cos \theta, \sin \theta)$ ,  $A(\theta) = \mathbf{u}^T \Sigma^{-1} \mathbf{u}$ ,  $B(\theta) = \mathbf{u}^T \Sigma^{-1} \boldsymbol{\mu}$ ,  $C = -\frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$  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 3.2 に示す．

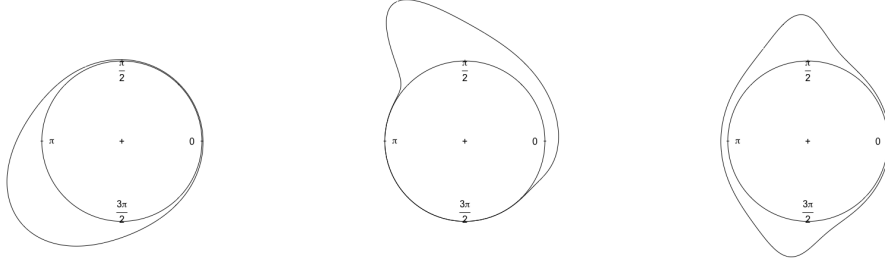


図 3.2:  $\boldsymbol{\mu} = \begin{pmatrix} -1.2 \\ -0.95 \end{pmatrix}$ ,  $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  とした場合の対称な射影正規分布 (左),  $\boldsymbol{\mu} = \begin{pmatrix} -0.19 \\ 2.09 \end{pmatrix}$ ,  $\Sigma = \begin{pmatrix} 2.49 & -1.85 \\ -1.85 & 1.96 \end{pmatrix}$  とした場合の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)

### 3.2 混合 Projected Normal 分布

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

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

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  $\boldsymbol{\eta} = (w_1, \dots, w_m, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \tau_1, \dots, \tau_m, \rho_1, \dots, \rho_m)$  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出来る. 尤度と事前分布の計算は簡単だが,  $p(\theta)$  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. 尤度は 1 より小さい多数の値の積であり, そのままでは扱いにくいので尤度の対数を取るのが一般的である. これを対数尤度と呼び, 式 3.1 に示す.

$$\begin{aligned}
\log p(\theta|\boldsymbol{\eta}) &= \sum_{j=1}^m \{\log w_j + \log \mathcal{PN}_2(\theta; \boldsymbol{\mu}_j, \Sigma_j)\} \\
&= \sum_{j=1}^m \left[ \log w_j - \log 2\pi - \log A - \frac{1}{2} \log |\Sigma| + C + \log \left\{ 1 + \frac{B(\theta)}{\sqrt{A(\theta)}} \frac{\Phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)}{\phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)} \right\} \right] \\
&\propto \sum_{j=1}^m \left[ \log w_j - \log A - \frac{1}{2} \log |\Sigma| + C + \log \left\{ 1 + \frac{B(\theta)}{\sqrt{A(\theta)}} \frac{\Phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)}{\phi\left(\frac{B(\theta)}{\sqrt{A(\theta)}}\right)} \right\} \right] \quad (3.1)
\end{aligned}$$

## 4 解析手法

### 4.1 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。HMC のアルゴリズムはハミルトン方程式の数学的性質によって正当性が保証されている。一般の確率分布からのサンプリングについて、ハミルトン方程式を定義するには、元の問題に存在しない仮想的な運動量を補助変数として導入する必要がある。

HMC は物理学の考えを用いて、メトロポリスアルゴリズムによるランダムウォークな振る舞いを局所的に抑圧する。従って、目標の分布までより素早く推移することを可能にする。対象となる各確率変数  $(\theta_j)$  に対して、速度変数  $\phi_j$  を付与する。このメトロポリスアルゴリズムでは  $\theta, \phi$  を共に更新する。 $\theta$  の分布の推移には  $\phi$  が大きく影響する。

HMC では、事後確率  $p(\theta|y)$  は  $p(\phi)$  の独立な分布によって増大する。したがって結合分布を以下のように定義する。

$$p(\theta, \phi|y) = p(\phi)p(\theta|y)$$

$\theta$  をシュミレーションするために、上記の結合分布によりパラメータをシュミレーションする。この補助変数  $\phi$  を導入することでパラメータ空間をより効率的に移動することを可能にする。HMC は対数事後密度の勾配を要求する。 $d$  次元のパラメータ  $\theta$  の勾配は以下で表す。

$$\frac{d \log p(\boldsymbol{\theta}|y)}{d\boldsymbol{\theta}} = \left( \frac{d \log p(\boldsymbol{\theta}|y)}{d\theta_1}, \dots, \frac{d \log p(\boldsymbol{\theta}|y)}{d\theta_d} \right)$$

$\phi$  については、平均 0、共分散行列  $M$  の多変量正規分布を与える。 $\phi$  の要素が独立の場合は各次元ごとに  $\phi_j \sim N(0, M_{jj})$  を与える。 $M$  は、事後分布の逆共分散行列  $(\text{var}(\theta|y))^{-1}$  で大まかにス

ケールすることができる。しかし、アルゴリズムはどんな場合でも動作するが、 $M$  のスケーリングが優れているだけで、HMC がより効率的となる。

HMC では 3 つの動作を繰り返し行う。

1. この反復は、 $\phi$  をその事後分布から無作為に引くことによって開始する。これは、 $\phi \sim N(0, M)$  の事前分布と同様である。

2.  $\theta, \phi$  を同時に更新する。

$$\phi_{t+\frac{1}{2}} \leftarrow \phi_t + \frac{1}{2}\epsilon \frac{d \log p(\theta|y)}{d\theta}, \theta_{t+1} \leftarrow \theta_t + \epsilon M^{-1} \phi_{t+\frac{1}{2}}, \phi_{t+1} \leftarrow \phi_{t+\frac{1}{2}} + \frac{1}{2}\epsilon \frac{d \log p(\theta|y)}{d\theta_{t+1}}.$$

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  $\theta^*, \phi^*$  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

$$r = \frac{p(\theta^*|y)p(\phi^*)}{p(\theta^{t-1}|y)p(\phi^{t-1})}$$

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  $\theta^*$  を採択する。サンプル数が定めた値に達したら、終了する。

## 5 シミュレーション

人工的データを用いて、混合 Projected Normal 分布によるクラスタリングの性能評価を行う。複数の von Mises 分布に従う乱数を発生させて、1 つのデータにまとめることで混合データを作成する。

### 5.1 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  $\theta = (\frac{7\pi}{4}, \frac{2\pi}{3}, \frac{5\pi}{4}, \frac{\pi}{6})$ , 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。

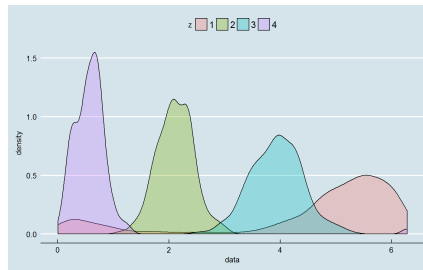


図 5.1: 混合データ (クラスター数:4)

## 5.2 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.

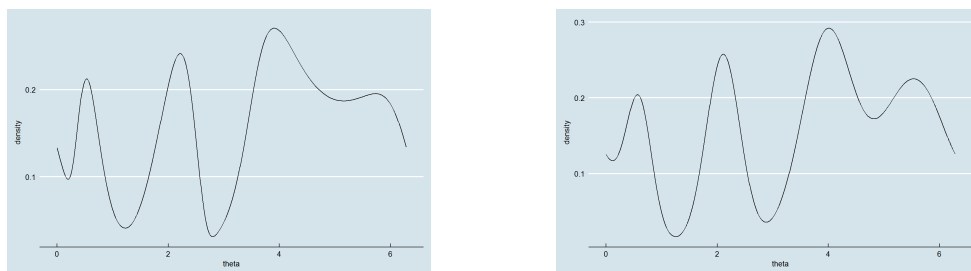


図 5.2: Mixture Projected Normal 分布により推定した混合分布 (左), Mixture von Mises 分布により推定した混合分布 (右)

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 5.1: 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



## 6 まとめと今後の課題

## 7 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] 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) [5]

D. Hemandez(2017) [2]

Dhillon and Modha(2001) [3]

Gopal and Yang(2014) [4]