

# Dirichlet Process Mixtureの導出

電気通信大学 電気通信学研究科 中村 友昭 (naka\_t@apple.ee.uec.ac.jp)

平成 23 年 2 月 23 日

## 1 Dirichlet Process Mixture

### 1.1 ノンパラメトリックベイズ

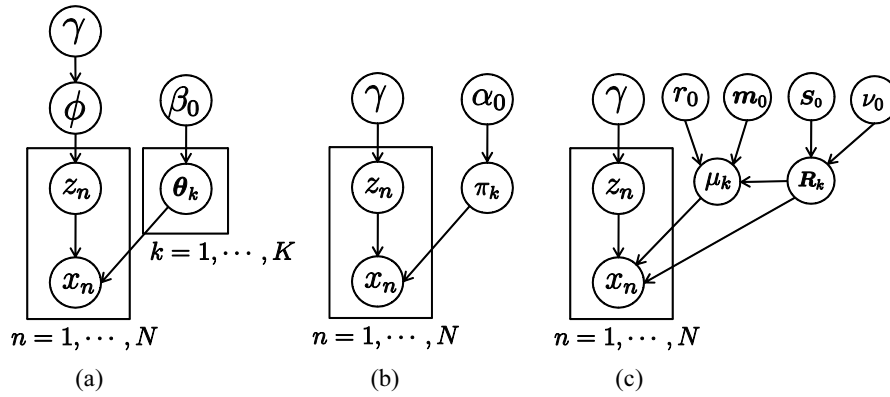


図 1: (a) パラメトリックベイズモデル (b) ディリクレ過程混合多項分布 (c) ディリクレ過程混合ガウス分布

図 1(a) に通常のパラメトリックベイズモデルのグラフィカルモデルを示す．パラメトリックベイズでは，あらかじめ  $K$  個の確率分布  $\theta_k$  を用意されており，確率分布  $P(z_n|\phi)$  に従い  $n$  番目のデータのクラス  $z_n$  が決定される．さらに，データ  $x_n$  は， $k = z_n$  と対応した確率分布  $P(\theta_k|\beta_0)$  から生成される．ただし， $\phi$  と  $\gamma$  はそれぞれ，多項分布とディリクレ事前分布のパラメタである．また， $\theta_k$  はデータを生成する確率分布のパラメタであり，そのハイパーパラメタが  $\beta_0$  である．

一方，ノンパラメトリックベイズである Dirichlet Process Mixture (DPM) では，事前分布にディリクレ過程を導入することで，クラス数  $K$  をあらかじめ設定する必要はなく，観測データの複雑さに応じて自動的にクラス数が決定される．

本稿では，ディリクレ過程の一つである Chinese Restaurant Process (CRP) を用いる．CRP では， $n$  番目のデータのクラスが  $z_n$  となる確率は以下ようになる．

$$P(z_n = k|\gamma) = \begin{cases} \frac{N_k}{\gamma+N-1} & (k = 1, \dots, K) \\ \frac{\gamma}{\gamma+N-1} & (k = K+1) \end{cases} \quad (1)$$

ただし,  $N$  はデータの総数,  $N_k$  はクラス  $k$  に属するデータの総数である. また,  $K$  は現在のクラス数であり,  $k = K+1$  とは新たなクラスを生成することを意味している. したがって,  $x_n$  が属するクラス  $z_n$  の事後確率は, ベイズの定理から次のようになる.

$$P(z_n = k|\mathbf{X}, \mathbf{Z}, \gamma) \propto P(\mathbf{x}_n|\mathbf{X}_k)P(z_n|\gamma) \quad (2)$$

$$= \begin{cases} P(\mathbf{x}_n|\mathbf{X}_k)\frac{n_k}{\gamma+n-1} & (k = 1, \dots, K) \\ P(\mathbf{x}_n|\mathbf{X}_k)\frac{\gamma}{\gamma+n-1} & (k = K+1) \end{cases} \quad (3)$$

ただし,  $\mathbf{X} = \{x_0, \dots, x_N\}$ ,  $\mathbf{Z} = \{z_0, \dots, z_N\}$ ,  $\mathbf{X}_k$  はクラス  $k$  に属するデータの集合である. より詳しい説明は, 参考文献を参照されたい [1, 2].

## 1.2 DPM の学習

DPM でのクラスタリングは, Gibbs サンプリングによって, モデルのパラメタ  $\theta_k$  と, 混合数  $K$  を推定することによって実現される. Gibbs Sampling では,  $x_n$  を除いたデータの集合を条件とした条件付確率

$$z_n \sim P(z_n|\mathbf{X}_{-n}, \mathbf{Z}_{-n}, \gamma) \quad (4)$$

から  $z_n$  をサンプリングし, クラス  $z_n$  のパラメタ  $\theta_{z_n}$  を更新する. この操作を, 全ての  $n$  について繰り返すことで真のパラメタ  $\theta_k$  を推定することができる.

## 2 DPM の具体的な導出

実際に DPM を適用する際に, 式 (3) 内の  $P(\mathbf{x}_n|\mathbf{X}_k)$  を, 仮定する分布に合わせて具体化する必要がある. 本稿では, ガウス分布と多項分布について導出を行う.

### 2.1 多項分布

図 1(b) が多項分布の DPM モデルである.  $\gamma$  は新しいクラスを生成する確率を制御するパラメタ,  $z_n$  は  $n$  番目のデータのクラス,  $x_n$  は  $n$  番目のデータである.

多項分布のモデルでは,  $\mathbf{x}_n$  のクラス  $z_n = k$  となるとき,  $\mathbf{x}_n$  はパラメタが  $\pi_k$  の多項分布から生成される.

$$P(\mathbf{x}_n | \pi_k) = \mathcal{M}(\mathbf{x}_n | \pi_k) = \prod_i^d \pi_{ki}^{x_{ni}} \quad (5)$$

ただし,  $d$  はデータの次元数を表している. 次に, クラス  $k$  に属する全データを  $\mathbf{X}_k$  と置くと, その尤度は,

$$P(\mathbf{X}_k | \pi_k) = \prod_n^{N_k} P(\mathbf{x}_n | \pi_k) = \prod_n^{N_k} \prod_i^d \pi_{ki}^{x_{ni}} = \prod_i^d \pi_{ki}^{\sum_n^{N_k} x_{ni}} \quad (6)$$

となる. さらに,  $\pi_k$  はディリクレ事前分布

$$P(\pi_k) = \mathcal{D}(\pi_k | \alpha_0) = \frac{1}{Z(\alpha_0)} \prod_i^d \pi_{ki}^{\alpha_{0i}-1} \quad (7)$$

によって生成される. ただし,  $Z(\alpha_0)$  は正規化項を表し以下のようになる.

$$Z(\alpha_0) = \frac{\prod_i^d \Gamma(\alpha_{0i})}{\Gamma(\sum_i^d \alpha_{0i})} \quad (8)$$

以上より,  $\mathbf{X}_k$  と  $\pi_k$  の同時確率は次式のようになる.

$$P(\mathbf{X}_k, \pi_k) = P(\mathbf{X}_k | \pi_k) P(\pi_k) = \prod_i^d \pi_{ki}^{\sum_n^{N_k} x_{ni}} \frac{1}{Z(\alpha_0)} \prod_i^d \pi_{ki}^{\alpha_{0i}-1} \quad (9)$$

$$= \frac{1}{Z(\alpha_0)} \prod_i^d \pi_{ki}^{\sum_n^{N_k} x_{ni} + \alpha_{0i} - 1} \quad (10)$$

次に,  $\pi_k$  で周辺化し, 周辺尤度を求める.

$$P(\mathbf{X}_k) = \int P(\mathbf{X}_k, \pi_k) d\pi_k \quad (11)$$

$$= \frac{1}{Z(\alpha_0)} \int \prod_i^d \pi_{ki}^{\sum_n^{N_k} x_{ni} + \alpha_{0i} - 1} d\pi_k \quad (12)$$

$$= \frac{Z(\alpha'_k)}{Z(\alpha_0)} \int \frac{1}{Z(\alpha'_k)} \prod_i^d \pi_{ki}^{\alpha'_{ki}-1} d\pi_k \quad (13)$$

ただし,  $\alpha'_{ki} = \sum_n^{N_k} x_{ni} + \alpha_{0i}$  である. この式の積分内がディリクレ分布と見なせ,

$$\int \frac{1}{Z(\alpha'_k)} \prod_i^d \pi_{ki}^{\alpha'_{ki}-1} d\pi_k = \int \mathcal{D}(\pi_k | \alpha'_k) d\pi_k = 1 \quad (14)$$

が成り立ち，最終的に以下の式を得る．

$$P(\mathbf{X}_k) = \frac{Z(\boldsymbol{\alpha}'_k)}{Z(\boldsymbol{\alpha}_0)} \quad (15)$$

DPM で必要となるのは，新たなデータ  $\mathbf{x}_{new}$  がクラス  $k$  に属する尤度であり，以下のようなになる．

$$P(\mathbf{x}_{new}|\mathbf{X}_k) = \frac{P(\mathbf{x}_{new}, \mathbf{X}_k)}{P(\mathbf{X}_k)} = \frac{Z(\boldsymbol{\alpha}'_k + \mathbf{x}_{new})}{Z(\boldsymbol{\alpha}'_k)} \quad (16)$$

以上より，Gibbs Sampling によるパラメタ推定アルゴリズムは次のようになる．  
 [Multinomial DPM algorithm]—

while 収束するまで

for  $n = 0, \dots, N$

1.  $\mathbf{x}_n$  をクラス  $k = z_n$  から削除し，パラメタを更新

$$\boldsymbol{\alpha}'_{k-} = \mathbf{x}_n \quad (17)$$

$$N_{k-} = 1 \quad (18)$$

2. 新たな  $z_n$  を事後分布からサンプリング

$$P(z_n = k | \mathbf{X}_k) \propto \begin{cases} P(\mathbf{x}_n | \mathbf{X}_k) \frac{N_k}{\gamma + N - 1} & (k = 1, \dots, K) \\ P(\mathbf{x}_n | \mathbf{X}_k) \frac{\gamma}{\gamma + N - 1} & (k = K + 1) \end{cases} \quad (19)$$

ただし， $\boldsymbol{\alpha}'_{K+1} = \boldsymbol{\alpha}_0$  とする．

3.  $\mathbf{x}_n$  をクラス  $k$  に追加し，パラメタを更新

$$\boldsymbol{\alpha}'_{k+} = \mathbf{x}_n \quad (20)$$

$$N_{k+} = 1 \quad (21)$$

$$z_n = k \quad (22)$$

新たなクラスが生成された場合  $K_+ = 1$  とする．

4. 空のクラスを削除

end for

end while

---

## 2.2 ガウス分布

図 1(c) がガウス分布の DPM モデルである．前節同様， $z_n$  は  $n$  番目のデータのクラス， $x_n$  は  $n$  番目のデータである．ガウス分布のモデルでは， $x_n$  のクラスが  $z_n = k$  となるときの， $x_n$  は平均  $\mu_k$ ，精度行列  $R_k$  のガウス分布から生成される．精度行列  $R_k$  の事前分布として  $S_0, \nu_0$  をパラメタとするウィシャート分布

$$\begin{aligned} P(\mathbf{R}_k) &= \mathcal{W}(\mathbf{R}_k | \mathbf{S}_0, \nu_0) \\ &= 2^{-\frac{\nu_0 d}{2}} \pi^{-\frac{d(d-1)}{4}} \mathbf{S}_0^{\frac{\nu_0}{2}} \prod_i^d \Gamma\left(\frac{\nu_0 + 1 - i}{2}\right)^{-1} \\ &\quad \times |\mathbf{R}_k|^{\frac{\nu_0 - d - 1}{2}} \exp\left(-\frac{1}{2} \text{Tr}(\mathbf{R}_k \mathbf{S}_0)\right) \end{aligned} \quad (23)$$

を，平均  $\mu_k$  の事前分布として  $r_0 \mathbf{R}_k, \nu_0$  をパラメタとするガウス分布

$$\begin{aligned} P(\mu_k | \mathbf{R}_k) &= \mathcal{N}(\mu_k | r_0 \mathbf{R}_k, \mathbf{m}_0) \\ &= (2\pi)^{-\frac{d}{2}} |r_0 \mathbf{R}_k|^{\frac{1}{2}} \exp\left(-\frac{1}{2} (\mu_k - \mathbf{m}_0)^T r_0 \mathbf{R}_k (\mu_k - \mathbf{m}_0)\right) \\ &= (2\pi)^{-\frac{d}{2}} r_0^{\frac{d}{2}} |\mathbf{R}_k|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \text{Tr}(r_0 \mathbf{R}_k (\mu_k - \mathbf{m}_0) (\mu_k - \mathbf{m}_0)^T)\right) \end{aligned} \quad (24)$$

を仮定している．ただし，2 行目から 3 行目の変形では， $\mathbf{a}^T \mathbf{A} \mathbf{a} = \text{Tr}(\mathbf{A} \mathbf{a} \mathbf{a}^T)$  と  $|\mathbf{a} \mathbf{A}| = a^d |\mathbf{A}|$  を用いた．以上の事前分布より， $\mu_k$  と  $R_k$  の同時確率は次のようになる．

$$\begin{aligned} P(\mu_k, \mathbf{R}_k) &= P(\mu_k | \mathbf{R}_k) P(\mathbf{R}_k) \\ &= (2\pi)^{-\frac{d}{2}} r_0^{\frac{d}{2}} |\mathbf{R}_k|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \text{Tr}(r_0 \mathbf{R}_k (\mu_k - \mathbf{m}_0) (\mu_k - \mathbf{m}_0)^T)\right) \\ &\quad \times 2^{-\frac{\nu_0 d}{2}} \pi^{-\frac{d(d-1)}{4}} \mathbf{S}_0^{\frac{\nu_0}{2}} \prod_i^d \Gamma\left(\frac{\nu_0 + 1 - i}{2}\right)^{-1} \\ &\quad \times |\mathbf{R}_k|^{\frac{\nu_0 - d - 1}{2}} \exp\left(-\frac{1}{2} \text{Tr}(\mathbf{R}_k \mathbf{S}_0)\right) \\ &= \frac{1}{Z(d, r_0, \nu_0, \mathbf{S}_0)} |\mathbf{R}_k|^{\frac{\nu_0 - d}{2}} \\ &\quad \times \exp\left[-\frac{1}{2} \text{Tr}\left[\mathbf{R}_k (r_0 (\mu_k - \mathbf{m}_0) (\mu_k - \mathbf{m}_0)^T + \mathbf{S}_0)\right]\right] \end{aligned} \quad (25)$$

$$\begin{aligned} &= \frac{1}{Z(d, r_0, \nu_0, \mathbf{S}_0)} |\mathbf{R}_k|^{\frac{\nu_0 - d}{2}} \\ &\quad \times \exp\left[-\frac{1}{2} \text{Tr}\left[\mathbf{R}_k (r_0 \mu_k \mu_k^T - 2r_0 \mu_k \mathbf{m}_0^T + r_0 \mathbf{m}_0 \mathbf{m}_0^T + \mathbf{S}_0)\right]\right] \end{aligned} \quad (26)$$

ただし，式変形にはベクトルの性質である  $ab^T = ba^T$  を用いた．また，

$$\frac{1}{Z(d, r_0, \nu_0, \mathbf{S}_0)} = 2^{-\frac{(\nu_0+1)d}{2}} \pi^{-\frac{d(d+1)}{2}} r_0^{\frac{d}{2}} |\mathbf{S}_0|^{\frac{\nu_0}{2}} \prod_i^d \Gamma\left(\frac{\nu_0+1-i}{2}\right)^{-1} \quad (27)$$

である．

次に，クラス  $k$  に属する全データ  $\mathbf{X}_k$  と置くと，その尤度は次のようになる．

$$\begin{aligned} P(\mathbf{X}_k | \boldsymbol{\mu}_k, \mathbf{R}_k) &= \prod_n^{N_k} \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \mathbf{R}_k) \\ &= \prod_n^{N_k} \left[ 2\pi^{-\frac{d}{2}} |\mathbf{R}_k|^{\frac{1}{2}} \right. \\ &\quad \left. \times \exp \left\{ -\frac{1}{2} \text{Tr} \left( \mathbf{R}_k (\mathbf{x}_n - \boldsymbol{\mu}_k) (\mathbf{x}_n - \boldsymbol{\mu}_k)^T \right) \right\} \right] \\ &= 2\pi^{-\frac{N_k d}{2}} |\mathbf{R}_k|^{\frac{N_k}{2}} \\ &\quad \times \exp \left\{ -\frac{1}{2} \text{Tr} \left( \mathbf{R}_k \sum_n^{N_k} (\mathbf{x}_n - \boldsymbol{\mu}_k) (\mathbf{x}_n - \boldsymbol{\mu}_k)^T \right) \right\} \\ &= 2\pi^{-\frac{N_k d}{2}} |\mathbf{R}_k|^{\frac{N_k}{2}} \\ &\quad \times \exp \left\{ -\frac{1}{2} \text{Tr} \left( \mathbf{R}_k (N_k \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T - 2\bar{\mathbf{X}}_k \boldsymbol{\mu}_k^T + \mathbf{C}_k) \right) \right\} \quad (28) \end{aligned}$$

ただし，

$$\begin{aligned} \bar{\mathbf{X}}_k &= \sum_n^{N_k} \mathbf{x}_n \\ \mathbf{C}_k &= \sum_n^{N_k} \mathbf{x}_n \mathbf{x}_n^T \end{aligned} \quad (29)$$

である．

以上より， $\mathbf{X}_k$ ,  $\boldsymbol{\mu}_k$ ,  $\mathbf{R}_k$  の同時確率は次のようになる．

$$\begin{aligned} P(\mathbf{X}_k, \boldsymbol{\mu}_k, \mathbf{R}_k) &= P(\mathbf{X}_k | \boldsymbol{\mu}_k, \mathbf{R}_k) P(\boldsymbol{\mu}_k, \mathbf{R}_k) \\ &= \frac{(2\pi)^{-\frac{N_k d}{2}}}{Z(d, r_0, \nu_0, \mathbf{S}_0)} |\mathbf{R}_k|^{\frac{N_k + \nu_0 - d}{2}} \\ &\quad \times \exp \left[ -\frac{1}{2} \text{Tr} \left[ \mathbf{R}_k \left\{ (r_0 + N_k) \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T \right. \right. \right. \\ &\quad \left. \left. \left. - 2\boldsymbol{\mu}_k (\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0)^T + r_0 \mathbf{m}_0 \mathbf{m}_0^T + \mathbf{C}_k + \mathbf{S}_0 \right\} \right] \right] \quad (30) \end{aligned}$$

ここで ,

$$\begin{aligned}
& (r_0 + N_k) \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T - 2 \boldsymbol{\mu}_k (\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0)^T + r_0 \mathbf{m}_0 \mathbf{m}_0^T + \mathbf{C}_k + \mathbf{S}_0 \\
&= (r_0 + N_k) \left\{ \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T - 2 \boldsymbol{\mu}_k \left( \frac{\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0}{r_0 + N_k} \right) \right. \\
&\quad \left. + \frac{(\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0) (\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0)^T}{(r_0 + N_k)^2} \right\} \\
&\quad - \frac{(\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0) (\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0)^T}{r_0 + N_k} + r_0 \mathbf{m}_0 \mathbf{m}_0^T + \mathbf{C}_k + \mathbf{S}_0 \\
&= r'_k (\boldsymbol{\mu}_k - \mathbf{m}'_k) (\boldsymbol{\mu}_k - \mathbf{m}'_k)^T + \mathbf{S}'_k
\end{aligned} \tag{31}$$

を代入する . ただし ,

$$\begin{aligned}
r'_k &= r_0 + N_k \\
\mathbf{m}'_k &= \frac{\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0}{r_0 + N_k} \\
\mathbf{S}'_k &= -r'_k \mathbf{m}'_k \mathbf{m}'_k{}^T + r_0 \mathbf{m}_0 \mathbf{m}_0^T + \mathbf{C}_k + \mathbf{S}_0
\end{aligned} \tag{32}$$

である . したがって ,

$$\begin{aligned}
P(\mathbf{X}_k, \boldsymbol{\mu}_k, \mathbf{R}_k) &= \frac{(2\pi)^{-\frac{N_k d}{2}}}{Z(d, r_0, \nu_0, \mathbf{S}_0)} |\mathbf{R}_k|^{\frac{\nu'_k - d}{2}} \\
&\quad \times \exp \left[ -\frac{1}{2} \text{Tr} \left[ \mathbf{R}_k \left\{ r'_k (\boldsymbol{\mu}_k - \mathbf{m}'_k) (\boldsymbol{\mu}_k - \mathbf{m}'_k)^T - \mathbf{S}'_k \right\} \right] \right]
\end{aligned} \tag{33}$$

である . ただし ,  $\nu'_k = N_k + \nu_0$  である .

次に ,  $P(\mathbf{X}_k, \boldsymbol{\mu}_k, \mathbf{R}_k)$  を  $\boldsymbol{\mu}_k$  と  $\mathbf{R}_k$  について周辺化し , 周辺尤度  $P(\mathbf{X}_k)$  を求める .



$$\begin{aligned}
P(\mathbf{X}_k) &= \int \int P(\mathbf{X}_k, \boldsymbol{\mu}_k, \mathbf{R}_k) d\boldsymbol{\mu}_k d\mathbf{R}_k \\
&= \frac{(2\pi)^{-\frac{N_k d}{2}}}{Z(d, r_0, \nu_0, \mathbf{S}_0)} \\
&\quad \times \int \int |\mathbf{R}_k|^{\frac{\nu'_k - d}{2}} \exp \left[ -\frac{1}{2} \text{Tr} \left[ \mathbf{R}_k \left\{ r'_k (\boldsymbol{\mu}_k - \mathbf{m}'_k) (\boldsymbol{\mu}_k - \mathbf{m}'_k)^T - \mathbf{S}'_k \right\} \right] \right] d\boldsymbol{\mu}_k d\mathbf{R}_k \\
&= \frac{(2\pi)^{-\frac{N_k d}{2}}}{Z(d, r_0, \nu_0, \mathbf{S}_0)} Z(d, r'_k, \nu'_k, \mathbf{S}'_k) \\
&\quad \times \int \int \frac{1}{Z(d, r'_k, \nu'_k, \mathbf{S}'_k)} |\mathbf{R}_k|^{\frac{\nu'_k - d}{2}} \\
&\quad \times \exp \left[ -\frac{1}{2} \text{Tr} \left[ \mathbf{R}_k \left\{ r'_k (\boldsymbol{\mu}_k - \mathbf{m}'_k) (\boldsymbol{\mu}_k - \mathbf{m}'_k)^T - \mathbf{S}'_k \right\} \right] \right] d\boldsymbol{\mu}_k d\mathbf{R}_k
\end{aligned} \tag{34}$$

式 (34) と式 (25) を比較すると , 式 (34) の積分内が

$$\begin{aligned}
\frac{1}{Z(d, r'_k, \nu'_k, \mathbf{S}'_k)} |\mathbf{R}_k|^{\frac{\nu'_k - d}{2}} \exp \left[ -\frac{1}{2} \text{Tr} \left[ \mathbf{R}_k \left\{ r'_k (\boldsymbol{\mu}_k - \mathbf{m}'_k) (\boldsymbol{\mu}_k - \mathbf{m}'_k)^T - \mathbf{S}'_k \right\} \right] \right] \\
= P(\boldsymbol{\mu}_k, \mathbf{R}_k | d, r'_k, \nu'_k, \mathbf{S}'_k) \tag{35}
\end{aligned}$$

となることが分かる . さらに ,

$$\int \int P(\boldsymbol{\mu}_k, \mathbf{R}_k | d, r'_k, \nu'_k, \mathbf{S}'_k) d\boldsymbol{\mu}_k d\mathbf{R}_k = 1 \tag{36}$$

が成り立つため , 式 (34) は次式のようになる .

$$\begin{aligned}
P(\mathbf{X}_k) &= \frac{(2\pi)^{-\frac{nd}{2}}}{Z(d, r_0, \nu_0, \mathbf{S}_0)} Z(d, r'_k, \nu'_k, \mathbf{S}'_k) \\
&= (2\pi)^{-\frac{N_k d}{2}} \frac{2^{-\frac{(\nu_0+1)d}{2}} \pi^{-\frac{d(d+1)}{2}} r_0^{\frac{d}{2}} |\mathbf{S}_0|^{\frac{\nu_0}{2}} \prod_i^d \Gamma\left(\frac{\nu_0+1-i}{2}\right)^{-1}}{2^{-\frac{(\nu'_k+1)d}{2}} \pi^{-\frac{d(d+1)}{2}} r'_k{}^{\frac{d}{2}} |\mathbf{S}'_k|^{\frac{\nu'_k}{2}} \prod_i^d \Gamma\left(\frac{\nu'_k+1-i}{2}\right)^{-1}} \\
&= (\pi)^{-\frac{N_k d}{2}} \frac{r_0^{\frac{d}{2}} |\mathbf{S}_0|^{\frac{\nu_0}{2}}}{r'_k{}^{\frac{d}{2}} |\mathbf{S}'_k|^{\frac{\nu'_k}{2}}} \prod_i^d \frac{\Gamma\left(\frac{\nu'_k+1-i}{2}\right)}{\Gamma\left(\frac{\nu_0+1-i}{2}\right)} \\
&\propto \frac{(\pi)^{-\frac{N_k d}{2}}}{r'_k{}^{\frac{d}{2}} |\mathbf{S}'_k|^{\frac{\nu'_k}{2}}} \prod_i^d \Gamma\left(\frac{\nu'_k+1-i}{2}\right)
\end{aligned} \tag{37}$$

以上より，新しいサンプルがクラス  $k$  から生成される尤度は次のように計算可能である．

$$P(\boldsymbol{x}_{new}|\boldsymbol{X}_k) = \frac{P(\boldsymbol{x}_{new}, \boldsymbol{X}_k)}{P(\boldsymbol{X}_k)} \quad (38)$$

以上より, Gibbs Sampling によるパラメタ推定アルゴリズムは次のようになる .  
 [Gaussian DPM algorithm]-----

while 収束するまで

for  $n = 0, \dots, N$

1.  $\mathbf{x}_n$  をクラス  $k = z_n$  から削除し, パラメタを更新

$$\bar{\mathbf{X}}_{k-} = \mathbf{x}_n, \quad \mathbf{C}_{k-} = \mathbf{x}_n \mathbf{x}_n^T, \quad (39)$$

$$r'_{k-} = 1, \quad \nu'_{k-} = 1, \quad N_{k-} = 1, \quad (40)$$

$$\mathbf{m}'_k = \frac{\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0}{r_0 + N_k} \quad (41)$$

$$\mathbf{S}'_k = -r'_k \mathbf{m}'_k \mathbf{m}'_k{}^T + r_0 \mathbf{m}_0 \mathbf{m}_0^T + \mathbf{C}_k + \mathbf{S}_0 \quad (42)$$

2. 新たな  $z_n$  を事後分布からサンプリング

$$P(z_n = k | \mathbf{X}_k) \propto \begin{cases} P(\mathbf{x}_n | \mathbf{X}_k) \frac{N_k}{\gamma + N - 1} & (k = 1, \dots, K) \\ P(\mathbf{x}_n | \mathbf{X}_k) \frac{\gamma}{\gamma + N - 1} & (k = K + 1) \end{cases} \quad (43)$$

ただし,

$$r'_{K+1} = r_0, \quad \nu'_{K+1} = \nu_0 \quad (44)$$

$$\mathbf{m}'_{K+1} = \mathbf{m}_0, \quad \mathbf{S}'_{K+1} = \mathbf{S}_0 \quad (45)$$

とする .

3.  $\mathbf{x}_n$  をクラス  $k$  に追加し, パラメタを更新

$$\bar{\mathbf{X}}_{k+} = \mathbf{x}_n, \quad \mathbf{C}_{k+} = \mathbf{x}_n \mathbf{x}_n^T \quad (46)$$

$$r'_{k+} = 1, \quad \nu'_{k+} = 1, \quad N_{k+} = 1 \quad (47)$$

$$\mathbf{m}'_k = \frac{\bar{\mathbf{X}}_k + r_0 \mathbf{m}_0}{r_0 + N_k} \quad (48)$$

$$\mathbf{S}'_k = -r'_k \mathbf{m}'_k \mathbf{m}'_k{}^T + r_0 \mathbf{m}_0 \mathbf{m}_0^T + \mathbf{C}_k + \mathbf{S}_0 \quad (49)$$

新たなクラスが生成された場合  $K+ = 1$  とする .

4. 空のクラスを削除

end for

end while

---

## 参考文献

- [1] 上田修功，山田武士，“ノンパラメトリックベイズモデル”，応用数理，vol.17，pp.196 – 214，2007.
- [2] 持橋大地，“最近のベイズ理論の進展と応用 [iii]：ノンパラメトリックベイズ”，電子情報通信学会誌，vol.93，no.1，pp.73-79，2010.