変分ベイズ法の混合正規分布への適用

中村 友昭

1 一般式の導出

変分ベイズ学習では，全ての未知量を周辺化した次式の周辺尤度を考える．

$$\mathcal{L}(D) = \log p(\mathbf{D}) = \log \sum_m \sum_{\mathbf{Z}} \int p(\mathbf{D}, \mathbf{Z}, \boldsymbol{\theta}, m) d\boldsymbol{\theta} \tag{1}$$

ここで，$\mathbf{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ は N 個の学習データ，$\mathbf{Z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\}$ は潜在変数，$\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I\}$ は I 個のモデルのパラメータ，m はモデル指標を表す．全ての確率変数の結合分布 $p(\mathbf{D}, \mathbf{Z}, \boldsymbol{\theta}, m)$ は

$$p(\mathbf{D}, \mathbf{Z}, \boldsymbol{\theta}, m) = p(m) p(\mathbf{D}, \mathbf{Z}|m) \prod_i p(\boldsymbol{\theta}_i|m) \tag{2}$$

と分解できる．ここで以下のような，各未知量毎に独立性を仮定した，新たな分布 q を導入する．

$$q(\mathbf{Z}, \boldsymbol{\theta}, m) = q(m) q(\mathbf{Z}|m) \prod_i q(\boldsymbol{\theta}_i|m) \tag{3}$$

q を用いて $\mathcal{L}(\mathbf{D})$ を以下のように変形する．

$$\begin{aligned}
\mathcal{L}(\mathbf{D}) &= \log \sum_m \sum_{\mathbf{Z}} \int q(\mathbf{Z}, \boldsymbol{\theta}, m) \frac{p(\mathbf{D}, \mathbf{Z}, \boldsymbol{\theta}, m)}{q(\mathbf{Z}, \boldsymbol{\theta}, m)} d\boldsymbol{\theta} \\
&= \sum_m \sum_{\mathbf{Z}} \int_\theta q(\mathbf{Z}, \boldsymbol{\theta}, m) \log \frac{q(\mathbf{Z}, \boldsymbol{\theta}, m)}{p(\mathbf{Z}, \boldsymbol{\theta}, m|\mathbf{D})} d\boldsymbol{\theta} \\
&\quad + \sum_m \sum_{\mathbf{Z}} \int_\theta q(\mathbf{Z}, \boldsymbol{\theta}, m) \log \frac{p(\mathbf{D}, \mathbf{Z}, \boldsymbol{\theta}, m)}{q(\mathbf{Z}, \boldsymbol{\theta}, m)} d\boldsymbol{\theta} \\
&\equiv \mathrm{KL}(q(\mathbf{Z}, \boldsymbol{\theta}, m), p(\mathbf{Z}, \boldsymbol{\theta}, m|\mathbf{D})) + \mathcal{F}[q]
\end{aligned} \tag{4}$$

ただし，$\mathcal{F}[q]$ は自由エネルギーと呼ばれ，分布 q を変関数とする汎関数，KL は分布間の距離を表すカルバック・ライブラー距離 (KL 距離) である．ここで，$\mathcal{L}(D)$ が変分事後分布 q に依存しないため，$F[q]$ を最大化することは，q と真の事後分布

p との KL 距離を最少化することと同義である．すなわち，$\mathcal{F}[q]$ を最大化することで，変分事後分布 q は真の事後分布 p の最良の近似となる．

式 (3) を用いて，$\mathcal{F}(q)$ を以下のように変形する．

$$\begin{aligned}\mathcal{F}[q] = \sum_m q(m) \Bigg\{ & \left\langle \log \frac{p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m)}{q(\mathbf{Z}|m)} \right\rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}|m)} \\ & + \sum_{i=1}^{I} \left\langle \log \frac{p(\boldsymbol{\theta}_i|m)}{q(\boldsymbol{\theta}_i|m)} \right\rangle_{q(\boldsymbol{\theta}_i|m)} + \log \frac{p(m)}{q(m)} \Bigg\} \end{aligned} \tag{5}$$

1.1 潜在変数の変分事後分布 $q(\mathrm{Z}|m)$ の導出

$q(\mathbf{Z}|m)$ は，制約条件 $\sum_{\mathbf{Z}} q(\mathbf{Z}|m) = 1$ の下で，$\mathcal{F}(q)$ を最大化することで得られる．最大化にはラグランジュの未定乗数法を用いる．式 (5) から，$q(\mathbf{Z}|m)$ に関する項を

$$\mathcal{F}[q(\mathbf{Z}|m)] = \left\langle \log \frac{p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m)}{q(\mathbf{Z}|m)} \right\rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}|m)} \tag{6}$$

とし，λ をラグランジュ乗数として，次式の汎関数 $\mathcal{J}[q(\mathbf{Z}|m)]$ の極値問題を解く．

$$\mathcal{J}[q(\mathbf{Z}|m)] = \mathcal{F}[q(\mathbf{Z}|m)] + \lambda(\sum_{\mathbf{Z}} q(\mathbf{Z}|m) - 1) \tag{7}$$

$$\begin{aligned}\frac{\partial \mathcal{J}[q(\mathbf{Z}|m)]}{\partial q(\mathbf{Z}|m)} &= \frac{\partial}{\partial q(\mathbf{Z}|m)} \Bigg[\sum_{\mathbf{Z}} \left\{ q(\mathbf{Z}|m) \int q(\boldsymbol{\theta}|m) \log \frac{p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m)}{q(\mathbf{Z}|m)} d\boldsymbol{\theta} \right\} \\ &\quad + \lambda(\sum_{\mathbf{Z}} q(\mathbf{Z}|m) - 1) \Bigg] \\ &= \frac{\partial}{\partial q(\mathbf{Z}|m)} \int q(\boldsymbol{\theta}|m) \{ q(\mathbf{Z}|m) \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \\ &\quad - q(\mathbf{Z}|m) \log q(\mathbf{Z}|m) \} d\boldsymbol{\theta} + \lambda \\ &= \int q(\boldsymbol{\theta}|m) \{ \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) - \log q(\mathbf{Z}|m) - 1 \} d\boldsymbol{\theta} + \lambda \\ &= \langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\boldsymbol{\theta}|m)} - \log q(\mathbf{Z}|m) - 1 + \lambda \\ &= 0 \end{aligned}$$

$$\therefore q(\mathbf{Z}|m) = \exp \left\{ \langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\boldsymbol{\theta}|m)} + \lambda - 1 \right\} \tag{8}$$

$$
\begin{aligned}
\frac{\partial \mathcal{J}[q(\mathbf{Z}|m)]}{\partial \lambda} &= \sum_{\mathbf{Z}} q(\mathbf{Z}|m) - 1 \\
&= \sum_{\mathbf{Z}} \exp\left\{ \langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\boldsymbol{\theta}|m)} + \lambda - 1 \right\} - 1 = 0 \\
& \therefore \exp(\lambda - 1) = \frac{1}{\sum_{\mathbf{Z}} \exp \langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\boldsymbol{\theta}|m)}} \qquad (9)
\end{aligned}
$$

式 (8), 式 (9) より，潜在変数の変分事後分布 $q(\mathbf{Z}|m)$ を得る．

$$
\begin{aligned}
q(\mathbf{Z}|m) &= \frac{\exp \langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\boldsymbol{\theta}|m)}}{\sum_{\mathbf{Z}} \exp \langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\boldsymbol{\theta}|m)}} \\
&= C \exp \langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\boldsymbol{\theta}|m)} \qquad (10)
\end{aligned}
$$

C は，規格化定数である．

1.2 パラメータの変分事後分布 $q(\boldsymbol{\theta}_i|m)$ の導出

前節と同様に，$q(\boldsymbol{\theta}_i|m)$ は，制約条件 $\sum_{\theta_i} q(\boldsymbol{\theta}_i|m) = 1$ の下で，$\mathcal{F}(q)$ を最大化することで得られる．式 (5) から，$q(\boldsymbol{\theta}_i|m)$ に関する項を

$$
\mathcal{F}[q(\boldsymbol{\theta}_i|m)] = \left\langle \log \frac{p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m)}{q(\mathbf{Z}|m)} \right\rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}|m)} + \sum_j \left\langle \log \frac{p(\boldsymbol{\theta}_j|m)}{q(\boldsymbol{\theta}_j|m)} \right\rangle_{q(\boldsymbol{\theta}_j|m)} \qquad (11)
$$

とし，λ をラグランジュ乗数として，次式の汎関数 $\mathcal{J}[q(\boldsymbol{\theta}_i|m)]$ の極値問題を解く．

$$
\begin{aligned}
\mathcal{J}[q(\boldsymbol{\theta}_i|m)] &= \mathcal{F}[q(\boldsymbol{\theta}_i|m)] + \lambda(\int q(\boldsymbol{\theta}_i|m) d\boldsymbol{\theta}_i - 1) \\
\frac{\partial \mathcal{J}[q(\boldsymbol{\theta}_i|m)]}{\partial q(\boldsymbol{\theta}_i|m)} &= \frac{\partial}{\partial q(\boldsymbol{\theta}_i|m))} \left[\sum_{\mathbf{Z}} \left\{ q(\mathbf{Z}|m) \int q(\boldsymbol{\theta}|m) \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \right.\right. \\
& \quad -q(\boldsymbol{\theta}|m) \log q(\mathbf{Z}|m) d\boldsymbol{\theta} \} \\
& \quad +q(\boldsymbol{\theta}_i|m) \log p(\boldsymbol{\theta}_i|m) - q(\boldsymbol{\theta}_i|m) \log q(\boldsymbol{\theta}_i|m)] + \lambda \\
&= \langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}_{-i}|m)} + \log p(\boldsymbol{\theta}_i|m) \\
& \quad - \log q(\boldsymbol{\theta}_i|m) - 1 = 0 \qquad (12)
\end{aligned}
$$

$$
\therefore q(\boldsymbol{\theta}_i|m) = p(\boldsymbol{\theta}_i|m) \exp \Big(\langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}_{-i}|m)} + \lambda - 1 \Big) \qquad (13)
$$

$$\begin{aligned}\frac{\partial \mathcal{J}[q(\boldsymbol{\theta}_i|m)]}{\partial \lambda} &= \int q(\boldsymbol{\theta}_i|m) d\boldsymbol{\theta}_i - 1 \\ &= \int p(\boldsymbol{\theta}_i|m) \exp\Big(\langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}_{-i}|m)} \\ &\quad +\lambda - 1 \Big) d\boldsymbol{\theta}_i - 1 = 0 \end{aligned} \tag{14}$$

$$\therefore \exp(\lambda - 1) = \frac{1}{\int p(\boldsymbol{\theta}_i|m) \exp\Big(\langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}_{-i}|m)} \Big) d\boldsymbol{\theta}_i} \tag{15}$$

ただし，$\boldsymbol{\theta}_{-i}$ は，$\boldsymbol{\theta}$ 中の，$\boldsymbol{\theta}_i$ 以外のパラメータの集合を表す.
式 (13) に式 (15) を代入することでパラメータの変分事後分布 $q(\boldsymbol{\theta}_i|m)$ を得る.

$$\begin{aligned} q(\boldsymbol{\theta}_i|m) &= \frac{p(\boldsymbol{\theta}_i|m) \exp\Big(\langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}_{-i}|m)} \Big)}{\int p(\boldsymbol{\theta}_i|m) \exp\Big(\langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}_{-i}|m)} \Big) d\boldsymbol{\theta}_i} \\ &= C_i p(\boldsymbol{\theta}_i|m) \exp\Big(\langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \rangle_{q(\mathbf{Z}|m), q(\boldsymbol{\theta}_{-i}|m)} \Big) \end{aligned} \tag{16}$$

ただし，C_i は規格化定数である.

1.3 モデル指標の変分事後分布の $q(m)$ の導出

$\mathcal{F}[q]$ の $q(m)$ を含まない項をまとめて $\mathcal{F}_m$ と置き，式 (5) を変形すると次式を得る.

$$\mathcal{F}[q] = \langle \mathcal{F}_m \rangle_{q(m)} + \left\langle \log \frac{p(m)}{q(m)} \right\rangle_{q(m)} \tag{17}$$

$q(\mathbf{Z}|m)$，$q(\boldsymbol{\theta}_i|m)$ は，式 (17) の $\mathcal{F}_m$ のみに依存するので，$\mathcal{F}[q]$ の $q(\mathbf{Z}|m)$，$q(\boldsymbol{\theta}_i|m)$ に関する最大化は，$\mathcal{F}_m$ の $q(\mathbf{Z}|m)$，$q(\boldsymbol{\theta}_i|m)$ に関する最大化と等価である．$\mathcal{F}_m$ の $q(\mathbf{Z}|m)$，$q(\boldsymbol{\theta}_i|m)$ に関する最大値を $\mathcal{F}_m^*$ と置くと，式 (17) は次のようになる.

$$\mathcal{F}[q] = \langle \mathcal{F}_m^* \rangle_{q(m)} + \left\langle \log \frac{p(m)}{q(m)} \right\rangle_{q(m)} \tag{18}$$

これを $\sum_m q(m) = 1$ の制約条件の下で，$q(m)$ に関して最大化することで m の最適値を得ることができる．ラグランジュ乗数を λ と置くと，次式の汎関数 $\mathcal{J}[q(m)]$

の極値問題となる．

$$\begin{aligned}
\mathcal{J}[q(m)] &= \langle \mathcal{F}_m^* \rangle_{q(m)} + \left\langle \log \frac{p(m)}{q(m)} \right\rangle_{q(m)} + \lambda \left(\sum_m q(m) - 1 \right) \\
&= \sum_m q(m) \left(\mathcal{F}_m^* + \log \frac{p(m)}{q(m)} \right) + \lambda \left(\sum_m q(m) - 1 \right) \\
\frac{\partial \mathcal{J}[q(m)]}{\partial q(m)} &= \mathcal{F}_m^* + \log p(m) - \log q(m) - 1 + \lambda = 0 \\
& \therefore q(m) = p(m) \exp\left(\mathcal{F}_m^* - 1 + \lambda\right) \qquad (19) \\
\frac{\partial \mathcal{J}[q(m)]}{\partial \lambda} &= \sum_m q(m) - 1 = 0 \\
& \therefore \exp(\lambda - 1) = \frac{1}{\sum_m p(m) \exp\left(\mathcal{F}_m^*\right)} \qquad (20)
\end{aligned}$$

式 (20) を，式 (19) に代入することで，モデル指標の変分事後分布 $q(m)$ を得る．

$$q(m) = \frac{p(m) \exp\left(\mathcal{F}_m^*\right)}{\sum_m p(m) \exp\left(\mathcal{F}_m^*\right)} = C_m p(m) \exp\left(\mathcal{F}_m^*\right) \tag{21}$$

ただし，C_m は規格化定数である．さらに簡単のため，事前分布 $p(m) =$ (一様) と仮定すると，$q(m)$ の最大化は，$\mathcal{F}_m$ の最大化と等価である．すなわち，$\mathcal{F}_m$ を $q(\mathbf{Z}|m)$, $q(\boldsymbol{\theta}_i|m)$ に関して最大化すると同時に，m に関しても最大化することにより，最適なモデル指標 m が求まる．

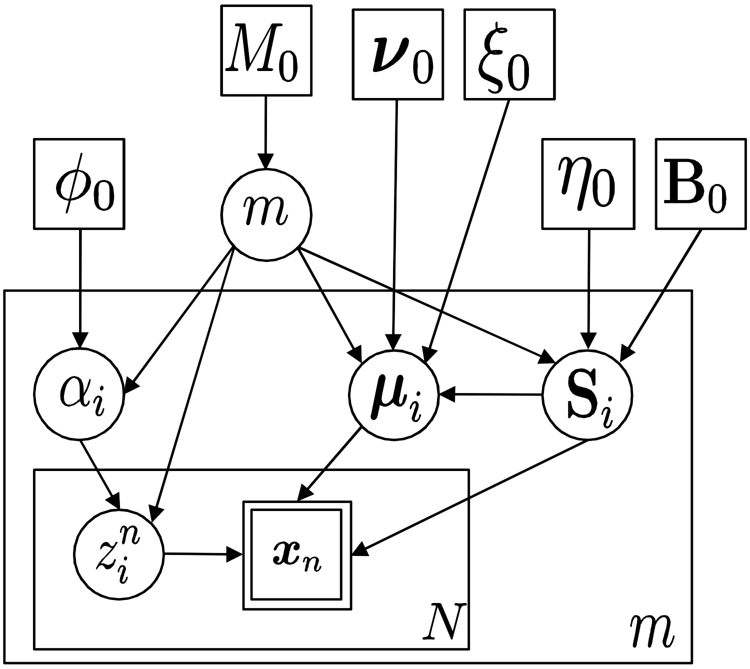

図 1: 混合正規分布のグラフィカルモデル

2　混合正規分布への適用

2.1　混合正規分布の確率モデル

図 1 に混合正規分布のグラフィカルモデルを示す．内側のプレートは学習データ数 N に対する反復計算を表し，$\boldsymbol{x}_n (n = 1, \cdots, N)$ は可観測の学習データを表す．また，$\boldsymbol{\mu}_i$，$\mathbf{S}_i$，α_i はそれぞれ，混合正規分布の第 i 要素 $(i = 1, \cdots, m)$ の平均，精度行列，混合比を表している．m はモデルの構造である混合数である．これらのパラメータはそれぞれ確率分布として与えられ，さらに各パラメータはハイパーパラメータとして共役事前分布を有している．$\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_m\}$ はディリクレ分布，$\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_m\}$ は正規分布，$\mathbf{S} = \{\mathbf{S}_1, \cdots, \mathbf{S}_m\}$ はウィシャート分布，混合数 m は一様分布をそれぞれ仮定している．

N 個の d 次元の学習データ集合を $\mathbf{D} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$ とする．各データ $\boldsymbol{x}_n$ が第 i 要素から生成されているならば，$z_i^n = 1$ さもなくば $z_i^n = 0$ となる潜在変数 $\mathbf{Z} = \{z_i^n\}_{n=1,i=1}^{N,m}$ を導入すると，完全データ集合の混合正規分布の結合分布は以下のようになる．

$$
\begin{aligned}
p(\mathbf{D},\mathbf{Z}|\boldsymbol{\theta},m) &= \prod_i^m \prod_n^N \left\{\alpha_i \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_i,\mathbf{S}_i^{-1})\right\}^{z_i^n} \\
&= \prod_i^m \prod_n^N \left\{\alpha_i (2\pi)^{-d/2}|\mathbf{S}_i|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}_n-\boldsymbol{\mu}_i)^T\mathbf{S}_i(\boldsymbol{x}_n-\boldsymbol{\mu}_i)\right)\right\}^{z_i^n} \\
&= \prod_i^m \prod_n^N \left\{\alpha_i (2\pi)^{-d/2}|\mathbf{S}_i|^{1/2} \exp\left(-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i(\boldsymbol{x}_n-\boldsymbol{\mu}_i)(\boldsymbol{x}_n-\boldsymbol{\mu}_i)^T\right)\right)\right\}^{z_i^n} \quad (22)
\end{aligned}
$$

$\mathcal{N}$ は正規分布を表し，右辺 2 式目から 3 式目への変形は,

$$
\boldsymbol{a}^T\mathbf{A}\boldsymbol{a} = \mathrm{Tr}\left(\mathbf{A}\boldsymbol{a}\boldsymbol{a}^T\right) \quad (23)
$$

を用いている．前節で計算した一般式を再掲する．

$$
q(\mathbf{Z}|m) \propto \exp\left\langle \log p(\boldsymbol{D},\mathbf{Z}|\boldsymbol{\theta},m)\right\rangle_{q(\boldsymbol{\theta}|m)} \quad (24)
$$

$$
\text{For} \quad i = 1,\dots,I
$$

$$
q(\boldsymbol{\theta}_i|m) \propto p(\boldsymbol{\theta}_i|m)\exp\left(\left\langle \log p(\boldsymbol{D},\mathbf{Z}|\boldsymbol{\theta},m)\right\rangle_{q(\mathbf{Z}|m),q(\boldsymbol{\theta}_{-i}|m)}\right) \quad (25)
$$

ただし，I は独立なパラメータの個数を表し，混合正規分布モデルの場合，式 (25) は次式となる．

$$
q(\boldsymbol{\alpha}|m) \propto p(\boldsymbol{\alpha}|m)\exp\left(\left\langle \log p(\boldsymbol{D},\mathbf{Z}|\boldsymbol{\theta},m)\right\rangle_{q(\mathbf{Z}|m),q(\boldsymbol{\mu},\mathbf{S}|m)}\right) \quad (26)
$$

$$
q(\boldsymbol{\mu},\mathbf{S}|m) \propto p(\boldsymbol{\mu},\mathbf{S}|m)\exp\left(\left\langle \log p(\boldsymbol{D},\mathbf{Z}|\boldsymbol{\theta},m)\right\rangle_{q(\mathbf{Z}|m),q(\boldsymbol{\alpha}|m)}\right) \quad (27)
$$

ここでの目標は，この式 (24)，式 (26)，式 (27) を具体化することである．

2.2 $\langle \log p(\mathbf{D},\mathbf{Z}|\boldsymbol{\theta},m)\rangle_{q(\mathbf{Z}|m)}$ の導出

まず，式 (26) と式 (27) に共通に現れる，完全データの対数尤度を計算する．

$$
\begin{aligned}
&\log p(\mathbf{D},\mathbf{Z}|\boldsymbol{\theta},m) \\
&\quad \propto \sum_i^m \sum_n^N z_i^n \left[\log\alpha_i + \log|\mathbf{S}_i|^{1/2} - \frac{1}{2}\mathrm{Tr}\left\{\mathbf{S}_i(\boldsymbol{x}_n-\boldsymbol{\mu}_i)(\boldsymbol{x}_n-\boldsymbol{\mu}_i)^T\right\}\right] \quad (28)
\end{aligned}
$$

さらに，式 (28) の $q(\mathbf{Z}|m)$ に関する期待値をとり式変形を行う．

$$
\begin{aligned}
\langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m)\rangle_{q(\mathbf{Z}|m)} &= \int q(\mathbf{Z}|m) \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) d\mathbf{Z} \\
&= \int q(\mathbf{Z}|m) \sum_i^m \sum_n^N z_i^n \left[\log \alpha_i + \frac{1}{2} \log |\mathbf{S}_i| \right. \\
&\quad \left. -\frac{1}{2} \mathrm{Tr} \left\{ \mathbf{S}_i \left(\boldsymbol{x}_n - \boldsymbol{\mu}_i\right) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_i\right)^T \right\} \right] d\mathbf{Z} \\
&= \int \sum_i^m \sum_n^N q(\mathbf{Z}|m) \left[z_i^n \log \alpha_i + \frac{1}{2} z_i^n \log |\mathbf{S}_i| \right] d\mathbf{Z} \\
&\quad - \int \sum_i^m \sum_n^N \frac{1}{2} q(\mathbf{Z}|m) z_i^n \mathrm{Tr} \left\{ \mathbf{S}_i \left(\boldsymbol{x}_n - \boldsymbol{\mu}_i\right) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_i\right)^T \right\} d\mathbf{Z}
\end{aligned}
\tag{29}
$$

以下，各項に分解し計算を行う．

$$
\begin{aligned}
(\text{第 1 項目}) &= \sum_i^m \left[\log \alpha_i \sum_n^N \int q(\mathbf{Z}|m) z_i^n d\mathbf{Z} \right. \\
&\quad \left. + \frac{1}{2} \log |\mathbf{S}_i| \sum_n^N \int q(\mathbf{Z}|m) z_i^n d\mathbf{Z} \right] \\
&= \sum_i^m \left[\bar{N}_i \log \alpha_i + \bar{N}_i \frac{1}{2} \log |\mathbf{S}_i| \right] \\
&= \sum_i^m \bar{N}_i (\log \alpha_i + \frac{1}{2} \log |\mathbf{S}_i|)
\end{aligned}
\tag{30}
$$

ただし，$\bar{z}_i^n = \int q(\mathbf{Z}|m) z_i^n d\mathbf{Z}$，$\bar{N}_i = \sum_n^N z_i^n$ とした．

$$
\begin{aligned}
(\text{第 2 項}) &= -\sum_i^m \sum_n^N \int \frac{1}{2} q(\mathbf{Z}|m) z_i^n \mathrm{Tr}\left\{\mathbf{S}_i \left(\boldsymbol{x}_n - \boldsymbol{\mu}_i\right) \left(\boldsymbol{x}_n - \boldsymbol{\mu}_i\right)^T\right\} d\mathbf{Z} \\
&= -\frac{1}{2} \sum_i^m \mathrm{Tr}\left[\sum_n^N \int q(\mathbf{Z}|m) z_i^n \mathbf{S}_i \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T\right. \\
&\quad + q(\mathbf{Z}|m) z_i^n \mathbf{S}_i \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right)^T \\
&\quad \left. - 2 q(\mathbf{Z}|m) z_i^n \mathbf{S}_i (\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i)(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i) d\mathbf{Z}\right] \\
&= -\frac{1}{2} \sum_i^m \mathrm{Tr}\left[\mathbf{S}_i \bar{N}_i \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T\right. \\
&\quad + \mathbf{S}_i \sum_n^N \bar{z}_i^n \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right)^T \\
&\quad \left. - 2 \mathbf{S}_i \left\{\sum_n^N \bar{z}_i^n (\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i)\right\} \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T\right] \\
&= \sum_i^m -\frac{1}{2} \mathrm{Tr}\left[\mathbf{S}_i \left\{\bar{N}_i \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T + \bar{\mathbf{C}}_i\right\}\right] \qquad (31)
\end{aligned}
$$

ただし，

$$
\bar{\boldsymbol{x}}_i = \frac{1}{\bar{N}_i} \sum_n^N \bar{z}_i^n \boldsymbol{x}_i \qquad (32)
$$

とし，右辺 1 式目から 2 式目への変形は以下の式変形を用いた．

$$
\begin{aligned}
(\boldsymbol{x}_n - \boldsymbol{\mu}_i)(\boldsymbol{x}_n - \boldsymbol{\mu}_i)^T &= \left((\boldsymbol{\mu}_i - \boldsymbol{x}_n) - (\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i)\right) \left((\boldsymbol{\mu}_i - \boldsymbol{x}_n) - (\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i)\right)^T \\
&= \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T + \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right)^T \\
&\quad -2 \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right)^T \qquad (33)
\end{aligned}
$$

また，右辺 3 式目から 4 式目への変形では，$\bar{\mathbf{C}}_i = \sum_n^N \bar{z}_i^n \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i\right)^T$ とおき，さらに以下のような変形を用いた．

$$
\begin{aligned}
\sum_n^N \bar{z}_i^n (\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i) &= \bar{N}_i \frac{1}{\bar{N}_i} \sum_n^N \bar{z}_i^n \boldsymbol{x}_n - \bar{\boldsymbol{x}}_i \sum_n^N \bar{z}_i^n \\
&= \bar{N}_i \bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_i \bar{N}_i = 0 \qquad (34)
\end{aligned}
$$

以上より完全データの $q(\mathbf{Z}|m)$ に関する期待値は以下のようになる．

$$
\begin{aligned}
\left\langle \log p(\mathbf{D}, \mathbf{Z}|\boldsymbol{\theta}, m) \right\rangle_{q(\mathbf{Z}|m)} &= \sum_i^m \left\{\bar{N}_i (\log \alpha_i + \frac{1}{2} \log |\mathbf{S}_i|)\right. \\
&\quad \left. -\frac{1}{2} \mathrm{Tr}\left[\mathbf{S}_i \left\{\bar{N}_i \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right) \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T + \bar{\mathbf{C}}_i\right\}\right]\right\} \qquad (35)
\end{aligned}
$$

2.3　$q(\boldsymbol{\alpha}|m)$ の導出

次に，式 (26) の導出を行う．混合正規分布のモデルでは，$\boldsymbol{\alpha}$ の事前分布として，次式のディリクレ分布を仮定している．

$$p(\boldsymbol{\alpha}|m) = \mathcal{D}(\{\alpha_i\}_{i=1}^{m}|\phi_0) \propto \prod_{i=1}^{m} \alpha_i^{\phi_o - 1} \tag{36}$$

ただし，$\mathcal{D}()$ はディリクレ分布を表している．式 (26) に，式 (35) と式 (36) を代入し $\boldsymbol{\alpha}$ に関する項のみを取り出すと次式になる．

$$\begin{aligned} q(\boldsymbol{\alpha}|m) &\propto \left(\prod_{i=1}^{m} \alpha_i^{\phi_o-1}\right) \exp\left(\int\int q(\boldsymbol{\mu}, \mathbf{S}|m) \sum_{i}^{m} \bar{N}_i \log \alpha_i d\boldsymbol{\mu} d\mathbf{S}\right) \\ &= \left(\prod_{i=1}^{m} \alpha_i^{\phi_o-1}\right) \exp\left(\sum_{i}^{m} \bar{N}_i \log \alpha_i\right) \\ &= \prod_{i}^{m} \alpha_i^{\phi_0 + \bar{N}_i - 1} = \mathcal{D}(\{\alpha_i\}_{i=1}^{m}|\{\phi_0 + \bar{N}_i\}_{i=1}^{m}) \end{aligned} \tag{37}$$

2.4　$q(\boldsymbol{\mu}|m)$，$q(\mathbf{S}|m)$ の導出

次に式 (27) の導出を行う．混合正規分布のモデルでは，$\boldsymbol{\mu}$ と $\mathbf{S}$ の事前分布として，それぞれ正規分布とウィシャート分布を仮定している．

$$p(\boldsymbol{\mu}, \mathbf{S}) = \prod_{i=1}^{m} \mathcal{N}(\boldsymbol{\mu}_i|\boldsymbol{\nu}_0, (\xi_0 \mathbf{S}_i)^{-1}) \mathcal{W}(\mathbf{S}_i|\eta_0, \mathbf{B}_0) \tag{38}$$

ただし，$\mathcal{W}()$ はウィシャート分布を表し，

$$\mathcal{W}(\mathbf{S}_i|\eta_0, \mathbf{B}_0) \propto |\mathbf{S}_i|^{1/2(\eta_0 - d - 1)} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i \mathbf{B}_0\right)\right\} \tag{39}$$

となる．式 (27) に式 (35) を代入し $\boldsymbol{\mu}$，$\mathbf{S}$ に関する項のみを取り出すと次式のようになる．

$$
\begin{aligned}
q(\boldsymbol{\mu}, \mathbf{S}|m) \;\propto\; & \left\{\prod_{i=1}^{m} \mathcal{N}(\boldsymbol{\mu}_i|\boldsymbol{\nu}_0, (\xi_0 \mathbf{S}_i)^{-1}) \mathcal{W}(\mathbf{S}_i|\eta_0, \mathbf{B}_0)\right\} \exp\left[\int \sum_i^m \left[\frac{1}{2}\bar{N}_i \log|\mathbf{S}_i|\right.\right. \\
& \left.\left. -\frac{1}{2}\mathrm{Tr}\left[\mathbf{S}_i\left\{\bar{N}_i\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T + \bar{\mathbf{C}}_i\right\}\right] q(\boldsymbol{\alpha}|m) d\boldsymbol{\alpha}\right]\right] \\
\propto\; & \prod_i^m \left[|\mathbf{S}_i|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_i - \boldsymbol{\nu}_0)^T \xi_0 \mathbf{S}_i (\boldsymbol{\mu}_i - \boldsymbol{\nu}_0)\right\}\right. \\
& \times |\mathbf{S}_i|^{1/2(\eta_0 - d - 1)} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i \mathbf{B}_0\right)\right\} \\
& \times \exp\left\{\frac{1}{2}\bar{N}_i \log|\mathbf{S}_i|\right. \\
& \left.\left. -\frac{1}{2}\mathrm{Tr}\left[\mathbf{S}_i\left(\bar{N}_i\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T + \bar{\mathbf{C}}_i\right)\right]\right\}\right] \\
=\; & \prod_i^m \left[|\mathbf{S}_i|^{1/2(\eta_0 + \bar{N}_i - d)}\right. \\
& \exp\left[-\frac{1}{2}\mathrm{Tr}\left\{\mathbf{S}_i\left(\xi_0\left(\boldsymbol{\mu}_i - \boldsymbol{\nu}_0\right)\left(\boldsymbol{\mu}_i - \boldsymbol{\nu}_0\right)^T\right.\right.\right. \\
& \left.\left.\left.\left. + \mathbf{B}_0 + \bar{N}_i\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T + \bar{\mathbf{C}}_i\right)\right\}\right]\right]
\end{aligned} \tag{40}
$$

右辺，2 式目から 3 式目への変形では，式 (23) を用いている．さらにここで，以下の式変形を使用する．

$$
\begin{aligned}
\bar{N}_i\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{x}}_i\right)^T \;+\; & \xi_0\left(\boldsymbol{\mu}_i - \boldsymbol{\nu}_0\right)\left(\boldsymbol{\mu}_i - \boldsymbol{\nu}_0\right)^T = \\
& (\bar{N}_i + \xi_0)\left(\boldsymbol{\mu}_i - \frac{\bar{N}_i \bar{\boldsymbol{x}}_i + \xi_0 \nu_0}{\bar{N}_i + \xi_0}\right)\left(\boldsymbol{\mu}_i - \frac{\bar{N}_i \bar{\boldsymbol{x}}_i + \xi_0 \nu_0}{\bar{N}_i + \xi_0}\right)^T \\
& + \frac{\bar{N}_i \xi_0}{\bar{N}_i + \xi_0}\left(\bar{\boldsymbol{x}}_i - \boldsymbol{\nu}_0\right)\left(\bar{\boldsymbol{x}}_i - \boldsymbol{\nu}_0\right)^T
\end{aligned} \tag{41}
$$

$$
\begin{aligned}
q(\boldsymbol{\mu}, \mathbf{S}|m) &\propto \prod_{i}^{m} \Big[|\mathbf{S}_i|^{1/2(\eta_0+\bar{N}_i-d)} \\
&\quad \times \exp\left[-\frac{1}{2}\mathrm{Tr}\left\{ \mathbf{S}_i \left((\bar{N}_i+\xi_0)\left(\boldsymbol{\mu}_i - \frac{\bar{N}_i\bar{\boldsymbol{x}}_i+\xi_0\nu_0}{\bar{N}_i+\xi_0}\right)\left(\boldsymbol{\mu}_i - \frac{\bar{N}_i\bar{\boldsymbol{x}}_i+\xi_0\nu_0}{\bar{N}_i+\xi_0}\right)^T \right.\right.\right. \\
&\quad \left.\left.\left. + \frac{\bar{N}_i\xi_0}{\bar{N}_i+\xi_0}(\bar{\boldsymbol{x}}_i-\boldsymbol{\nu}_0)(\bar{\boldsymbol{x}}_i-\boldsymbol{\nu}_0)^T + \mathbf{B}_0 + \bar{\mathbf{C}}_i \right)\right\}\right]\Big] \\
&= \prod_{i}^{m} \Big[|\mathbf{S}_i|^{1/2(\eta_0+\bar{N}_i-d)} \\
&\quad \times \exp\left[-\frac{1}{2}\mathrm{Tr}\left\{ \mathbf{S}_i \left((\bar{N}_i+\xi_0)(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}}_i)(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}}_i)^T + \mathbf{B}_i \right)\right\}\right]\Big] \\
&= \prod_{i}^{m} \left[|\mathbf{S}_i|^{1/2} \exp\left[-\frac{1}{2}\mathrm{Tr}\left\{ \mathbf{S}_i \left((\bar{N}_i+\xi_0)(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}}_i)(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}}_i)^T \right)\right\}\right]\right] \\
&\quad \times \prod_{i}^{m} \left[|\mathbf{S}_i|^{1/2(\eta_0+\bar{N}_i-d-1)} \exp\left\{ -\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i\mathbf{B}_i\right)\right\}\right]
\end{aligned}
\tag{42}
$$

ただし，

$$
\bar{\boldsymbol{\mu}}_i = \frac{\bar{N}_i\bar{\boldsymbol{x}}_i+\xi_0\boldsymbol{\nu}_0}{\bar{N}_i+\xi_0} \tag{43}
$$

$$
\mathbf{B}_i = \frac{\bar{N}_i\xi_0}{\bar{N}_i+\xi_0}(\bar{\boldsymbol{x}}_i-\boldsymbol{\nu}_0)(\bar{\boldsymbol{x}}_i-\boldsymbol{\nu}_0)^T + \mathbf{B}_0 + \bar{\mathbf{C}}_i \tag{44}
$$

とする．これは，定数項を無視すれば，正規分布とウィシャート分布の積と見ることができる．すなわち，$\boldsymbol{\mu}_i$ と $\mathbf{S}_i$ の最適事後分布は以下のようになる．

$$
q(\boldsymbol{\mu}_i|\mathbf{S}_i, m) = \mathcal{N}(\boldsymbol{\mu}_i|\bar{\boldsymbol{\mu}}_i, ((\bar{N}_i+\xi_0)\mathbf{S}_i)^{-1}) \tag{45}
$$

$$
q(\mathbf{S}_i|m) = \mathcal{W}(\mathbf{S}_i|\eta_0+\bar{N}_i, \mathbf{B}_i) \tag{46}
$$

さらに，式 (42) を $\mathbf{S}_i$ に関して周辺化することにより，$q(\boldsymbol{\mu}_i|m)$ を得る．

$$
\begin{aligned}
q(\boldsymbol{\mu}_i|m) &= \int q(\boldsymbol{\mu}_i, \mathbf{S}_i|m)d\mathbf{S}_i \\
&= \int |\mathbf{S}_i|^{1/2(\eta_0+\bar{N}_i-d)} \\
&\quad \times \exp\left\{ -\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i \left((\bar{N}_i+\xi_0)(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}}_i)(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}}_i)^T + \mathbf{B}_i \right)\right)\right\} d\mathbf{S}_i
\end{aligned}
\tag{47}
$$

ここで，

$$\mathbf{B}_i' = (\bar{N}_i + \xi_0)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i\right)^T + \mathbf{B}_i \tag{48}$$

$$\eta_i' = \eta_0 + \bar{N}_i + 1 \tag{49}$$

とおき式変形すると，

$$q(\boldsymbol{\mu}_i|m) = \int |\mathbf{S}_i|^{1/2(\eta_i'-d-1)} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i\mathbf{B}_i'\right)\right\} d\mathbf{S}_i \tag{50}$$

となる．ここで，$\int \mathcal{W}(\mathbf{S}_i|\eta_i', \mathbf{B}_i')d\mathbf{S}_i = 1$ が成り立つので，

$$\begin{aligned}\int \mathcal{W}(\mathbf{S}_i|\eta_i', \mathbf{B}_i')d\mathbf{S}_i &= \int |\mathbf{B}_i'|^{\eta_i'/2}\left(2^{\frac{\eta_i' d}{2}}\pi^{\frac{d(d-1)}{4}}\prod_j^d \Gamma\left(\frac{\eta_i'+i-j}{2}\right)\right)^{-1} \\ &\quad \times |\mathbf{S}_i|^{1/2(\eta_i'-d-1)}\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i\mathbf{B}_i'\right)\right)d\mathbf{S}_i \\ &= |\mathbf{B}_i'|^{\eta_i'/2}\left(2^{\frac{\eta_i' d}{2}}\pi^{\frac{d(d-1)}{4}}\prod_j^d \Gamma\left(\frac{\eta_i'+i-j}{2}\right)\right)^{-1} q(\boldsymbol{\mu}_i|m) \\ &= 1\end{aligned} \tag{51}$$

となる．したがって，

$$\begin{aligned}q(\boldsymbol{\mu}_i|m) &= |\mathbf{B}_i'|^{-\eta_i'/2}\left(2^{\frac{\eta_i' d}{2}}\pi^{\frac{d(d-1)}{4}}\prod_j^d \Gamma\left(\frac{\eta_i'+i-j}{2}\right)\right) \\ &\propto |\mathbf{B}_i'|^{-\eta_i'/2} \\ &= |(\bar{N}_i+\xi_0)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i\right)^T + \mathbf{B}_i|^{-\frac{\eta_0+\bar{N}_i+1}{2}}\end{aligned} \tag{52}$$

となる．ここで行列式の公式 $|A + aa^T| = |A|(1 + a^T A^{-1} a)$ を用いて変形すると，以下のようなティ分布となる．

$$\begin{aligned}q(\boldsymbol{\mu}_i|m) &= \left[|\mathbf{B}_i|\left\{1 + (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T(\bar{N}_i+\xi_0)\mathbf{B}_i^{-1}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)\right\}\right]^{-\frac{\eta_0+\bar{N}_i+1}{2}} \\ &\propto \mathcal{T}(\boldsymbol{\mu}_i|\bar{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i})\end{aligned} \tag{53}$$

ただし，

$$\begin{aligned}\boldsymbol{\Sigma}_{\mu_i} &= \frac{\mathbf{B}_i}{(\bar{N}_i+\xi_0)f_{\mu_i}} \\ f_{\mu_i} &= \eta_0 + \bar{N}_i + 1 - d\end{aligned} \tag{54}$$

また，ティ分布の確率密度関数は次式で定義される．

$$\mathcal{T}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},f)=\frac{\Gamma((f+d)/2)}{(f\pi)^{(d/2)}\Gamma(f/2)|\boldsymbol{\Sigma}|^{1/2}}\left\{1+(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{\Sigma}f)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}^{-\frac{d+f}{2}} \tag{55}$$

ただし，Γ はガンマ関数を表す．

2.5 $q(\mathbf{Z}|m)$ の導出

次に潜在変数の最適事後分布の導出を行う．式 (24) に式 (28) を代入すると次式を得る．

$$\begin{aligned} q(\mathbf{Z}|m) &\propto \exp\left[\int\int\int q(\mathbf{S}|m)q(\boldsymbol{\mu}|m)q(\boldsymbol{\alpha}|m)\sum_i^m\sum_n^N z_i^n\left\{\log\alpha_i+\log|\mathbf{S}_i|^{1/2}\right.\right. \\ &\qquad \left.\left.-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i\left(\boldsymbol{x}_n-\boldsymbol{\mu}_i\right)\left(\boldsymbol{x}_n-\boldsymbol{\mu}_i\right)^T\right)\right\}d\mathbf{S}d\boldsymbol{\mu}d\boldsymbol{\alpha}\right] \\ &= \prod_i^m\prod_n^N\exp\left[z_i^n\left\{\langle\log\alpha_i\rangle_{q(\boldsymbol{\alpha}|m)}+\frac{1}{2}\langle\log|\mathbf{S}_i|\rangle_{q(\mathbf{S}_i|m)}\right.\right. \\ &\qquad \left.\left.-\frac{1}{2}\mathrm{Tr}\left(\langle\mathbf{S}_i\rangle_{q(\mathbf{S}_i|m)}\left\langle\left(\boldsymbol{x}_n-\boldsymbol{\mu}_i\right)\left(\boldsymbol{x}_n-\boldsymbol{\mu}_i\right)^T\right\rangle_{q(\boldsymbol{\mu}_i|m)}\right)\right\}\right] \end{aligned} \tag{56}$$

すなわち，

$$\begin{aligned} q(z_i^n=1|m) &\propto \exp\left\{\langle\log\alpha_i\rangle_{q(\boldsymbol{\alpha}|m)}+\frac{1}{2}\langle\log|\mathbf{S}_i|\rangle_{q(\mathbf{S}_i|m)}\right. \\ &\qquad \left.-\frac{1}{2}\mathrm{Tr}\left(\langle\mathbf{S}_i\rangle_{q(\mathbf{S}_i|m)}\left\langle\left(\boldsymbol{x}_n-\boldsymbol{\mu}_i\right)\left(\boldsymbol{x}_n-\boldsymbol{\mu}_i\right)^T\right\rangle_{q(\boldsymbol{\mu}_i|m)}\right)\right\} \end{aligned} \tag{57}$$

となる．続いて，この式内の各期待値

$$\langle\log\alpha_i\rangle_{q(\boldsymbol{\alpha}|m)} \tag{58}$$

$$\langle\mathbf{S}_i\rangle_{q(\mathbf{S}_i|m)} \tag{59}$$

$$\langle\log|\mathbf{S}_i|\rangle_{q(\mathbf{S}_i|m)} \tag{60}$$

$$\left\langle\left(\boldsymbol{x}_n-\boldsymbol{\mu}_i\right)\left(\boldsymbol{x}_n-\boldsymbol{\mu}_i\right)^T\right\rangle_{q(\boldsymbol{\mu}_i|m)} \tag{61}$$

の 4 つを計算する．

2.5.1 $\langle \log \alpha_i \rangle_{q(\boldsymbol{\alpha}|m)}$ の計算

式 (37) を式 (58) に代入することで次式を得る．

$$\langle \log \alpha_i \rangle_{q(\boldsymbol{\alpha}|m)} = \frac{\Gamma(\sum_j^m \phi'_j)}{\prod_j^m \Gamma(\phi'_j)} \int \log \alpha_i \prod_j^m \alpha_j^{\phi'_j - 1} d\boldsymbol{\alpha} \tag{62}$$

ただし，$\phi'_i = \phi_0 + \bar{N}_i$ とする．ここで，次式で定義される diggama 関数

$$\Psi(\phi'_i) = \frac{\partial \log \Gamma(\phi'_i)}{\partial \phi'_i} = \frac{1}{\Gamma(\phi'_i)} \frac{\partial \Gamma(\phi'_i)}{\partial \phi'_i} \tag{63}$$

より，

$$\frac{\partial \Gamma(\phi'_i)}{\partial \phi'_i} = \Gamma(\phi'_i)\Psi(\phi'_i) \tag{64}$$

が成り立つ．また，

$$\Psi(\sum_i^m \phi'_i) = \frac{\partial \log \Gamma(\sum_i^m \phi'_i)}{\partial (\sum_i^m \phi'_i)} = \frac{1}{\Gamma(\sum_i^m \phi'_i)} \frac{\frac{\partial \Gamma(\sum_i^m \phi'_i)}{\partial \phi'_i}}{\frac{\partial (\sum_i^m \phi'_i)}{\partial \phi'_i}} = \frac{1}{\Gamma(\sum_i^m \phi'_i)} \frac{\partial \Gamma(\sum_i^m \phi'_i)}{\partial \phi'_i} \tag{65}$$

より，

$$\frac{\partial \Gamma(\sum_i^m \phi'_i)}{\partial \phi'_i} = \Gamma(\sum_i^m \phi'_i)\Psi(\sum_i^m \phi'_i) \tag{66}$$

が成り立つ．さらに，$\int \mathcal{D}(\{\alpha_i\}_{i=1}^m | \{\phi'_i\}_{i=1}^m) d\boldsymbol{\alpha} = 1$ より，以下の式が成り立つ．

$$\begin{aligned} \int \frac{\Gamma(\sum_j^m \phi'_j)}{\prod_j^m \Gamma(\phi'_j)} \prod_j^m \alpha_j^{\phi'_j - 1} d\boldsymbol{\alpha} = 1 \\ \therefore \quad \int \prod_j^m \alpha_j^{\phi'_j - 1} d\boldsymbol{\alpha} = \frac{\prod_j^m \Gamma(\phi'_j)}{\Gamma(\sum_j^m \phi'_j)} \end{aligned} \tag{67}$$

この式の両辺を ϕ'_i で微分する．

$$\begin{aligned} \int (\log \alpha_i) \prod_j^m \alpha_j^{\phi'_j - 1} d\boldsymbol{\alpha} &= \frac{\partial}{\partial \phi'_i} \frac{\prod_j^m \Gamma(\phi'_j)}{\Gamma(\sum_j^m \phi'_j)} \\ &= \frac{\prod_{j \neq i} \Gamma(\phi'_j)}{\Gamma(\sum_j^m \phi'_j)} \frac{\partial \Gamma(\phi'_i)}{\partial \phi'_i} - \frac{\prod_j^m \Gamma(\phi'_j)}{\left\{\Gamma(\sum_j^m \phi'_j)\right\}^2} \frac{\partial \Gamma(\sum_j^m \phi'_j)}{\partial \phi'_i} \end{aligned} \tag{68}$$

さらに，式 (64)，式 (66) を代入すると，

$$\begin{aligned}\int (\log\alpha_i)\prod_j^m \alpha_j^{\phi'_j-1}d\boldsymbol{\alpha} &= \frac{\prod_{j\neq i}\Gamma(\phi'_j)}{\Gamma(\sum_j^m \phi'_j)}\Gamma(\phi'_i)\Psi(\phi'_i) - \frac{\prod_j^m \Gamma(\phi'_j)}{\left\{\Gamma(\sum_j^m \phi'_j)\right\}^2}\Gamma(\sum_j^m \phi'_j)\Psi(\sum_j^m \phi'_j) \\ &= \frac{\prod_j \Gamma(\phi'_j)}{\Gamma(\sum_j^m \phi'_j)}\left\{\Psi(\phi'_i) - \Psi(\sum_j^m \phi'_j)\right\} \qquad (69)\end{aligned}$$

これを，式 (62) に代入すると，以下の式を得る．

$$\langle\log\alpha_i\rangle_{q(\boldsymbol{\alpha}|m)} = \Psi(\phi'_i) - \Psi(\sum_j^m \phi'_j) = \Psi(\phi_0+\bar{N}_i) - \Psi(\sum_j^m(\phi_0+\bar{N}_j)) \qquad (70)$$

2.5.2 $\langle\mathbf{S}_i\rangle_{q(\mathbf{S}_i|m)}$ の計算

式 (59) に，式 (46) を代入すると，

$$\langle\mathbf{S}_i\rangle_{q(\mathbf{S}_i|m)} = \int \mathbf{S}_i\mathcal{W}(\mathbf{S}_i|\eta',\mathbf{B}_i)d\mathbf{S}_i \qquad (71)$$

ただし，

$$\begin{aligned}\mathcal{W}(\mathbf{S}_i|\eta'_i,\mathbf{B}_i) &= C_{\mathcal{W}}(\eta'_i,\mathbf{B}_i)|\mathbf{S}_i|^{1/2(\eta'_i-d-1)}\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i\mathbf{B}_i\right)\right) &(72)\\ C_{\mathcal{W}}(\eta'_i,\mathbf{B}_i) &= |\mathbf{B}_i|^{\eta'_i/2}\left\{2^{\eta'_i d/2}\pi^{d(d-1)/4}\prod_j^d\Gamma(\frac{\eta'_i+1-j}{2})\right\}^{-1} &(73)\\ \eta'_i &= \eta_0+\bar{N}_i &(74)\end{aligned}$$

とする．また，$\int\mathcal{W}(\mathbf{S}_i|\eta'_i,\mathbf{B}_i)d\mathbf{S}_i = 1$ より次式が成り立つ．

$$\int |\mathbf{S}_i|^{1/2(\eta'_i-d-1)}\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i\mathbf{B}_i\right)\right)d\mathbf{S}_i = \frac{1}{C_{\mathcal{W}}(\eta'_i,\mathbf{B}_i)} \qquad (75)$$

この両辺を，トレースの微分の公式 $\frac{\partial\mathrm{Tr}(\boldsymbol{AB})}{\partial\boldsymbol{A}} = \frac{\partial\mathrm{Tr}(\boldsymbol{BA})}{\partial\boldsymbol{A}} = \boldsymbol{B}^T$ を用いて，$\mathbf{S}_i$ が対象行列であることに注意し，$\mathbf{B}_i$ で微分する．

$$-\frac{1}{2}\int \mathbf{S}_i|\mathbf{S}_i|^{1/2(\eta'_i-d-1)}\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}_i\mathbf{B}_i\right)\right)d\mathbf{S}_i = \frac{\partial}{\partial\mathbf{B}_i}C_{\mathcal{W}}(\eta'_i,\mathbf{B}_i)^{-1} \qquad (76)$$

さらに，行列式の微分公式 $\frac{\partial\log|\boldsymbol{A}|}{\partial\boldsymbol{A}} = (\boldsymbol{A}^{-1})^T$ を用いて，$\log C_{\mathcal{W}}(\eta'_i,\mathbf{B}_i)^{-1}$ を $\mathbf{B}_i$ で微分し，$\mathbf{B}_i$ が対称行列であることに注意し，変形する．

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{B}_i} \log C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)^{-1} &= \frac{\partial}{\partial \mathbf{B}_i} \log |\mathbf{B}_i|^{-\frac{\eta_i'}{2}} \\
C_{\mathcal{W}}(\eta_i', \mathbf{B}_i) \frac{\partial}{\partial \mathbf{B}_i} C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)^{-1} &= -\frac{\eta_i'}{2} \mathbf{B}_i^{-1} \\
\therefore \quad \frac{\partial}{\partial \mathbf{B}_i} C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)^{-1} &= -\frac{\eta_i'}{2C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)} \mathbf{B}_i^{-1} \qquad (77)
\end{aligned}
$$

この式を式 (76) に代入すると次式を得る．

$$
\int \mathbf{S}_i |\mathbf{S}_i|^{1/2(\eta_i'-d-1)} \exp\left(-\frac{1}{2} \mathrm{Tr}\,(\mathbf{S}_i \mathbf{B}_i)\right) d\mathbf{S}_i \;=\; \frac{\eta_i'}{C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)} \mathbf{B}_i^{-1} \qquad (78)
$$

この式を，式 (71) へ代入すると，$\langle \mathbf{S}_i \rangle_{q(\mathbf{S}_i|m)}$ は以下のようになる．

$$
\langle \mathbf{S}_i \rangle_{q(\mathbf{S}_i|m)} = \eta_i' \mathbf{B}_i^{-1} = (\eta_0 + \bar{N}_i) \mathbf{B}_i^{-1} \qquad (79)
$$

2.5.3 $\langle \log |\mathbf{S}_i| \rangle_{q(\mathbf{S}_i|m)}$ の計算

式 (60) に，式 (46) を代入すると，次式を得る．

$$
\langle \log |\mathbf{S}_i| \rangle_{q(\mathbf{S}_i|m)} = \int \log |\mathbf{S}_i| \mathcal{W}(\mathbf{S}_i | \eta_i', \mathbf{B}_i) d\mathbf{S}_i \qquad (80)
$$

式 (75) を，η_i' で微分すると，

$$
\frac{1}{2} \int (\log |\mathbf{S}_i|) |\mathbf{S}_i|^{1/2(\eta_i'-d-1)} \exp\left(-\frac{1}{2} \mathrm{Tr}\,(\mathbf{S}_i \mathbf{B}_i)\right) d\mathbf{S}_i = \frac{\partial}{\partial \eta_i'} C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)^{-1} \qquad (81)
$$

となる．また，$\Psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$ に注意し，$\log C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)^{-1}$ を η_i' で微分する．

$$
\begin{aligned}
\frac{\partial}{\partial \eta_i'} \log C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)^{-1} &= \frac{\partial}{\partial \eta_i'} \log \left(|\mathbf{B}_i|^{-\eta_i'/2} 2^{\eta_i' d/2} \pi^{d(d-1)/4} \prod_j^d \Gamma(\frac{\eta_i' + 1 - j}{2}) \right) \\
C_{\mathcal{W}}(\eta_i', \mathbf{B}_i) \frac{\partial}{\partial \eta_i'} C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)^{-1} &= -\frac{1}{2} \log |\mathbf{B}_i| + \frac{d}{2} \log 2 + \frac{1}{2} \sum_j^d \Psi(\frac{\eta_i' + 1 + j}{2}) \\
\therefore \quad \frac{\partial}{\partial \eta_i'} C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)^{-1} &= C_{\mathcal{W}}(\eta_i', \mathbf{B}_i)^{-1} \left\{ -\frac{1}{2} \log |\mathbf{B}_i| + \frac{d}{2} \log 2 + \frac{1}{2} \sum_j^d \Psi(\frac{\eta_i' + 1 + j}{2}) \right\}
\end{aligned} \qquad (82)
$$

式 (82) と式 (81) より，式 (80) は次式となる．

$$\begin{aligned}\left\langle \log|\mathbf{S}_i|^{1/2}\right\rangle_{q(\mathbf{S}_i|m)} &= -\log|\mathbf{B}_i| + d\log 2 + \sum_j^d \Psi(\frac{\eta_i' + 1 + j}{2}) \\ &= -\log|\mathbf{B}_i| + d\log 2 + \sum_j^d \Psi(\frac{\eta_0 + \bar{N}_i + 1 + j}{2}) \qquad (83)\end{aligned}$$

2.5.4 $\left\langle (\boldsymbol{x}_n - \boldsymbol{\mu}_i)(\boldsymbol{x}_n - \boldsymbol{\mu}_i)^T \right\rangle_{q(\boldsymbol{\mu}_i|m)}$ の計算

まず，以下のようなティ分布の積分を考える．

$$\int \mathcal{T}\left(\boldsymbol{\mu}_i|\bar{\boldsymbol{\mu}}_i, \frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2\right) d\boldsymbol{\mu}_i = 1 \qquad (84)$$

$$\int C_{\mathcal{T}}(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2)\left\{1 + (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T (f_{\mu_i}\boldsymbol{\Sigma}_{\mu_i})^{-1}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)\right\}^{-\frac{f_{\mu_i}+d-2}{2}} d\boldsymbol{\mu}_i = 1 \qquad (85)$$

ただし，$C_{\mathcal{T}}(\boldsymbol{\Sigma}, f)$ は以下のような正規化項である．

$$C_{\mathcal{T}}(\boldsymbol{\Sigma}, f) = \frac{\Gamma((f+d)/2)}{(f\pi)^{d/2}\Gamma(f/2)|\Sigma|^{1/2}} \qquad (86)$$

式 (85) を変形し，次式を得る．

$$\int \left\{1 + (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T (f_{\mu_i}\boldsymbol{\Sigma}_{\mu_i})^{-1}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)\right\}^{-\frac{f_{\mu_i}+d-2}{2}} d\boldsymbol{\mu}_i = C_{\mathcal{T}}(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2)^{-1} \qquad (87)$$

この左辺を，行列の微分公式 $\frac{\partial \boldsymbol{a}^T\boldsymbol{A}\boldsymbol{b}}{\partial \boldsymbol{A}} = \boldsymbol{a}\boldsymbol{b}^T$ を用いて，$\boldsymbol{\Sigma}_{\mu_i}^{-1}$ で微分し変形する．

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\Sigma}_{\mu_i}^{-1}} \int \left\{1 + (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T (f_{\mu_i} \boldsymbol{\Sigma}_{\mu_i})^{-1} (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)\right\}^{-\frac{f_{\mu_i}+d-2}{2}} d\boldsymbol{\mu}_i \\
&= -\frac{f_{\mu_i}+d-2}{2} \int \left\{1 + (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T (f_{\mu_i} \boldsymbol{\Sigma}_{\mu_i})^{-1} (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)\right\}^{-\frac{f_{\mu_i}+d}{2}} \\
&\qquad \times \frac{1}{f_{\mu_i}} (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T d\boldsymbol{\mu}_i \\
&= -\frac{f_{\mu_i}+d-2}{2} \int \left\{1 + (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T (f_{\mu_i} \boldsymbol{\Sigma}_{\mu_i})^{-1} (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)\right\}^{-\frac{f_{\mu_i}+d}{2}} \\
&\qquad \times \frac{1}{f_{\mu_i}} \Big\{(\boldsymbol{x}_n - \boldsymbol{\mu}_i)(\boldsymbol{x}_n - \boldsymbol{\mu}_i)^T + (\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)(\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)^T \\
&\qquad - 2(\boldsymbol{x}_n - \boldsymbol{\mu}_i)(\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)^T\Big\} d\boldsymbol{\mu}_i \\
&= -\frac{f_{\mu_i}+d-2}{2f_{\mu_i}} C_{\mathcal{T}}(\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i})^{-1} \left[\int (\boldsymbol{x}_n - \boldsymbol{\mu}_i)(\boldsymbol{x}_n - \boldsymbol{\mu}_i)^T \mathcal{T}(\boldsymbol{\mu}_i | \bar{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}) d\boldsymbol{\mu}_i \right. \\
&\qquad + (\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)(\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)^T \int \mathcal{T}(\boldsymbol{\mu}_i | \bar{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}) d\boldsymbol{\mu}_i \\
&\qquad - 2\boldsymbol{x}_n (\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)^T \int \mathcal{T}(\boldsymbol{\mu}_i | \bar{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}) d\boldsymbol{\mu}_i \\
&\qquad \left. + 2 \int \boldsymbol{\mu}_i \mathcal{T}(\boldsymbol{\mu}_i | \bar{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}) d\boldsymbol{\mu}_i (\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)\right] \\
&= -\frac{f_{\mu_i}+d-2}{2f_{\mu_i}} C_{\mathcal{T}}(\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i})^{-1} \left[\left\langle (\boldsymbol{x}_n - \boldsymbol{\mu}_i)(\boldsymbol{x}_n - \boldsymbol{\mu}_i)^T \right\rangle_{q(\boldsymbol{\mu}_i|m)} \right. \\
&\qquad \left. + (\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)(\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)^T - 2\boldsymbol{x}_n (\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)^T + 2\bar{\boldsymbol{\mu}}_i (\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)^T\right] \\
&= -\frac{f_{\mu_i}+d-2}{2f_{\mu_i}} C_{\mathcal{T}}(\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i})^{-1} \left[\left\langle (\boldsymbol{x}_n - \boldsymbol{\mu}_i)(\boldsymbol{x}_n - \boldsymbol{\mu}_i)^T \right\rangle_{q(\boldsymbol{\mu}_i|m)} \right. \\
&\qquad \left. - (\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)(\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i)^T\right]
\end{aligned} \tag{88}
$$

次に，式 (87) の右辺の対数をとり，$\boldsymbol{\Sigma}_{\mu_i}^{-1}$ で微分する．

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\Sigma}_{\mu_i}^{-1}} \log C_{\mathcal{T}}\left(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2\right)^{-1} = \frac{\partial}{\partial \boldsymbol{\Sigma}_{\mu_i}^{-1}} - \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mu_i}^{-1}| \\
&C_{\mathcal{T}}\left(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2\right) \frac{\partial}{\partial \boldsymbol{\Sigma}_{\mu_i}^{-1}} C_{\mathcal{T}}\left(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2\right)^{-1} = -\frac{1}{2}\boldsymbol{\Sigma}_{\mu_i} \\
&\therefore \frac{\partial}{\partial \boldsymbol{\Sigma}_{\mu_i}^{-1}} C_{\mathcal{T}}\left(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2\right)^{-1} = -\frac{1}{2} C_{\mathcal{T}}\left(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2\right)^{-1} \boldsymbol{\Sigma}_{\mu_i}
\end{aligned} \tag{89}
$$

また，行列式の性質 $|a\boldsymbol{A}| = a^d|\boldsymbol{A}|$ とガンマ関数の性質 $\Gamma(x+1) = x\Gamma(x)$ に注意し，式変形を行うと，次式が成り立つ．

$$
\begin{aligned}
\frac{C_{\mathcal{T}}(\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i})}{C_{\mathcal{T}}(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2)} &= \frac{\Gamma((f_{\mu_i}+d)/2)}{(f_{\mu_i}\pi)^{d/2}\Gamma(f_{\mu_i}/2)|\boldsymbol{\Sigma}_{\mu_i}|^{1/2}}\left\{\frac{\Gamma((f_{\mu_i}+d-2)/2)}{((f_{\mu_i}-2)\pi)^{d/2}\,\Gamma((f_{\mu_i}-2)/2)|\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}|^{1/2}}\right\} \\
&= \frac{\Gamma(\frac{f_{\mu_i}+d-2}{2}+1)}{\Gamma(\frac{f_{\mu_i}+d-2}{2})}\frac{\Gamma(\frac{f_{\mu_i}-2}{2})}{\Gamma(\frac{f_{\mu_i}-2}{2}+1)}\left(\frac{f_{\mu_i}-2}{f_{\mu_i}}\right)^{d/2}\left(\frac{f_{\mu_i}}{f_{\mu_i}-2}\right)^{d/2} \\
&= \frac{f_{\mu_i}+d-2}{f_{\mu_i}-2} \qquad (9
\end{aligned}
$$

式 (88)=式 (89) より，

$$
\begin{aligned}
&-\frac{f_{\mu_i}+d-2}{2f_{\mu_i}}C_{\mathcal{T}}(\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i})^{-1}\left[\left\langle(\boldsymbol{x}_n-\boldsymbol{\mu}_i)(\boldsymbol{x}_n-\boldsymbol{\mu}_i)^T\right\rangle_{q(\boldsymbol{\mu}_i|m)}-(\boldsymbol{x}_n-\bar{\boldsymbol{\mu}}_i)(\boldsymbol{x}_n-\bar{\boldsymbol{\mu}}_i)^T\right] \\
&\qquad = -\frac{1}{2}C_{\mathcal{T}}(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2)^{-1}\boldsymbol{\Sigma}_{\mu_i} \\
&\left\langle(\boldsymbol{x}_n-\boldsymbol{\mu}_i)(\boldsymbol{x}_n-\boldsymbol{\mu}_i)^T\right\rangle_{q(\boldsymbol{\mu}_i|m)} \\
&\qquad = \frac{f_{\mu_i}}{f_{\mu_i}+d-2}\frac{C_{\mathcal{T}}(\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i})}{C_{\mathcal{T}}(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}-2)}\boldsymbol{\Sigma}_{\mu_i}+(\boldsymbol{x}_n-\bar{\boldsymbol{\mu}}_i)(\boldsymbol{x}_n-\bar{\boldsymbol{\mu}}_i)^T \quad (91)
\end{aligned}
$$

最終的に，式 (90) を代入することで，求めるべき期待値は以下のようになる．

$$
\left\langle(\boldsymbol{x}_n-\boldsymbol{\mu}_i)(\boldsymbol{x}_n-\boldsymbol{\mu}_i)^T\right\rangle_{q(\boldsymbol{\mu}_i|m)} = \frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}+(\boldsymbol{x}_n-\bar{\boldsymbol{\mu}}_i)(\boldsymbol{x}_n-\bar{\boldsymbol{\mu}}_i)^T \quad (92)
$$

2.5.5 $q(z_i^n=1|m)$ の計算

ここまでで計算した，式 (70)，式 (79)，式 (83)，式 (92) を式 (57) へ代入することで次式を得る．

$$
\begin{aligned}
q(z_i^n=1|m) &= \frac{\exp(\gamma_i^n)}{\sum_j^m \exp(\gamma_j^n)} \\
\gamma_i^n &= \Psi(\phi_0+\bar{N}_i)-\Psi(\sum_j^m \phi_0+\bar{N}_i) \\
&\quad -\frac{1}{2}\log|\mathbf{B}_i|+\frac{1}{2}\sum_j^m \Psi(\frac{\eta_0+\bar{N}_i+1+j}{2}) \\
&\quad -\frac{1}{2}\mathrm{Tr}\left\{(\eta_0+\bar{N}_i)\mathbf{B}_i^{-1}\right. \\
&\quad \left.\times(\frac{f_{\mu_i}}{f_{\mu_i}-2}\boldsymbol{\Sigma}_{\mu_i}+(\boldsymbol{x}_n-\bar{\boldsymbol{\mu}}_i)(\boldsymbol{x}_n-\bar{\boldsymbol{\mu}}_i)^T)\right\} \qquad (93)
\end{aligned}
$$

2.6 アルゴリズム

以上の導出結果に基づいて，アルゴリズム形式で以下にまとめる．

[VB algorithm]――――――――――――――――――――

Step1:初期化

事前分布のパラメータを初期値 $(\phi_0, \xi_0, \eta_0, \boldsymbol{\nu}_0, \mathbf{B}_0)$ とし，$\bar{N}_i^{(0)} = N/m$ として，変分事後分布のパラメータを初期化する．

$$\bar{\boldsymbol{\mu}}_i^{(0)} = \boldsymbol{\nu}_0, \quad \mathbf{B}_i^{(0)} = \mathbf{B}_0, \tag{94}$$

$$f_{\mu_i}^{(0)} = \eta_0 + \bar{N}_i + 1 - d, \quad \boldsymbol{\Sigma}_{\mu_i}^{(0)} = \frac{\mathbf{B}_i^{(0)}}{(\bar{N}_i^{(0)} + \xi_0) f_{\mu_i}^{(0)}} \tag{95}$$

Step2:収束するまで以下を繰り返す

2-1:潜在変数の事後分布の更新

$i = 0, \ldots, m, \quad n = 1, \ldots, N$ に対し，

$$\bar{\boldsymbol{z}}_i^n = \frac{\exp\left(\gamma_i^n\right)}{\sum_j^m \exp\left(\gamma_j^n\right)} \tag{96}$$

を計算する．ただし，γ_i^n は以下のようになる．

$$\begin{aligned}
\gamma_i^n \quad = \quad & \Psi(\phi_0 + \bar{N}_i^{(t)}) - \Psi(\sum_j^m (\phi_0 + \bar{N}_j^{(t)})) \\
& -\frac{1}{2}\log|\mathbf{B}_i^{(t)}| + \frac{1}{2}\sum_j^d \Psi(\frac{\eta_0 + \bar{N}_i^{(t)} + 1 + j}{2}) \\
& -\frac{1}{2}\mathrm{Tr}\left\{(\eta_0 + \bar{N}_i^{(t)})(\mathbf{B}_i^{(t)})^{-1}\right. \\
& \left.\times(\frac{f_{\mu_i}^{(t)}}{f_{\mu_i}^{(t)} - 2}\boldsymbol{\Sigma}_{\mu_i}^{(t)} + \left(\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i^{(t)}\right)\left(\boldsymbol{x}_n - \bar{\boldsymbol{\mu}}_i^{(t)}\right)^T)\right\}
\end{aligned} \tag{97}$$

2-2:変分事後分布の更新

$i = 0, \ldots, m, \quad n = 1, \ldots, N$ に対し，

$$\bar{N}_i^{(t)} \leftarrow \sum_{n=1}^{N} \bar{z}_i^n, \quad \bar{\boldsymbol{x}}_i^{(t)} \leftarrow \frac{1}{\bar{N}_i^{(t)}} \sum_{n=1}^{N} \bar{z}_i^n \boldsymbol{x}_n \tag{98}$$

$$\bar{\mathbf{C}}_i^{(t)} = \sum_{n=1}^{N} \bar{z}_i^n \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i^{(t)}\right) \left(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i^{(t)}\right)^T \tag{99}$$

$$\boldsymbol{\mu}_i^{(t)} \leftarrow \frac{\bar{N}_i^{(t)} \bar{\boldsymbol{x}}_i^{(t)} + \xi_0 \boldsymbol{\nu}_0}{\bar{N}_i^{(t)} + \xi_0}, \quad f_{\mu_i}^{(t)} \leftarrow \eta_i^{(t)} + 1 - d \tag{100}$$

$$\mathbf{B}_i^{(t)} \leftarrow \mathbf{B}_0 + \bar{\mathbf{C}}_i^{(t)} + \frac{\bar{N}_i^{(t)} \xi_0}{\bar{N}_i^{(t)} + \xi_0} \left(\bar{\boldsymbol{x}}_i^{(t)} - \boldsymbol{\nu}_0\right) \left(\bar{\boldsymbol{x}}_i^{(t)} - \boldsymbol{\nu}_0\right)^T \tag{101}$$

$$\boldsymbol{\Sigma}_{\mu_i}^{(t)} \leftarrow \frac{\mathbf{B}_i^{(t)}}{(\bar{N}_i^{(t)} + \xi_0) f_{\mu_i}^{(t)}} \tag{102}$$

を計算し，$t \leftarrow t + 1$ とする．
