【付録】5 章における正規分布、GMM、HMM パラメータ推定式の導出

本付録資料では、ページの都合上で本に載せきれなかった、多変量正規分布のパラメータ（平均値ベクトルおよび分散共分散行列）、GMM のパラメータ（混合重み、平均値ベクトル、分散共分散行列）、そして HMM の状態遷移確率を求める式を導出しています。

【補足 1】5.2 節 最尤推定法による平均値ベクトルと分散共分散行列の導出

ここでは、式 (5.11) から式 (5.13) の導出と、式 (5.12) から式 (5.14) の導出を行います。まず、平均値ベクトルについて導出します。ここでは、行列に関する以下の性質を用います。

$$(\boldsymbol{ABC})^\top = \boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A}^\top \tag{5.90}$$

$$\frac{\partial}{\partial \boldsymbol{a}} \boldsymbol{a}^\top \boldsymbol{A}\boldsymbol{a} = (\boldsymbol{A} + \boldsymbol{A}^\top)\boldsymbol{a} \tag{5.91}$$

$$\frac{\partial}{\partial \boldsymbol{a}} \boldsymbol{b}^\top \boldsymbol{a} = \boldsymbol{b} \tag{5.92}$$

$\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$ はそれぞれ行列、$\boldsymbol{a}$、$\boldsymbol{b}$ はそれぞれベクトルです。これらの性質についての導出は割愛しますが、適当な行列とベクトルを作って実際に計算すれば、正しいことが確認できると思います。式 (5.11) を展開していきます。

$$\frac{\partial}{\partial \boldsymbol{\mu}} \sum_{n=0}^{N-1} (\boldsymbol{x}(n) - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}(n) - \boldsymbol{\mu})$$

$$= \frac{\partial}{\partial \boldsymbol{\mu}} \sum_{n=0}^{N-1} \left\{ (\boldsymbol{x}(n))^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}(n) - (\boldsymbol{x}(n))^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}(n) + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right\} \tag{5.93}$$

$$= \frac{\partial}{\partial \boldsymbol{\mu}} \sum_{n=0}^{N-1} \left\{ - (\boldsymbol{x}(n))^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \left(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}(n)\right)^\top + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right\} \tag{5.94}$$

$$= \frac{\partial}{\partial \boldsymbol{\mu}} \sum_{n=0}^{N-1} \left\{ -2 (\boldsymbol{x}(n))^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right\} \qquad (\because \text{式 } (5.90)) \tag{5.95}$$

$$= \frac{\partial}{\partial \boldsymbol{\mu}} \left\{ -2 \left(\sum_{n=0}^{N-1} \boldsymbol{x}(n) \right)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + N \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right\} \tag{5.96}$$

$$= -2 \left(\left(\sum_{n=0}^{N-1} \boldsymbol{x}(n) \right)^\top \boldsymbol{\Sigma}^{-1} \right)^\top + N \left(\boldsymbol{\Sigma}^{-1} + (\boldsymbol{\Sigma}^{-1})^\top \right) \boldsymbol{\mu} \qquad (\because \text{式 } (5.91), (5.92)) \tag{5.97}$$

$$= -2 \boldsymbol{\Sigma}^{-1} \sum_{n=0}^{N-1} \boldsymbol{x}(n) + 2N \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \qquad \left(\because (\boldsymbol{\Sigma}^{-1})^\top = \boldsymbol{\Sigma}^{-1}\right) \tag{5.98}$$

$$= 0 \tag{5.99}$$

これを解くと、

$$
\begin{aligned}
N\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} &= \boldsymbol{\Sigma}^{-1}\sum_{n=0}^{N-1}\boldsymbol{x}(n) && (5.100)\\
\hat{\boldsymbol{\mu}} &= \frac{1}{N}\sum_{n=0}^{N-1}\boldsymbol{x}(n) && (5.101)
\end{aligned}
$$

と導出されます。

つぎに、分散共分散行列について導出します。ここでは、行列に関する以下の性質を用います。

$$
\begin{aligned}
\frac{\partial|\boldsymbol{A}|}{\partial\boldsymbol{A}} &= |\boldsymbol{A}|(\boldsymbol{A}^{-1})^{\mathsf{T}} && (5.102)\\
\frac{\partial\boldsymbol{a}^{\mathsf{T}}\boldsymbol{A}^{-1}\boldsymbol{b}}{\partial\boldsymbol{A}} &= -\left(\boldsymbol{A}^{-1}\right)^{\mathsf{T}}\boldsymbol{a}\boldsymbol{b}^{\mathsf{T}}\left(\boldsymbol{A}^{-1}\right)^{\mathsf{T}} && (5.103)
\end{aligned}
$$

式 (5.12) を展開していきます。

$$
\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{\Sigma}}\left\{N\log|\boldsymbol{\Sigma}| + \sum_{n=0}^{N-1}(\boldsymbol{x}(n)-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}(n)-\boldsymbol{\mu})\right\} && (5.104)\\
&= \frac{N}{|\boldsymbol{\Sigma}|}\frac{\partial|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\Sigma}} + \sum_{n=0}^{N-1}\frac{\partial}{\partial\boldsymbol{\Sigma}}(\boldsymbol{x}(n)-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}(n)-\boldsymbol{\mu}) && (5.105)\\
&= \frac{N}{|\boldsymbol{\Sigma}|}|\boldsymbol{\Sigma}|\left(\boldsymbol{\Sigma}^{-1}\right)^{\mathsf{T}} + \sum_{n=0}^{N-1}\frac{\partial}{\partial\boldsymbol{\Sigma}}(\boldsymbol{x}(n)-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}(n)-\boldsymbol{\mu}) \quad (\because \text{式 } (5.102)) && (5.106)\\
&= N\boldsymbol{\Sigma}^{-1} - \sum_{n=0}^{N-1}\left(\boldsymbol{\Sigma}^{-1}\right)^{\mathsf{T}}(\boldsymbol{x}(n)-\boldsymbol{\mu})(\boldsymbol{x}(n)-\boldsymbol{\mu})^{\mathsf{T}}\left(\boldsymbol{\Sigma}^{-1}\right)^{\mathsf{T}} \quad (\because \text{式 } (5.103)) && (5.107)\\
&= N\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\left(\sum_{n=0}^{N-1}(\boldsymbol{x}(n)-\boldsymbol{\mu})(\boldsymbol{x}(n)-\boldsymbol{\mu})^{\mathsf{T}}\right)\boldsymbol{\Sigma}^{-1} && (5.108)\\
&= 0 && (5.109)
\end{aligned}
$$

これを解くと、

$$
\begin{aligned}
N\boldsymbol{\Sigma}^{-1} &= \boldsymbol{\Sigma}^{-1}\left(\sum_{n=0}^{N-1}(\boldsymbol{x}(n)-\boldsymbol{\mu})(\boldsymbol{x}(n)-\boldsymbol{\mu})^{\mathsf{T}}\right)\boldsymbol{\Sigma}^{-1} && (5.110)\\
N &= \boldsymbol{\Sigma}^{-1}\left(\sum_{n=0}^{N-1}(\boldsymbol{x}(n)-\boldsymbol{\mu})(\boldsymbol{x}(n)-\boldsymbol{\mu})^{\mathsf{T}}\right) && (5.111)\\
N\boldsymbol{\Sigma} &= \sum_{n=0}^{N-1}(\boldsymbol{x}(n)-\boldsymbol{\mu})(\boldsymbol{x}(n)-\boldsymbol{\mu})^{\mathsf{T}} && (5.112)\\
\boldsymbol{\Sigma} &= \frac{1}{N}\sum_{n=0}^{N-1}(\boldsymbol{x}(n)-\boldsymbol{\mu})(\boldsymbol{x}(n)-\boldsymbol{\mu})^{\mathsf{T}} && (5.113)
\end{aligned}
$$

と導出されます。

【補足 2】**5.3 節 GMM における各パラメータ更新式の導出**

ここでは、式 (5.30) から式 (5.31)、(5.32)、(5.33) の導出を行います。まず w_m についての導出ですが、w_m は総和が 1 になるように解く必要があります。従って、$\sum_{m=0}^{M-1} w_m = 1$ という条件付きの最大化問題となり、これはラグランジュの未定乗数法を使って解くことになります。式 (5.30) から w_m に関する項のみに注目し、ラグランジュ関数を作成すると以下のようになります。

$$L = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n),\theta')\log w_m - \lambda\left(\sum_{m=0}^{M-1} w_m - 1\right) \tag{5.114}$$

λ はラグランジュ乗数です。これを w_m で偏微分し、イコール 0 を解きます。

$$\frac{\partial L}{\partial w_m} = \frac{1}{w_m}\sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n),\theta') - \lambda \tag{5.115}$$

$$= 0 \tag{5.116}$$

これを変形し、以下の式が得られます。

$$w_m = \frac{1}{\lambda}\sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n),\theta') \tag{5.117}$$

ここで、$\sum_{m=0}^{M-1} w_m = 1$ より、上記の式の両辺を m で総和すると、以下のように変形できます。

$$\sum_{m=0}^{M-1} w_m = \frac{1}{\lambda}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} P(z_m|\boldsymbol{x}(n),\theta') \tag{5.118}$$

$$= \frac{N}{\lambda} \tag{5.119}$$

$$= 1 \tag{5.120}$$

よって、$\lambda = N$ が得られ、これを式 (5.117) に代入すると、式 (5.31) が導出されます。

$$w_m = \frac{1}{N}\sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n),\theta') \tag{5.121}$$

$\boldsymbol{\mu}_m$ と $\boldsymbol{\Sigma}_m$ については、基本的には前節の補足部分にて説明した、正規分布の各パラメータの導出と同じですので、詳しくはそちらをご覧ください。n についての総和部分において、$P(z_m|\boldsymbol{x}(n),\theta')$

が掛けられている点のみが異なります。$\boldsymbol{\mu}_m$ について解くと以下のようになります。

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\mu}_m} \sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta')(\boldsymbol{x}(n) - \boldsymbol{\mu}_m)^\mathsf{T} \boldsymbol{\Sigma}_m^{-1} (\boldsymbol{x}(n) - \boldsymbol{\mu}_m) \\
&\quad = -2\boldsymbol{\Sigma}_m^{-1} \sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta')\boldsymbol{x}(n) + 2\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m \sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta') \qquad (5.122) \\
&\quad = 0 \qquad (5.123)
\end{aligned}
$$

これを解くと、

$$
\hat{\boldsymbol{\mu}}_m = \frac{\sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta')\boldsymbol{x}(n)}{\sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta')} \qquad (5.124)
$$

が得られます。

$\boldsymbol{\Sigma}_m$ について解くと以下のようになります。

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\Sigma}_m} \left\{ \sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta') \log |\boldsymbol{\Sigma}_m| + \sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta')(\boldsymbol{x}(n) - \boldsymbol{\mu}_m)^\mathsf{T} \boldsymbol{\Sigma}_m^{-1} (\boldsymbol{x}(n) - \boldsymbol{\mu}_m) \right\} \qquad (5.125) \\
&\quad = \sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta') \boldsymbol{\Sigma}_m^{-1} \\
&\qquad - \boldsymbol{\Sigma}_m^{-1} \left(\sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta')(\boldsymbol{x}(n) - \boldsymbol{\mu}_m)(\boldsymbol{x}(n) - \boldsymbol{\mu}_m)^\mathsf{T} \right) \boldsymbol{\Sigma}_m^{-1} \qquad (5.126) \\
&\quad = 0 \qquad (5.127)
\end{aligned}
$$

これを解くと、

$$
\hat{\boldsymbol{\Sigma}}_m = \frac{\sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta')(\boldsymbol{x}(n) - \hat{\boldsymbol{\mu}}_m)(\boldsymbol{x}(n) - \hat{\boldsymbol{\mu}}_m)^\mathsf{T}}{\sum_{n=0}^{N-1} P(z_m|\boldsymbol{x}(n), \theta')} \qquad (5.128)
$$

が得られます。

【補足】5.4 節 HMM における状態遷移確率の更新式の導出

ここでは、式 (5.47) から式 (5.48) の導出を行います。

式 (5.47) において、状態遷移確率に関する項は第一項のみです。この項では、あらゆる遷移パスで総和を計算していますが、このうち、解きたい遷移確率、すなわち $a_{s^u(n-1)=(p,i),s^u(n)=(p,j)} = a^p_{i,j}$ の最適化に関する部分のみを取り出すと、以下のようになります。

$$Q(a^p_{i,j}|\theta') = \sum_{u=0}^{U-1}\sum_{n=1}^{N^u-1}\sum_{(\boldsymbol{s}^u,\boldsymbol{z}^u)\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n-1)=(p,i), s^u(n)=(p,j)|\boldsymbol{x}^u,\theta'\right)\log a^p_{i,j} \quad (5.129)$$

この $Q(a^p_{i,j}|\theta')$ を最大化するわけですが、ここで、状態遷移確率 $a^p_{i,j}$ は、全ての遷移先 j（left-to-right HMM の場合は自己ループか次の状態）への遷移確率で総和すると 1 になるように解く必要があります。従って、$\sum_j a^p_{i,j} = 1$ という条件付きの最大化問題となります。これは GMM における混合重みを求める時と同様、ラグランジュの未定乗数法を使って解くことになります。ラグランジュ関数は以下のようになります。

$$\begin{aligned} L &= \sum_{u=0}^{U-1}\sum_{n=1}^{N^u-1}\sum_{(\boldsymbol{s}^u,\boldsymbol{z}^u)\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n-1)=(p,i), s^u(n)=(p,j)|\boldsymbol{x}^u,\theta'\right)\log a^p_{i,j} \\ &- \lambda\left(\sum_j a^p_{i,j} - 1\right) \end{aligned} \quad (5.130)$$

λ はラグランジュ乗数です。これを $a^p_{i,j}$ で偏微分してイコール 0 を解きます。

$$\begin{aligned} \frac{\partial L}{\partial a^p_{i,j}} &= \frac{1}{a^p_{i,j}}\sum_{u=0}^{U-1}\sum_{n=1}^{N^u-1}\sum_{(\boldsymbol{s}^u,\boldsymbol{z}^u)\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n-1)=(p,i), s^u(n)=(p,j)|\boldsymbol{x}^u,\theta'\right) - \lambda \\ &= 0 \end{aligned} \quad (5.131)$$

これを変形し、以下の式が得られます。

$$a^p_{i,j} = \frac{1}{\lambda}\sum_{u=0}^{U-1}\sum_{n=1}^{N^u-1}\sum_{(\boldsymbol{s}^u,\boldsymbol{z}^u)\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n-1)=(p,i), s^u(n)=(p,j)|\boldsymbol{x}^u,\theta'\right) \quad (5.132)$$

ここで、$\sum_j a_{i,j}^p = 1$ より、上式の両辺を j で総和すると、以下のように変形できます。

$$\sum_j a_{i,j}^p = \frac{1}{\lambda}\sum_{u=0}^{U-1}\sum_{n=1}^{N^u-1}\sum_j \sum_{(\boldsymbol{s}^u,\boldsymbol{z}^u)\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n-1)=(p,i), s^u(n)=(p,j)|\boldsymbol{x}^u,\theta'\right) \tag{5.133}$$

$$= \frac{1}{\lambda}\sum_{u=0}^{U-1}\sum_{n=1}^{N^u-1}\sum_{(\boldsymbol{s}^u,\boldsymbol{z}^u)\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n-1)=(p,i)|\boldsymbol{x}^u,\theta'\right) \tag{5.134}$$

$$= 1 \tag{5.135}$$

よって、

$$\lambda = \sum_{u=0}^{U-1}\sum_{n=1}^{N^u-1}\sum_{(\boldsymbol{s}^u,\boldsymbol{z}^u)\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n-1)=(p,i)|\boldsymbol{x}^u,\theta'\right) \tag{5.136}$$

が得られます。これを式 (5.132) に代入すると、以下の式が導出されます。

$$\hat{a}_{i,j}^p = \frac{\sum_{u=0}^{U-1}\sum_{n=1}^{N^u-1}\sum_{\boldsymbol{s}^u\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n-1)=(p,i), s^u(n)=(p,j)|\boldsymbol{x}^u,\theta'\right)}{\sum_{u=0}^{U-1}\sum_{n=1}^{N^u-1}\sum_{\boldsymbol{s}^u\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n-1)=(p,i)|\boldsymbol{x}^u,\theta'\right)} \tag{5.137}$$

この式に対して、$n-1$ を中心、つまり $n-1\rightarrow n$ として見直すと、$\sum_{n=1}^{N^u-1}$ は $\sum_{n=0}^{N^u-2}$ と書き換わり、式 (5.48) が導出されます。

$$\hat{a}_{i,j}^p = \frac{\sum_{u=0}^{U-1}\sum_{n=0}^{N^u-2}\sum_{\boldsymbol{s}^u\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n)=(p,i), s^u(n+1)=(p,j)|\boldsymbol{x}^u,\theta'\right)}{\sum_{u=0}^{U-1}\sum_{n=0}^{N^u-2}\sum_{\boldsymbol{s}^u\in\boldsymbol{l}^u} P\left(\boldsymbol{s}^u, s^u(n)=(p,i)|\boldsymbol{x}^u,\theta'\right)} \tag{5.138}$$