Variational Bayes-Hidden Markov Model Analysis for Time Series Data

Kenji Okamoto

March 24, 2016

Contents

Chapter 1

General Solution

1.1 Solution of Variational Bayes

1.1.1 Basic principle

最尤法では、得られた観測データ $\mathbf{X}$ に対して尤度関数 $p(\mathbf{X}|\Theta)$ を最大化するパラメータセット Θ を求めることを目的とする。

変分ベイズ法では、尤度関数をパラメータで周辺化したエビデンス

$$p(\mathbf{X}) = \int d\Theta\ p(\Theta)p(\mathbf{X}|\Theta) \tag{1.1}$$

を取り扱う。

エビデンス $p(\mathbf{X})$ は次のように変形できる。

$$\ln p(\mathbf{X}) = \mathcal{L}(q) + \mathrm{KL}(q||p) \tag{1.2}$$

ただし、

$$\mathcal{L}(q) = \sum_{\mathbf{Z}} \int d\Theta\ q(\mathbf{Z},\Theta) \ln \left\{ \frac{p(\mathbf{X},\mathbf{Z},\Theta)}{q(\mathbf{Z},\Theta)} \right\} \tag{1.3}$$

$$\mathrm{KL}(q||p) = -\sum_{\mathbf{Z}} \int d\Theta\ q(\mathbf{Z},\Theta) \ln \left\{ \frac{p(\mathbf{Z},\Theta|\mathbf{X})}{q(\mathbf{Z},\Theta)} \right\} \tag{1.4}$$

とする。ただし、$\mathbf{Z}$ は潜在変数、Θ はパラメータ。q は $\mathbf{Z}$ および Θ の分布関数であり、$\mathcal{L}(q)$, $\mathrm{KL}(q||p)$ は q の汎関数 (関数を引数とする関数)。

潜在変数とパラメータに関して、真の分布関数 p に近似できる分布関数 q を求めたい。この時、

- Kullback-Leibler ダイバージェンス $\mathrm{KL}(q||p)$ は、２つの分布関数 p, q の相似の程度を表す (小さいほど似通っている)。

- エビデンスの下限 $\mathcal{L}(q)$ 項と、$\mathrm{KL}(q||p)$ 項は、ともに非負。

- エビデンス $p(\mathbf{X})$ は既知の $\mathbf{X}$ にのみ依存しており、(未知ではあるが)１つの固定値を持っている。

といった点に注目すると、KL 項を最小化することと、 $\mathcal{L}(q)$ 項を最大化することが等価な目標であることが分かる。

変分ベイズ法では、エビデンスの下限値 (変分下限) $\mathcal{L}(q)$ を最大化する分布関数 q を求める。$\mathcal{L}(q)$ の q での汎関数微分 (変分) に基づいて解くので、変分ベイズと呼ぶ。

1.1.2 Mean field approximation

実際に解くには、物理学における平均場近似に近い考え方を用いる。

パラメータと潜在変数をまとめて $\mathbf{Z}$ とした時、

$$q(\mathbf{Z}) = \prod_{i=1}^{M} q_i(\mathbf{Z}_i) \tag{1.5}$$

に分解できるとすると、

$$\begin{aligned}
\mathcal{L}(q) &= \int q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{X}, \mathbf{Z}|K)}{q(\mathbf{Z})} \right\} d\mathbf{Z} && (1.6)\\
&= \int \prod_i q_i \left\{ \ln p(\mathbf{X}, \mathbf{Z}|K) - \sum_i \ln q_i \right\} d\mathbf{Z} && (1.7)\\
&= \int q_j \left\{ \int \ln p(\mathbf{X}, \mathbf{Z}|K) \prod_{i \neq j} q_i d\mathbf{Z}_i \right\} d\mathbf{Z}_j - \int q_i \ln q_i d\mathbf{Z}_i + \text{const.} && (1.8)\\
&= \int q_j \ln \tilde{p}(\mathbf{X}, \mathbf{Z}_j|K) d\mathbf{Z}_j - \int q_j \ln q_j d\mathbf{Z}_j + \text{const.} && (1.9)
\end{aligned}$$

ただし、$\mathbb{E}_{i \neq j}[\cdots]$ は期待値を表すとして、

$$\tilde{p}(\mathbf{X}, \mathbf{Z}_j) = \mathbb{E}_{i \neq j}[\ln p(\mathbf{X}, \mathbf{Z}|K)] + \text{const.} \tag{1.10}$$

$\{q_{i \neq j}\}$ を固定化 (平均場近似) した上で $\mathcal{L}(q)$ が最大化するのは、式 (1.9) が負の KL-ダイバージェンスの関係であることに注意すると、$q_i(\mathbf{Z}_j) \cong \tilde{p}(\mathbf{X}, \mathbf{Z}_j)$ の時なので、最適解 $q_j^*(\mathbf{Z}_j)$ は、

$$\ln q_j^*(\mathbf{Z}_j) = \mathbb{E}_{i \neq j}[\ln p(\mathbf{X}, \mathbf{Z}|K)] + \text{const.} \tag{1.11}$$

で与えられる。

1.1.3 変分ベイズ法の EM アルゴリズム

潜在変数 $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_N\}$ とパラメータ θ について

$$q(\mathbf{Z}, \theta) = q(\mathbf{Z})q(\theta) \tag{1.12}$$

に分離できるとすると、平均場近似によってそれぞれ、

$$\ln q^*(\mathbf{Z}) = \mathbb{E}_\theta[\ln p(\mathbf{X}, \mathbf{Z}, \theta|K)] + \text{const.} \tag{1.13}$$

を求める E-step と

$$\ln q^*(\theta) = \mathbb{E}_\mathbf{Z}[\ln p(\mathbf{X}, \mathbf{Z}, \theta|K)] + \text{const.} \tag{1.14}$$

を求める M-step を繰り返せばよい。

1.1.4 下限値 $\mathcal{L}(q)$ の算出

式 (1.2) で表されるエビデンスの中で、最大値の下限を表す $\mathcal{L}(q)$ 項は、最適化ステップが進むとともに単調に増加する。そこでこの値をステップ毎に計算することで、計算が正しく行われていることの確認になる。

また、最適化が完了した時にはエビデンスの近似値となるので、収束の確認、モデル間の比較等をおこなうためにも $\mathcal{L}(q)$ の値を計算する必要がある。

$\mathcal{L}(q)$ はさらに、

$$\mathcal{L}(q) = \sum_{\mathbf{z}} \int d\theta \; q(\mathbf{Z}, \theta) \ln p(\mathbf{X}, \mathbf{Z}, \theta) - \sum_{\mathbf{z}} \int d\theta \; q(\mathbf{Z}, \theta) \cdot \ln q(\mathbf{Z}, \theta) \tag{1.15}$$

$$= \mathbb{E}_{\mathbf{Z},\theta}\big[\ln p(\mathbf{X}, \mathbf{Z}, \theta)\big] - \mathbb{E}_{\mathbf{Z},\theta}\big[\ln q(\mathbf{Z}, \theta)\big] \tag{1.16}$$

と展開でき、実際にはこれを計算する。

EM-アルゴリズムでは、E-step で $\mathbf{Z}$ 分布を更新した直後に変分下限 $\mathcal{L}(q)$ を計算し、収束判定等をおこなう。

1.1.5 モデル選択、混合要素数の決定

変分ベイズ法は複数のモデル間での最適モデル選択に用いることもできる。モデル m に対して、規格化された変分下限 $\mathcal{L}(q|m)$ を求め、

$$p(m)\mathcal{L}(q|m) \tag{1.17}$$

を計算して比較すればよい。

$p(m)$ は各モデルについての事前分布であり、各 m について $p(m) = 1$ とすることができれば、直接 $\mathcal{L}(q|m)$ を比較すればよいことになる。

混合ガウス分布のように、等価な複数の分布の混合分布を考える場合、最適解から混合要素同士を入れ替えた別の設定もまた最適解となる。そのような等価なパラメータ設定の組み合わせは、K 個の混合要素を考えた場合、$K!$ 通りとなる。

ちなみに、最尤推定の場合には、この冗長性は問題にならない。最適解は、パラメータの初期値に依存して1つだけ求められ、それ以外の等価な解は無関係となる。

この影響を取り除くためには、最も簡単な近似は、得られた変分下限に $\ln K!$ を加えることである。

したがって、要素数の異なる混合分布の間でモデル選択する場合には、

$$p(K)\{\mathcal{L}(q|K)+\ln K!\} \tag{1.18}$$

を各 K について求めて比較すればよい。

1.2 Hidden Markov Model の解法

1.2.1 状態を表す潜在変数の定義

全部で N 個あるデータ点 $\mathbf{X}=\{x_1,x_2,\ldots,x_N\}$ それぞれが、$1\ldots K$ 状態のいずれかに属するとする。

$N\times K$ ベクトルとして潜在変数 $\mathbf{Z}=\{\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_N\}, \mathbf{z}_n=\{z_{n1},z_{n2},\ldots,z_{nK}\}$ を定義する。各データ点がどの状態に属するかを表すため、

$$z_{nk}=\begin{cases}0\\1\end{cases} \tag{1.19}$$

$$\sum_{i=1}^{K} z_{ni}=1 \tag{1.20}$$

とする。つまり、各 n に対して1つだけが 1 となり、残りは 0 となる。

1.2.2 一般式

一般的な (たとえば混合 Gauss 分布) の HMM を解く場合には、$\mathbf{X},\mathbf{Z}$ の同時分布を以下のように書き下す。

$$p(\mathbf{X},\mathbf{Z}|\theta)=p(\mathbf{z}_1|\theta)\times\prod_{n=2}^{N}p(\mathbf{z}_n|\mathbf{z}_{n-1},\theta)\times\prod_{m=1}^{N}p(x_m|\mathbf{z}_m,\theta) \tag{1.21}$$

このうち、$p(\mathbf{z}_1|\theta)$ と $p(\mathbf{z}_n|\mathbf{z}_{n-1},\theta)$ に関しては一般的に、

$$p(\mathbf{z}_1|\boldsymbol{\pi})=\prod_{i=1}^{K}{\pi_i}^{z_{1i}} \tag{1.22}$$

$$p(\mathbf{z}_n|\mathbf{z}_{n-1},\mathbf{A})=\prod_{i=1}^{K}\prod_{j=1}^{K}{A_{ij}}^{z_{n-1,i}z_{nj}} \tag{1.23}$$

とすることができる。ただし、π_i は初期状態が i 状態である確率を表すパラメータで、$\sum_{i=1}^{K}\pi_i = 1$ 。A_{ij} は i 状態から次のステップで j 状態に遷移する確率 ($i = j$ の場合 i 状態に留まる確率) を表すパラメータ行列で、$\sum_{j=1}^{K} A_{ij} = 1$ 。

$\boldsymbol{\pi}, \mathbf{A}$ 以外の、主に $p(x_m|\mathbf{z}_m)$ に関わるパラメータを、まとめて ϕ で表す。

1.2.3 EM アルゴリズムによる HMM 解法

HMM を最尤法で解くためには、式 (1.21) 尤度関数を最大化するパラメータセット θ を求める必要がある。しかし、$\mathbf{Z}$ が未知のため、このままでは解くことが出来ない。

そこで、仮に与えたパラメータ θ^{old} を基にして $\mathbf{Z}$ の期待値を求めることにして、式 (1.21) の対数尤度を書き直した

$$Q(\theta,\theta^{\mathrm{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\theta^{\mathrm{old}})\cdot \ln p(\mathbf{X},\mathbf{Z}|\theta) \tag{1.24}$$

を求める。

ここで新たな変数 γ,ξ を導入する。

$$\gamma(\mathbf{z}_n) = p(\mathbf{z}_n|\mathbf{X},\theta^{\mathrm{old}}) \tag{1.25}$$

$$\gamma(z_{nk}) = \mathbb{E}[z_{nk}] = \sum_{\mathbf{z}_n}\gamma(\mathbf{z}_n)\cdot z_{nk} \tag{1.26}$$

$$\xi(\mathbf{z}_n,\mathbf{z}_{n-1}) = p(\mathbf{z}_{n-1},\mathbf{z}_n|\mathbf{X},\theta^{\mathrm{old}}) \tag{1.27}$$

$$\xi(z_{n-1,i},z_{nj}) = \mathbb{E}[z_{n-1,i}z_{nj}] = \sum_{\mathbf{z}_n,\mathbf{z}_{n-1}}\xi(\mathbf{z}_n,\mathbf{z}_{n-1})\cdot z_{n-1,i}z_{nj} \tag{1.28}$$

$\gamma(z_{nk})$ は、n 番目のデータ点が k 状態に属する確率、$\xi(z_{n-1,i},z_{nj})$ は $n-1$ から n の間に i 状態から j 状態に遷移する確率を表す。

これにより、

$$\begin{aligned} Q(\theta,\theta^{\mathrm{old}}) = &\sum_{i=1}^{K}\gamma(z_{1i})\ln\pi_i + \sum_{n=2}^{N}\sum_{i=1}^{K}\sum_{j=1}^{K}\xi(z_{n-1,i},z_{nj})\ln A_{ij} \\ &+ \sum_{n=1}^{N}\sum_{i=1}^{K}\gamma(z_{ni})\ln p(x_n|\theta) \end{aligned} \tag{1.29}$$

と書き直す。

EM アルゴリズムでは、仮の θ_{old} を基に γ,ξ 分布を求める E-step と、得られた γ,ξ 分布から θ の最尤値を更新する M-step とを、M-step で得られた θ を次の E-step の θ_{old} として与えることで、値が収束するまで反復計算する。

E-step

E-step では、γ,ξ を求める。

Forward-backward (Baum-Welch) アルゴリズム：
γ は

$$\gamma(\mathbf{z}_n) = \frac{p(\mathbf{X}|\mathbf{z}_n) \cdot p(\mathbf{z}_n)}{p(\mathbf{X})} \tag{1.30}$$

$$= \frac{\alpha(\mathbf{z}_n) \cdot \beta(\mathbf{z}_n)}{p(\mathbf{X})} \tag{1.31}$$

と書くことができる。ただし、

$$\alpha(\mathbf{z}_n) \equiv p(x_1, \ldots, x_n, \mathbf{z}_n) \tag{1.32}$$

$$\beta(\mathbf{z}_n) \equiv p(x_{n+1}, \ldots, x_N|\mathbf{z}_n) \tag{1.33}$$

と定義する。ここで、スケーリング係数 c_n を導入する。

$$c_n = p(x_n|x1, x2, \ldots, x_{n-1}) \tag{1.34}$$

とすると、乗法定理より

$$p(x1, x2, \ldots, x_n) = \prod_{m=1}^{n} c_m \tag{1.35}$$

これを用いて α, β をそれぞれ

$$\hat{\alpha}(\mathbf{z}_n) = \frac{\alpha(\mathbf{z}_n)}{\prod_{m=1}^{n} c_m} = p(\mathbf{z}_n|x_1, x_2, \ldots, x_n) \tag{1.36}$$

$$\hat{\beta}(\mathbf{z}_n) = \frac{\beta(\mathbf{z}_n)}{\prod_{m=n+1}^{N} c_m} = \frac{p(\mathbf{x}_{n+1}, \ldots, x_N|\mathbf{z}_n)}{\prod_{m=n+1}^{N} c_m} \tag{1.37}$$

$\hat{\alpha}$ に関しては、

$$\hat{\alpha}(\mathbf{z}_n) = \frac{p(x_n|\mathbf{z}_n)}{c_n} \sum_{\mathbf{z}_{n-1}} \hat{\alpha}(\mathbf{z}_{n-1}) \cdot p(\mathbf{z}_n|\mathbf{z}_{n-1}) \tag{1.38}$$

と再帰式に展開することができ、

$$\hat{\alpha}(\mathbf{z}_1) = \frac{p(x_1|\mathbf{z}_1)}{p(x_1)} = \frac{p(x_1|\mathbf{z}_1)}{\sum_{\mathbf{z}_1} p(x_1|\mathbf{z}_1)} \tag{1.39}$$

を得ることができるので、全領域で $\hat{\alpha}$ が得られる。β に関しても、

$$\hat{\beta}(\mathbf{z}_n) = \frac{1}{c_n} \sum_{\mathbf{z}_{n+1}} \hat{\beta}(\mathbf{z}_{n+1}) \cdot p(x_{n+1}|\mathbf{z}_{n+1}) \cdot p(\mathbf{z}_{n+1}|\mathbf{z}_n) \tag{1.40}$$

と再帰式に展開することができ、

$$\hat{\beta}(z_{Nk}) = 1 \tag{1.41}$$

とすればよいので、全領域で $\hat{\beta}$ が得られる。

$\hat{\alpha}, \hat{\beta}$ を用いると、式 (1.31) を書き直して、γ に関して、

$$\gamma(\mathbf{z}_n) = \hat{\alpha}(\mathbf{z}_n) \cdot \hat{\beta}(\mathbf{z}_n) \tag{1.42}$$

ξ に関しても

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = \frac{1}{c_n} \hat{\alpha}(\mathbf{z}_{n-1}) \cdot p(x_n|\mathbf{z}_n) \cdot p(\mathbf{z}_n|\mathbf{z}_{n-1}) \cdot \hat{\beta}(\mathbf{z}_n) \tag{1.43}$$

により、全領域で得ることができる。

また、

$$p(\mathbf{X}) = \prod_{n=1}^{N} c_n \tag{1.44}$$

から尤度を得ることが出来る。

M-step

E-step で得られた γ, ξ を用いて、パラメータ値を更新する。

$$\pi_k = \frac{\gamma(z_{1i})}{\sum_{j=1}^{K} \gamma(z_{1j})} \tag{1.45}$$

$$A_{ij} = \frac{\sum_{n=2}^{N} \xi(z_{n-1,i}, z_{nj})}{\sum_{n=2}^{N} \sum_{k=1}^{K} \xi(z_{n-1,i}, z_{nk})} \tag{1.46}$$

ϕ に関しては、$p(x_n|\mathbf{z}_n)$ の関数形に依存する。ラグランジュ乗数を用いるなどして更新式を求める。

Max-Sum (Viterbi) アルゴリズム

各 n について最大の $\gamma(z_{nk})$ を選ぶことが、最適な軌跡を復元することにはならない。$\hat{\theta}$ を最尤パラメータとして、$p(\mathbf{Z}|\mathbf{X}, \hat{\theta})$ を最大化する $\mathbf{Z}$ を選ぶ必要がある。

そのために、次の再帰式を計算する。

$$\omega(\mathbf{z}_n) = \ln p(x_n|\mathbf{z}_n) + \max_{\mathbf{z}_{n-1}}\left\{\ln p(\mathbf{z}_n|\mathbf{z}_{n-1}) + \omega(\mathbf{z}_{n-1})\right\} \tag{1.47}$$

$$\omega(\mathbf{z}_1) = \ln p(\mathbf{z}_1) + \ln p(x_1|\mathbf{z}_1) \tag{1.48}$$

またその再帰計算の際、

$$\phi(z_{nj}^{\max}) = z_{n-1,i}^{\max} \tag{1.49}$$

を記録しておく。

最後に、$n = N$ から ϕ を利用して順に遡りながら、$p(\mathbf{Z}|\mathbf{X}, \hat{\theta})$ を最大化する経路を復元する。

1.3 Solution of Variational Bayes-Hidden Markov Model

1.3.1 E-step

最尤法の場合と同様に、forward-backward アルゴリズムにより、γ, ξ 分布を得ることが目的。

一般的な HMM で用いられる forward-backward アルゴリズムとは、要するに、$\mathbf{Z}$ 分布を規定する同時分布 $p(\mathbf{X}, \mathbf{Z}|\theta)$ から γ, ξ 分布を得る計算法である。

変分ベイズでは、$\mathbf{Z}$ 分布は $q(\mathbf{Z})$ 関数によって与えられるので、(平均場近似のためパラメータで期待値をとった後の) $q(\mathbf{Z})$ 関数に対して forward-backward アルゴリズムを適用する。

$q^*(\mathbf{Z})$ を計算すると、

$$\begin{aligned}\ln q^*(\mathbf{Z}) =& \mathbb{E}_\pi\left[\sum_i^K z_{1i}\ln\pi_i\right] + \mathbb{E}_A\left[\sum_{n=2}^N\sum_i^K\sum_j^K z_{n-1,i}z_{nj}\ln A_{ij}\right] \\ &+ \mathbb{E}_\theta\left[\sum_{m=1}^N \ln p(x_m|\mathbf{z}_m, \theta, K)\right] + \text{const.} \end{aligned} \tag{1.50}$$

$$\begin{aligned}=& \sum_i^K z_{1i}\overline{\ln\pi_i} + \sum_{n=2}^N\sum_i^K\sum_j^K z_{n-1,i}z_{nj}\overline{\ln A_{ij}} \\ &+ \sum_{m=1}^N \overline{\ln p(x_m|\mathbf{z}_m, \theta, K)} + \text{const.}\end{aligned} \tag{1.51}$$

したがって、

$$
\begin{aligned}
q^*(\mathbf{Z}) \propto & \prod_{i}^{K} \exp(\overline{\ln \pi_i})^{z_{1i}} \times \prod_{n=2}^{N} \prod_{i}^{K} \prod_{j}^{K} \exp\big(\overline{\ln A_{ij}}\big)^{z_{n-1,i} z_{nj}} \\
& \times \prod_{m=1}^{N} \exp\left(\overline{\ln p(x_m|\mathbf{z}_m, \theta, K)}\right) \qquad (1.52)
\end{aligned}
$$

となる。そこで、

$$
\tilde{p}(\mathbf{z}_1|\theta, K) = \prod_{i}^{K} \exp(\overline{\ln \pi_i})^{z_{1i}} \qquad (1.53)
$$

$$
\tilde{p}(\mathbf{z}_n|\mathbf{z}_{n-1}, \theta, K) = \prod_{i}^{K} \prod_{j}^{K} \exp\big(\overline{\ln A_{ij}}\big)^{z_{n-1,i} z_{nj}} \qquad (1.54)
$$

$$
\tilde{p}(x_m|\mathbf{z}_m, \theta, K) = \exp\left(\overline{\ln p(x_m|\mathbf{z}_m, \theta, K)}\right) \qquad (1.55)
$$

をそれぞれ $p(\mathbf{z}_1, \theta, K)$, $p(\mathbf{z}_n|\mathbf{z}_{n-1}, \theta, K)$, $p(x_m|\mathbf{z}_m, \theta, K)$ の代わりに用いて forward-backward アルゴリズムを実行し、γ, ξ 分布を得る。

各パラメータには期待値 ($\overline{\ln \pi_i}$, $\overline{\ln A_{ij}}$ 等) を与えなければならないが、計算の１ステップ目では、乱数等で適当に与えればよい。反復計算中は、直前の M-step で得られた値を用いる。

1.3.2 M-step

各パラメータについて分布関数 q を更新し、期待値を求める。

$\mathbf{X}, \mathbf{Z}, \theta$ の同時分布は、

$$
p(\mathbf{X}, \mathbf{Z}, \theta|K) = p(\theta|K) \times p(\mathbf{X}, \mathbf{Z}|\theta, K) \qquad (1.56)
$$

$$
\begin{aligned}
&= p(\theta|K) \times p(\mathbf{z}_1|\theta, K) \times \prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}, \theta, K) \\
&\quad \times \prod_{m=1}^{N} p(x_m|\mathbf{z}_m, \theta, K) \qquad (1.57)
\end{aligned}
$$

と書けるので、$\mathbf{Z}$ を固定する平均場近似を用いて

$$
\begin{aligned}
\ln q^*(\theta) = & \mathbb{E}_{\mathbf{Z}}\Bigg[\ln p(\theta) + \ln p(\mathbf{z}_1) + \sum_{n=2}^{N} \ln p(\mathbf{z}_n|\mathbf{z}_{n-1}) \\
& + \sum_{m=1}^{N} \ln p(x_m|\mathbf{z}_m)\Bigg] \qquad (1.58)
\end{aligned}
$$

により q^* を得る。

パラメータ $\boldsymbol{\pi}$, $\mathbf{A}$, ϕ がそれぞれ独立、かつ、$p(x_m|\mathbf{z}_m, \theta, K)$ は $\boldsymbol{\pi}$, $\mathbf{A}$ に依存しないとすると、まず $\boldsymbol{\pi}$ に関して、

$$\begin{aligned}
\ln q^*(\boldsymbol{\pi}) =& \ln p(\boldsymbol{\pi}) + \mathbb{E}_{\mathbf{Z}}\left[\ln p(\mathbf{z}_1)\right] + \text{const.} && (1.59)\\
=& \ln p(\boldsymbol{\pi}) + \mathbb{E}_{\mathbf{Z}}\left[\sum_i^K z_{1i} \ln \pi_i\right] + \text{const.} && (1.60)\\
=& \ln p(\boldsymbol{\pi}) + \sum_i^K \overline{z_{1i}} \cdot \ln \pi_i + \text{const.} && (1.61)
\end{aligned}$$

となり、

$$q^*(\boldsymbol{\pi}) \propto p(\boldsymbol{\pi}) \times \prod_i^K {\pi_i}^{\overline{z_{1i}}} \qquad (1.62)$$

ここで、事前分布として Dirichlet 分布 $p(\boldsymbol{\pi}) \propto \prod_i^K {\pi_i}^{(u_i^\pi - 1)}$ を与えて規格化して、

$$q^*(\boldsymbol{\pi}) = \frac{\Gamma\left(u_0^\pi + 1\right)}{\prod_i^K \Gamma\left(u_i^\pi + \overline{z_{1i}}\right)} \prod_i^K {\pi_i}^{(u_i^\pi + \overline{z_{1i}} - 1)} \qquad (1.63)$$

ただし、u_i^π は Dirichlet 分布のハイパーパラメータ。特に理由が無ければ、$u_i^\pi = 1$ としておけばよい。

Dirichlet 分布の性質から、期待値

$$\begin{aligned}
\overline{\pi_i} =& \frac{u_i^\pi + \overline{z_{1i}}}{u_0^\pi + 1} && (1.64)\\
\overline{\ln \pi_i} =& \psi\left(u_i^\pi + \overline{z_{1i}}\right) - \psi\left(u_0^\pi + 1\right) && (1.65)
\end{aligned}$$

を得る。ただし、$u_0^\pi = \sum_i^K u_i^\pi$、$\psi(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ は digamma 関数。

次に $\mathbf{A}$ に関して、

$$\begin{aligned}
\ln q^*(\mathbf{A}) =& \ln p(\mathbf{A}) + \mathbb{E}_{\mathbf{Z}}\left[\sum_{n=2}^N \ln p(\mathbf{z}_n|\mathbf{z}_{n-1})\right] + \text{const.} && (1.66)\\
=& \ln p(\mathbf{A}) + \mathbb{E}_{\mathbf{Z}}\left[\sum_{n=2}^N \sum_i^K \sum_j^K z_{n-1,i} z_{nj} \ln A_{ij}\right] + \text{const.} && (1.67)\\
=& \ln p(\mathbf{A}) + \sum_{n=2}^N \sum_i^K \sum_j^K \overline{z_{n-1,i} z_{nj}} \cdot \ln A_{ij} + \text{const.} && (1.68)
\end{aligned}$$

となり、

$$q^*(\mathbf{A}) \propto p(\mathbf{A}) \times \prod_i^K \prod_j^K {A_{ij}}^{N_{ij}} \tag{1.69}$$

となる。ただし、$N_{ij} = \sum_{n=2}^N \overline{z_{n-1,i} z_{nj}}$。ここで、事前分布として各 i での j に関する Dirichlet 分布 $p(\mathbf{A}_i) \propto \prod_j^K {A_{ij}}^{(u_{ij}^A - 1)}$ を与えて規格化して、

$$q^*(\mathbf{A}_i) = \frac{\Gamma\left(u_{i0}^A + M_i\right)}{\prod_j^K \Gamma\left(u_{ij}^A + N_{ij}\right)} \prod_j^K {A_{ij}}^{(u_{ij}^A + N_{ij} - 1)} \tag{1.70}$$

を得る。ただし、u_{ij}^A は Dirichlet 分布のハイパーパラメータ、$u_{i0}^A = \sum_j^K u_{ij}^A$, $M_i = \sum_j^K N_{ij} = \sum_{n=2}^N \sum_j^K \overline{z_{n-1,i} z_{nj}}$。

Dirichlet 分布の性質から、期待値

$$\overline{A_{ij}} = \frac{u_{ij}^A + N_{ij}}{u_{i0}^A + M_i} \tag{1.71}$$

$$\overline{\ln A_{ij}} = \psi\left(u_{ij}^A + N_{ij}\right) - \psi\left(u_{i0}^A + M_i\right) \tag{1.72}$$

を得る。

ϕ に関しては、$p(x_n|\mathbf{z}_n)$ の関数形に依存する。同様に適宜計算。

1.3.3 変分下限 $\mathcal{L}(q)$ の算出

変分下限 $\mathcal{L}(q)$ の計算において、$\mathbf{Z}, \theta$ が独立であると見なせる場合、式 (1.16) は、

$$\mathcal{L}(q) = \mathbb{E}\left[\ln p(\theta)\right] - \mathbb{E}\left[\ln q(\theta)\right] + \mathbb{E}\left[\ln p(\mathbf{X}, \mathbf{Z}|\theta)\right] - \mathbb{E}\left[\ln q(\mathbf{Z})\right] \tag{1.73}$$

$$= \mathbb{E}\left[\ln p(\theta)\right] - \mathbb{E}\left[\ln q(\theta)\right] + \sum_{n=1}^N \ln \tilde{c}_n \tag{1.74}$$

ただし $\tilde{c}_n$ は、$\tilde{p}(\mathbf{z}_1)$, $\tilde{p}(\mathbf{z}_n|\mathbf{z}_{n-1})$, $\tilde{p}(x_m|\mathbf{z}_m)$ を用いて α-β アルゴリズムを実行した際に得られるスケーリング係数であり、E-step の計算過程で既に得られている。

式 (1.74) を既知のパラメータについて書き下すと、

$$\begin{aligned}\mathcal{L}(q) =& \mathbb{E}\left[\ln p(\boldsymbol{\pi})\right] + \mathbb{E}\left[\ln p(\mathbf{A})\right] + \mathbb{E}\left[\ln p(\phi)\right] \\ &- \mathbb{E}\left[\ln q(\boldsymbol{\pi})\right] - \mathbb{E}\left[\ln q(\mathbf{A})\right] - \mathbb{E}\left[\ln q(\phi)\right] + \sum_{n=1}^N \ln \tilde{c}_n\end{aligned} \tag{1.75}$$

となるので、その他のパラメータに関して必要な項をそれぞれ計算する。

$\mathbb{E}\big[\ln p(\boldsymbol{\pi})\big]$ **:**　$\boldsymbol{\pi}$ の事前分布は Dirichlet 分布で与えられるので、

$$
\begin{aligned}
\mathbb{E}\big[\ln p(\boldsymbol{\pi})\big] =& \mathbb{E}_\pi\big[\ln p(\boldsymbol{\pi})\big] && (1.76)\\
=& \mathbb{E}_\pi\left[\ln\left\{\frac{\Gamma(u_0^\pi)}{\prod_i^K \Gamma(u_i^\pi)}\prod_i^K \pi_i^{u_i^\pi - 1}\right\}\right] && (1.77)\\
=& \mathbb{E}_\pi\left[\left\{\ln\Gamma(u_0^\pi) - \sum_i^K \ln\Gamma(u_i^\pi) + \sum_i^K (u_i^\pi - 1)\cdot \ln \pi_i\right\}\right] && (1.78)\\
=& \ln\Gamma(u_0^\pi) + \sum_i^K \left\{(u_i^\pi - 1)\cdot \overline{\ln \pi_i} - \ln\Gamma(u_i^\pi)\right\} && (1.79)
\end{aligned}
$$

$\mathbb{E}\big[\ln p(\mathbf{A})\big]$ **:**　$\mathbf{A}$ についても、事前分布は Dirichlet 分布で与えられるので、

$$
\begin{aligned}
\mathbb{E}\big[\ln p(\mathbf{A})\big] =& \mathbb{E}_A\big[\ln p(\mathbf{A})\big] && (1.80)\\
=& \mathbb{E}_A\left[\ln\left\{\prod_i^K\left(\frac{\Gamma(u_{i0}^A)}{\prod_j^K \Gamma(u_{ij}^A)}\prod_j^K A_{ij}^{u_{ij}^A - 1}\right)\right\}\right] && (1.81)\\
=& \mathbb{E}_A\left[\sum_i^K\left\{\ln\Gamma(u_{i0}^A) - \sum_j^K \ln\Gamma(u_{ij}^A) + \sum_j^K (u_{ij}^A - 1)\ln A_{ij}\right\}\right] && (1.82)\\
=& \sum_i^K\left[\ln\Gamma(u_{i0}^A) + \sum_j^K\left\{(u_{ij}^A - 1)\cdot\overline{\ln A_{ij}} - \ln\Gamma(u_{ij}^A)\right\}\right] && (1.83)
\end{aligned}
$$

$\mathbb{E}\big[\ln q(\boldsymbol{\pi})\big]$ **:**　$q(\boldsymbol{\pi}) = \mathrm{Dir}(\pi_i|u_i^\pi + \overline{z_{1i}})$ なので、Dirichlet 分布の性質から、

$$
\begin{aligned}
\mathbb{E}\big[\ln q(\boldsymbol{\pi})\big] =& \mathbb{E}_\pi\big[\ln q(\boldsymbol{\pi})\big] && (1.84)\\
=& \sum_i^K\left[(u_i^\pi + \overline{z_{1i}} - 1)\left\{\psi(u_i^\pi + \overline{z_{1i}}) - \psi\left(u_0^\pi + \sum_i^K \overline{z_{1i}}\right)\right\}\right]\\
&+ \ln\left\{\frac{\Gamma(u_0^\pi + \sum_i^K \overline{z_{1i}})}{\prod_i^K \Gamma(u_i^\pi + \overline{z_{1i}})}\right\} && (1.85)\\
=& \ln\Gamma(u_0^\pi + 1) + \sum_i^K\Big[(u_i^\pi + \overline{z_{1i}} - 1)\big\{\psi(u_i^\pi + \overline{z_{1i}})\\
&- \psi(u_0^\pi + 1)\big\} - \ln\Gamma(u_i^\pi + \overline{z_{1i}})\Big] && (1.86)
\end{aligned}
$$

$\mathbb{E}\big[\ln q(\mathbf{A})\big]$ **:** $q(\mathbf{A}_i) = \mathrm{Dir}(A_{ij}|u^A_{ij} + N_{ij})$ なので、Dirichlet 分布の性質から、

$$\mathbb{E}\big[\ln q(\mathbf{A})\big] = \mathbb{E}_A\big[\ln q(\mathbf{A})\big] \tag{1.87}$$

$$= \sum_i^K \mathbb{E}_{A_i}\big[\ln q(\mathbf{A}_i)\big] \tag{1.88}$$

$$= \sum_i^K \left[\sum_j^K \left\{ (u^A_{ij} + N_{ij} - 1)\Big[\psi(u^A_{ij} + N_{ij}) - \psi(u^A_{i0} + M_i)\Big] \right\} + \ln\left\{ \frac{\Gamma(u^A_{i0} + M_i)}{\prod_j^K \Gamma\big(u^A_{ij} + N_{ij}\big)} \right\} \right] \tag{1.89}$$

$$= \sum_i^K \Bigg[\ln\Gamma(u^A_{i0} + M_i) + \sum_j^K \Big\{ \big(u^A_{ij} + N_{ij} - 1\big)\Big[\psi\big(u^A_{ij} + N_{ij}\big) - \psi\big(u^A_{i0} + M_i\big)\Big] - \ln\Gamma\big(u^A_{ij} + N_{ij}\big)\Big\} \Bigg] \tag{1.90}$$

$\mathbb{E}\big[\ln p(\phi)\big], \mathbb{E}\big[\ln q(\phi)\big]$ については、関数形に依存するので、$\boldsymbol{\pi}$, $\mathbf{A}$ と同様にそれぞれ計算する。

1.4 Global Hidden Markov Model の変分ベイズ解法

R 本の時系列データからグローバルに解析する方法。パラメータは共通とする。

データ $\mathbf{X}$ が、

$$\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(R)}\} \tag{1.91}$$

$$\mathbf{X}^{(r)} = \{x^{(r)}_1, x^{(r)}_2, \ldots, x^{(r)}_{N^{(r)}}\} \tag{1.92}$$

とする。ただし、$1 \leq r \leq R$。潜在変数 $\mathbf{Z}$ も同様に

$$\mathbf{Z} = \{\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots, \mathbf{Z}^{(R)}\} \tag{1.93}$$

$$\mathbf{Z}^{(r)} = \{\mathbf{z}^{(r)}_1, \mathbf{z}^{(r)}_2, \ldots, \mathbf{z}^{(r)}_{N^{(r)}}\} \tag{1.94}$$

に拡張する。

この時 $\mathbf{X},\mathbf{Z},\theta$ の同時分布は、

$$
\begin{aligned}
p(\mathbf{X},\mathbf{Z},\theta) &= p(\theta)\cdot p(\mathbf{X},\mathbf{Z}|\theta) && (1.95)\\
&= p(\theta)\cdot \prod_{r=1}^{R} p(\mathbf{X}^{(r)},\mathbf{Z}^{(r)}|\theta) && (1.96)\\
&= p(\theta)\cdot \prod_{r=1}^{R} \left\{ p(\mathbf{z}_1^{(r)}|\theta) \times \prod_{n=2}^{N^{(r)}} p(\mathbf{z}_n^{(r)}|\mathbf{z}_{n-1}^{(r)},\theta) \times \prod_{m=1}^{N^{(r)}} p(x_m^{(r)}|\mathbf{z}_m^{(r)},\theta) \right\} && (1.97)\\
& && (1.98)
\end{aligned}
$$

$$
\ln p(\mathbf{X},\mathbf{Z},\theta) = \ln p(\theta) + \sum_{r=1}^{R}\left\{ \ln p(\mathbf{z}_1^{(r)}|\theta) + \sum_{n=2}^{N^{(r)}} \ln p(\mathbf{z}_n^{(r)}|\mathbf{z}_{n-1}^{(r)},\theta) + \sum_{m=1}^{N^{(r)}} \ln p(x_m^{(r)}|\mathbf{z}_m^{(r)},\theta) \right\} \tag{1.99}
$$

1.4.1 E-step

θ の平均場近似で $q^*(\mathbf{Z})$ を計算すると、

$$
\ln q^*(\mathbf{Z}) = \mathbb{E}_\theta\left[\sum_{r=1}^{R}\left\{ \ln p(\mathbf{z}_1^{(r)}|\theta) + \sum_{n=2}^{N^{(r)}} \ln p(\mathbf{z}_n^{(r)}|\mathbf{z}_{n-1}^{(r)},\theta) + \sum_{m=1}^{N^{(r)}} \ln p(x_m^{(r)}|\mathbf{z}_m^{(r)},\theta) \right\}\right] + \text{const.} \tag{1.100}
$$

$$
\begin{aligned}
&= \mathbb{E}_\theta\left[\sum_{r=1}^{R} \ln p(\mathbf{z}_1^{(r)}|\theta)\right] + \mathbb{E}_\theta\left[\sum_{r=1}^{R}\sum_{n=2}^{N^{(r)}} \ln p(\mathbf{z}_n^{(r)}|\mathbf{z}_{n-1}^{(r)},\theta)\right]\\
&\quad + \mathbb{E}_\theta\left[\sum_{r=1}^{R}\sum_{m=1}^{N^{(r)}} \ln p(x_m^{(r)}|\mathbf{z}_m^{(r)},\theta)\right] + \text{const.} \qquad (1.101)
\end{aligned}
$$

$p(\mathbf{z}_1^{(r)}|\theta)=\prod_{i=1}^K \pi_i^{z_{1i}^{(r)}}$, $p(\mathbf{z}_n^{(r)}|\mathbf{z}_{n-1}^{(r)},\theta)=\prod_{i=1}^K\prod_{j=1}^K A_{ij}^{z_{n-1,i}^{(r)},z_{nj}^{(r)}}$ とすると、

$$=\mathbb{E}_\pi\left[\sum_{r=1}^R\sum_{i=1}^K z_{1i}^{(r)}\ln\pi_i\right]+\mathbb{E}_A\left[\sum_{r=1}^R\sum_{n=2}^{N^{(r)}}\sum_{i=1}^K\sum_{j=1}^K z_{n-1,i}^{(r)}z_{nj}^{(r)}\ln A_{ij}\right]$$
$$+\mathbb{E}_\theta\left[\sum_{r=1}^R\sum_{m=1}^{N^{(r)}}\ln p(x_m^{(r)}|\mathbf{z}_m^{(r)},\theta)\right]+\text{const.} \tag{1.102}$$
$$=\sum_{r=1}^R\left\{\sum_{i=1}^K z_{1i}^{(r)}\overline{\ln\pi_i}+\sum_{n=2}^{N^{(r)}}\sum_{i=1}^K\sum_{j=1}^K z_{n-1,i}^{(r)}z_{nj}^{(r)}\overline{\ln A_{ij}}+\sum_{m=1}^{N^{(r)}}\overline{\ln p(x_m^{(r)}|\mathbf{z}_m^{(r)},\theta)}\right\}+\text{const.} \tag{1.103}$$

したがって、

$$q^*(\mathbf{Z})\propto\prod_{r=1}^R\times\left\{\prod_{i=1}^K\exp\left(\overline{\ln\pi_i}\right)^{z_{1i}^{(r)}}\times\prod_{n=2}^{N^{(r)}}\prod_{i=1}^K\prod_{j=1}^K\exp\left(\overline{\ln A_{ij}}\right)^{z_{n-1,i}^{(r)}z_{nj}^{(r)}}\right.$$
$$\left.\times\prod_{m=1}^{N^{(r)}}\exp\left(\overline{\ln p(x_m^{(r)}|\mathbf{z}_m^{(r)},\theta)}\right)\right\} \tag{1.104}$$
$$\tag{1.105}$$

各 r について独立なので、単独の HMM と同様にそれぞれ $\gamma^{(r)},\xi^{(r)}$ を求める。

1.4.2 M-step

各パラメータについて分布関数 q を更新し、期待値を求める。

$\mathbf{X},\mathbf{Z},\theta$ の同時分布

$$p(\mathbf{X},\mathbf{Z},\theta)=p(\theta)\cdot p(\mathbf{X},\mathbf{Z}|\theta) \tag{1.106}$$
$$=p(\theta)\cdot\prod_{r=1}^R\left\{p(\mathbf{X}^{(r)},\mathbf{Z}^{(r)}|\theta)\right\} \tag{1.107}$$
$$=p(\theta)\cdot\prod_{r=1}^R\left\{p(\mathbf{z}_1^{(r)}|\theta)\times\prod_{n=2}^{N^{(r)}}p(\mathbf{z}_n^{(r)}|\mathbf{z}_{n-1}^{(r)},\theta)\times\prod_{m=1}^{N^{(r)}}p(x_m^{(r)}|\mathbf{z}_m^{(r)},\theta)\right\} \tag{1.108}$$

から、$\mathbf{Z}$ を固定する平均場近似を用いて

$$\ln q^*(\theta)=\mathbb{E}_\mathbf{Z}\left[\ln p(\theta)+\sum_{r=1}^R\left\{\ln p(\mathbf{z}_1^{(r)})+\sum_{n=2}^N\ln p(\mathbf{z}_n^{(r)}|\mathbf{z}_{n-1}^{(r)})+\sum_{m=1}^N\ln p(x_m^{(r)}|\mathbf{z}_m^{(r)})\right\}\right] \tag{1.109}$$

により q^* を得る。

パラメータ $\boldsymbol{\pi}$, $\mathbf{A}$, ϕ がそれぞれ独立、かつ、$p(x_m^{(r)}|\mathbf{z}_m^{(r)},\theta)$ は $\boldsymbol{\pi}$, $\mathbf{A}$ に依存しないとすると、まず $\boldsymbol{\pi}$ に関して、

$$\ln q^*(\boldsymbol{\pi}) = \ln p(\boldsymbol{\pi}) + \mathbb{E}_{\mathbf{Z}}\left[\sum_{r=1}^{R} \ln p(\mathbf{z}_1^{(r)})\right] + \text{const.} \tag{1.110}$$

$$= \ln p(\boldsymbol{\pi}) + \mathbb{E}_{\mathbf{Z}}\left[\sum_{r=1}^{R}\sum_{i}^{K} z_{1i}^{(r)} \ln \pi_i\right] + \text{const.} \tag{1.111}$$

$$= \ln p(\boldsymbol{\pi}) + \sum_{i}^{K}\left(\sum_{r}^{R} \overline{z_{1i}^{(r)}}\right) \ln \pi_i + \text{const.} \tag{1.112}$$

となり、

$$q^*(\boldsymbol{\pi}) \propto p(\boldsymbol{\pi}) \times \prod_{i}^{K} \pi_i^{\overline{z_{1i}^R}} \tag{1.113}$$

ただし、$\overline{z_{1i}^R} = \sum_r^R \overline{z_{1i}^{(r)}}$。ここで、事前分布として Dirichlet 分布 $p(\boldsymbol{\pi}) \propto \prod_i^K \pi_i^{(u_i^\pi - 1)}$ を与えて規格化して、

$$q^*(\boldsymbol{\pi}) = \frac{\Gamma\left(u_0^\pi + R\right)}{\prod_{i}^{K} \Gamma\left(u_i^\pi + \overline{z_{1i}^R}\right)} \prod_{i}^{K} \pi_i^{\left(u_i^\pi + \overline{z_{1i}^R} - 1\right)} \tag{1.114}$$

ただし、u_i^π は Dirichlet 分布のハイパーパラメータ。特に理由が無ければ、$u_i^\pi = 1$ としておけばよい。

Dirichlet 分布の性質から、期待値

$$\overline{\pi_i} = \frac{u_i^\pi + \overline{z_{1i}^R}}{u_0^\pi + R} \tag{1.115}$$

$$\overline{\ln \pi_i} = \psi\left(u_i^\pi + \overline{z_{1i}^R}\right) - \psi\left(u_0^\pi + R\right) \tag{1.116}$$

を得る。

次に $\mathbf{A}$ に関して、

$$\ln q^*(\mathbf{A}) = \ln p(\mathbf{A}) + \mathbb{E}_{\mathbf{Z}}\left[\sum_{r=1}^{R}\sum_{n=2}^{N^{(r)}} \ln p(\mathbf{z}_n^{(r)}|\mathbf{z}_{n-1}^{(r)})\right] + \text{const.} \tag{1.117}$$

$$= \ln p(\mathbf{A}) + \mathbb{E}_{\mathbf{Z}}\left[\sum_{r=1}^{R}\sum_{n=2}^{N^{(r)}}\sum_{i}^{K}\sum_{j}^{K} z_{n-1,i}^{(r)} z_{nj}^{(r)} \ln A_{ij}\right] + \text{const.} \tag{1.118}$$

$$= \ln p(\mathbf{A}) + \sum_{r=1}^{R}\sum_{n=2}^{N^{(r)}}\sum_{i}^{K}\sum_{j}^{K} \overline{z_{n-1,i}^{(r)} z_{nj}^{(r)}} \cdot \ln A_{ij} + \text{const.} \tag{1.119}$$

となり、

$$q^*(\mathbf{A}) \propto p(\mathbf{A}) \times \prod_{i}^{K}\prod_{j}^{K} {A_{ij}}^{N_{ij}^{R}} \tag{1.120}$$

となる。ただし、$N_{ij}^R = \sum_{r=1}^{R}\sum_{n=2}^{N^{(r)}} \overline{z_{n-1,i}^{(r)} z_{nj}^{(r)}}$。ここで、事前分布として各 i での j に関する Dirichlet 分布 $p(\mathbf{A}_i) \propto \prod_j^K {A_{ij}}^{(u_{ij}^A - 1)}$ を与えて規格化して、

$$q^*(\mathbf{A}_i) = \frac{\Gamma\left(u_{i0}^A + M_i\right)}{\prod_j^K \Gamma\left(u_{ij}^A + N_{ij}^R\right)} \prod_j^K {A_{ij}}^{\left(u_{ij}^A + N_{ij}^R - 1\right)} \tag{1.121}$$

を得る。ただし、u_{ij}^A は Dirichlet 分布のハイパーパラメータ、$u_{i0}^A = \sum_j^K u_{ij}^A$, $M_i^R = \sum_j^K N_{ij}^R = \sum_{r=1}^{R}\sum_{n=2}^{N^{(r)}}\sum_j^K \overline{z_{n-1,i}^{(r)} z_{nj}^{(r)}}$。

Dirichlet 分布の性質から、期待値

$$\overline{A_{ij}} = \frac{u_{ij}^A + N_{ij}^R}{u_{i0}^A + M_i^R} \tag{1.122}$$

$$\overline{\ln A_{ij}} = \psi\left(u_{ij}^A + N_{ij}^R\right) - \psi\left(u_{i0}^A + M_i^R\right) \tag{1.123}$$

を得る。

ϕ に関しては、$p(x_m^{(r)}|\mathbf{z}_m^{(r)})$ の関数形に依存するが、r に関して独立な場合には、一般化できる。

$$\ln q^*(\phi) = \ln p(\phi) + \mathbb{E}_{\mathbf{Z}}\left[\sum_{r=1}^{R}\sum_{m=1}^{N^{(r)}} \ln p(x_m^{(r)}|\mathbf{z}_m^{(r)}, \phi)\right] + \text{const.} \tag{1.124}$$

$$= \ln p(\phi) + \sum_{r=1}^{R} \mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^{N^{(r)}} \ln p(x_m^{(r)}|\mathbf{z}_m^{(r)}, \phi)\right] + \text{const.} \tag{1.125}$$

各 r について、単独の HMM と同様に $\ln q^*(\phi)$ を求めて (事前分布は除いて) 和をとればよい (?)。

1.4.3 変分下限 $\mathcal{L}(q)$ の算出

式 (1.74) を既知のパラメータについて書き下すと、

$$\begin{aligned}\mathcal{L}(q) =&\mathbb{E}\big[\ln p(\boldsymbol{\pi})\big] + \mathbb{E}\big[\ln p(\mathbf{A})\big] + \mathbb{E}\big[\ln p(\phi)\big] + \mathbb{E}\big[\ln p(\mathbf{X},\mathbf{Z}|\phi)\big] \\ &- \mathbb{E}\big[\ln q(\boldsymbol{\pi})\big] - \mathbb{E}\big[\ln q(\mathbf{A})\big] - \mathbb{E}\big[\ln q(\phi)\big] - \mathbb{E}\big[\ln q(\mathbf{Z})\big] \qquad (1.126)\\ =&\mathbb{E}\big[\ln p(\boldsymbol{\pi})\big] + \mathbb{E}\big[\ln p(\mathbf{A})\big] + \mathbb{E}\big[\ln p(\phi)\big] + \mathbb{E}\left[\sum_{r=1}^{R} \ln p(\mathbf{X}^{(r)},\mathbf{Z}^{(r)}|\phi)\right] \\ &- \mathbb{E}\big[\ln q(\boldsymbol{\pi})\big] - \mathbb{E}\big[\ln q(\mathbf{A})\big] - \mathbb{E}\big[\ln q(\phi)\big] - \mathbb{E}\left[\sum_{r=1}^{R} \ln q(\mathbf{Z}^{(r)})\right] \\ &\qquad (1.127)\\ =&\mathbb{E}\big[\ln p(\boldsymbol{\pi})\big] + \mathbb{E}\big[\ln p(\mathbf{A})\big] + \mathbb{E}\big[\ln p(\phi)\big] \\ &- \mathbb{E}\big[\ln q(\boldsymbol{\pi})\big] - \mathbb{E}\big[\ln q(\mathbf{A})\big] - \mathbb{E}\big[\ln q(\phi)\big] + \sum_{r=1}^{R}\sum_{n=1}^{N} \ln \tilde{c}_n^{(r)} \\ &\qquad (1.128)\end{aligned}$$

となるので、その他のパラメータに関して必要な項をそれぞれ計算する。

$\mathbb{E}\big[\ln q(\boldsymbol{\pi})\big]$ **:** $q(\boldsymbol{\pi}) = \mathrm{Dir}(\pi_i | u_i^{\pi} + \overline{z_{1i}^R})$ なので、Dirichlet 分布の性質から、

$$\begin{aligned}\mathbb{E}\big[\ln q(\boldsymbol{\pi})\big] &= \mathbb{E}_{\pi}\big[\ln q(\boldsymbol{\pi})\big] \qquad (1.129)\\ &= \sum_{i}^{K}\left[(u_i^{\pi} + \overline{z_{1i}^R} - 1)\left\{\psi(u_i^{\pi} + \overline{z_{1i}^R}) - \psi\left(u_0^{\pi} + \sum_{i}^{K}\overline{z_{1i}^R}\right)\right\}\right] + \ln\left\{\frac{\Gamma(u_0^{\pi} + \sum_i^K \overline{z_{1i}^R})}{\prod_i^K \Gamma(u_i^{\pi} + \overline{z_{1i}^R})}\right\} \\ &\qquad (1.130)\\ &= \ln\Gamma(u_0^{\pi} + R) + \sum_{i}^{K}\Big[(u_i^{\pi} + \overline{z_{1i}^R} - 1)\big\{\psi(u_i^{\pi} + \overline{z_{1i}^R}) - \psi(u_0^{\pi} + R)\big\} - \ln\Gamma(u_i^{\pi} + \overline{z_{1i}^R})\Big] \\ &\qquad (1.131)\end{aligned}$$

$\mathbb{E}\bigl[\ln q(\mathbf{A})\bigr]$ **:** $q(\mathbf{A}_i) = \mathrm{Dir}(A_{ij}|u^A_{ij} + N^R_{ij})$ なので、Dirichlet 分布の性質から、

$$
\begin{aligned}
\mathbb{E}\bigl[\ln q(\mathbf{A})\bigr] &= \mathbb{E}_A\bigl[\ln q(\mathbf{A})\bigr] && (1.132)\\
&= \sum_i^K \mathbb{E}_{A_i}\bigl[\ln q(\mathbf{A}_i)\bigr] && (1.133)\\
&= \sum_i^K \Biggl[\sum_j^K \Biggl\{(u^A_{ij} + N^R_{ij} - 1)\Bigl[\psi(u^A_{ij} + N^R_{ij}) - \psi(u^A_{i0} + M^R_i)\Bigr]\Biggr\}\\
&\quad + \ln\Biggl\{\frac{\Gamma(u^A_{i0} + M^R_i)}{\prod_j^K \Gamma\bigl(u^A_{ij} + N^R_{ij}\bigr)}\Biggr\}\Biggr] && (1.134)\\
&= \sum_i^K \Biggl[\ln\Gamma(u^A_{i0} + M^R_i) + \sum_j^K \Bigl\{\bigl(u^A_{ij} + N^R_{ij} - 1\bigr)\Bigl[\psi\bigl(u^A_{ij} + N^R_{ij}\bigr) - \psi\bigl(u^A_{i0} + M^R_i\bigr)\Bigr]\\
&\quad - \ln\Gamma\bigl(u^A_{ij} + N^R_{ij}\bigr)\Bigr\}\Biggr] && (1.135)
\end{aligned}
$$

Chapter 2

Gaussian Time Series

2.1 汎用 Gauss 分布

2.1.1 Gaussian 信号

m 番目の時間ビンでの信号を x_m とする。状態 i の平均を μ_i、精度 $\lambda_i(\equiv {\sigma_i}^{-2})$ とすると、$p(x_m|\mathbf{z}_m)$ は Gauss 分布

$$\begin{aligned} p(x_m|\mathbf{z}_m) =& \prod_{i=1}^{K}\left\{\frac{1}{\sqrt{2\pi{\sigma_i}^2}}\exp\left(-\frac{(x_m-\mu_i)^2}{2{\sigma_i}^2}\right)\right\}^{z_{mi}} \\ =& \prod_{i=1}^{K}\left\{\sqrt{\frac{\lambda_i}{2\pi}}\exp\left(-\frac{\lambda_i}{2}(x_m-\mu_i)^2\right)\right\}^{z_{mi}} \end{aligned} \tag{2.1}$$

で与えられる。

状態遷移確率を、遷移確率行列 $\mathbf{A}$ で与えることにすると、$\mathbf{X}, \mathbf{Z}, \theta$ の同時分布は、

$$\begin{aligned} p(\mathbf{X}, \mathbf{Z}, \theta) =& p(\theta) \cdot \prod_{i}^{K} \pi^{z_{1i}} \cdot \prod_{n=2}^{N}\prod_{i}^{K}\prod_{j}^{K} {A_{ij}}^{z_{n-1,i}z_{nj}} \\ & \times \prod_{m=1}^{N}\prod_{i=1}^{K}\left\{\sqrt{\frac{\lambda_i}{2\pi}}\exp\left(-\frac{\lambda_i}{2}(x_m-\mu_i)^2\right)\right\}^{z_{mi}} \end{aligned} \tag{2.2}$$

となる。

2.1.2 E-step :

$$\ln p(x_m|\mathbf{z}_m, \mu_i, \lambda_i) = \ln\left\{\sqrt{\frac{\lambda_i}{2\pi}}\exp\left(-\frac{\lambda_i}{2}(x_m-\mu_i)^2\right)\right\} \tag{2.3}$$

$$= \frac{1}{2}\left\{\ln\lambda_i - \ln 2\pi - \lambda_i (x_m-\mu_i)^2\right\} \tag{2.4}$$

$$\overline{\ln p(x_m|\mathbf{z}_m, \mu_i, \lambda_i)} = \frac{1}{2}\left(\overline{\ln\lambda_i} - \ln 2\pi - \overline{\lambda_i (x_m-\mu_i)^2}\right) \tag{2.5}$$

式 (1.55) に式 (2.1) を適用し、

$$\tilde{p}(x_m|\mathbf{z}_m) = \prod_i^K\left\{\frac{1}{\sqrt{2\pi}}\exp\left(\frac{1}{2}\left[\overline{\ln\lambda_i} - \overline{\lambda_i (x_m-\mu_i)^2}\right]\right)\right\}^{z_{mi}} \tag{2.6}$$

後で得られる μ_{0i}, β_i^μ, a_i^λ, b_i^λ を使って

$$= \prod_i^K\left\{\frac{1}{\sqrt{2\pi}}\exp\left(\frac{1}{2}\left[\overline{\ln\lambda_i} - \frac{1}{\beta_i^\mu} - \frac{a_i^\lambda}{b_i^\lambda}(x_m-\mu_{0i})^2\right]\right)\right\}^{z_{mi}} \tag{2.7}$$

を得る。

これを、式 (1.53), (1.54) とともに用いて forward-backward アルゴリズムを実行し、γ, ξ 分布を得る。

2.1.3 M-step :

$\boldsymbol{\pi}$, $\mathbf{A}$ に関しては、1.3 節と同様に式 (1.64–1.65, 1.71–1.72) を用いて期待値を求める。

$\boldsymbol{\mu}, \boldsymbol{\lambda}$ に関しては独立でないので、各 i に関しては独立として、Gauss-Gamma 事前分布

$$p(\mu_i, \lambda_i) = \mathcal{N}\left(\mu_i|u_i^\mu, (u_i^\beta\lambda_i)^{-1}\right)\mathrm{Gam}(\lambda_i|u_i^a, u_i^b) \tag{2.8}$$

を与えて (ただし、u_i^μ, u_i^β, u_i^a, u_i^b はハイパーパラメータ)、

$$\ln q^*(\mu_i, \lambda_i) = \ln\mathcal{N}\left(\mu_i|u_i^\mu, (u_i^\beta\lambda_i)^{-1}\right) + \ln\mathrm{Gam}(\lambda_i|u_i^a, u_i^b) + \mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}\cdot\ln\mathcal{N}(x_m|\mu_i, {\lambda_i}^{-1})\right] + \mathrm{const.} \tag{2.9}$$

μ_i, λ_i に依存する項だけ取り出すと、

$$
\begin{aligned}
&= \frac{1}{2}\ln\lambda_i - \frac{u_i^\beta\lambda_i}{2}(\mu_i - u_i^\mu)^2 + (u_i^a - 1)\ln\lambda_i - u_i^b\lambda_i \\
&\quad + \mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}\cdot\frac{1}{2}\left\{\ln\lambda_i - \lambda_i\left(x_m - \mu_i\right)^2\right\}\right] + \text{const.}
\end{aligned} \tag{2.10}
$$

$\ln q^*(\mu_i, \lambda_i)$ は $\ln q^*(\mu_i|\lambda_i) + \ln q^*(\lambda_i)$ と書けるので、μ_i に依存する項だけ考えると、

$$
\begin{aligned}
\ln q^*(\mu_i|\lambda_i) =& -\frac{u_i^\beta\lambda_i}{2}({\mu_i}^2 - 2u_i^\mu\mu_i) \\
& - \mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}\cdot\frac{\lambda_i}{2}\left({\mu_i}^2 - 2x_m\mu_i\right)\right] + \text{const.} \qquad (2.11) \\
=& -\frac{\lambda_i}{2}\left(u_i^\beta + \mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}\right]\right){\mu_i}^2 \\
& + \lambda_i\left(u_i^\beta u_i^\mu + \mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}x_m\right]\right)\mu_i + \text{const.} \qquad (2.12) \\
=& -\frac{\lambda_i}{2}\left(u_i^\beta + \sum_{m=1}^N \overline{z_{mi}}\right){\mu_i}^2 \\
& + \lambda_i\left(u_i^\beta u_i^\mu + \sum_{m=1}^N \overline{z_{mi}}x_m\right)\mu_i + \text{const.} \qquad (2.13) \\
=& -\frac{\lambda_i}{2}\left(u_i^\beta + N_i\right){\mu_i}^2 + \lambda_i\left(u_i^\beta u_i^\mu + N_i\overline{x}_i\right)\mu_i + \text{const.} \qquad (2.14)
\end{aligned}
$$

ただし、$N_i = \sum_m^N \overline{z_{mi}}$、$\overline{x}_i = \frac{1}{N_i}\sum_m^N \overline{z_{mi}}\cdot x_m$。

$\ln q^*(\mu_i|\lambda_i)$ が μ_i の二乗に依存しているので、$q^*(\mu_i|\lambda_i)$ は Gauss 分布。平方完成から、

$$
\beta_i^\mu = u_i^\beta + N_i \tag{2.15}
$$

$$
\mu_{0i} = \frac{1}{\beta_i^\mu}\left(u_i^\beta u_i^\mu + N_i\overline{x}_i\right) \tag{2.16}
$$

を使って、

$$
\ln q^*(\mu_i|\lambda_i) = -\frac{\beta_i^\mu\lambda_i}{2}{\mu_i}^2 + \beta_i^\mu\lambda_i\mu_{0i}\cdot\mu_i + \text{const.} \tag{2.17}
$$

$$
= -\frac{\beta_i^\mu\lambda_i}{2}\left(\mu_i - \mu_{0i}\right)^2 + \text{const.} \tag{2.18}
$$

$$
\begin{aligned}
q^*(\mu_i|\lambda_i) \propto & \exp\left(-\frac{\beta_i^\mu \lambda_i}{2}\left(\mu_i-\mu_{0i}\right)^2\right) && (2.19)\\
\equiv & \mathcal{N}\left(\mu_i|\mu_{0i},(\beta_i^\mu\lambda_i)^{-1}\right) && (2.20)
\end{aligned}
$$

を得る。

$q^*(\lambda_i)$ に関しては、

$$
\ln q^*(\lambda_i) = \ln q^*(\mu_i,\lambda_i) - \ln q^*(\mu_i|\lambda_i) \tag{2.21}
$$

の関係から、式 (2.10) と式 (2.20) を使って、

$$
\begin{aligned}
\ln q^*(\lambda_i) =& \frac{1}{2}\ln\lambda_i - \frac{u_i^\beta \lambda_i}{2}(\mu_i-u_i^\mu)^2 + (u_i^a-1)\ln\lambda_i - u_i^b\lambda_i \\
&+ \mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}\cdot\frac{1}{2}\left\{\ln\lambda_i - \lambda_i\left(x_m-\mu_i\right)^2\right\}\right] \\
&- \left\{\frac{1}{2}\ln\lambda_i - \frac{\beta_i^\mu\lambda_i}{2}\left(\mu_i-\mu_{0i}\right)^2\right\} + \text{const.} && (2.22)
\end{aligned}
$$

λ_i に依存する項だけ取り出して、

$$
\begin{aligned}
\ln q^*(\lambda_i) =& \left(u_i^a - 1 + \frac{1}{2}\mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}\right]\right)\ln\lambda_i - \left\{\frac{u_i^\beta}{2}(\mu_i-u_i^\mu)^2 + u_i^b\right. \\
&\left.+\frac{1}{2}\mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}\left(x_m-\mu_i\right)^2\right] - \frac{\beta_i^\mu}{2}\left(\mu_i-\mu_{0i}\right)^2\right\}\lambda_i + \text{const.} && (2.23)\\
=& (a_i^\lambda-1)\ln\lambda_i - b_i^\lambda\lambda_i + \text{const.} && (2.24)
\end{aligned}
$$

ただし、

$$
\begin{aligned}
a_i^\lambda =& u_i^a + \frac{1}{2}\mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}\right] + \frac{1}{2} && (2.25)\\
=& u_i^a + \frac{N_i}{2} && (2.26)\\
b_i^\lambda =& u_i^b + \frac{1}{2}\mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^N z_{mi}\left(x_m-\mu_i\right)^2\right] + \frac{u_i^\beta}{2}\left(\mu_i-u_i^\mu\right)^2 \\
&- \frac{u_i^\beta+N_i}{2}\left(\mu_i - \frac{u_i^\beta u_i^\mu + N_i\overline{x}_i}{u_i^\beta+N_i}\right)^2 && (2.27)
\end{aligned}
$$

$$
\begin{aligned}
=& u_i^b + \frac{1}{2}\sum_{m=1}^{N} \overline{z_{mi}} \left({x_m}^2 - 2\mu_i x_m + {\mu_i}^2 \right) + \frac{u_i^\beta}{2} \left({\mu_i}^2 - 2u_i^\mu \mu_i + u_i^{\mu 2} \right) \\
&- \frac{u_i^\beta + N_i}{2} \left\{ {\mu_i}^2 - 2\frac{u_i^\beta u_i^\mu + N_i \overline{x}_i}{u_i^\beta + N_i}\mu_i + \left(\frac{u_i^\beta u_i^\mu + N_i \overline{x}_i}{u_i^\beta + N_i} \right)^2 \right\}
\end{aligned} \tag{2.28}
$$

$$
\begin{aligned}
=& u_i^b + \frac{1}{2}\sum_{m=1}^{N} \overline{z_{mi}} \left({x_m}^2 - 2\overline{x}_i x_m + {\overline{x}_i}^2 \right) \\
&- \frac{1}{2}\sum_{m=1}^{N} \overline{z_{mi}} \left({\overline{x}_i}^2 - 2\overline{x}_i x_m + 2\mu_i x_m - {\mu_i}^2 \right) - \frac{N_i}{2}{\mu_i}^2 \\
&- \left\{ u_i^\beta u_i^\mu - \left(u_i^\beta u_i^\mu + N_i \overline{x}_i \right) \right\} \mu_i + \frac{u_i^\beta u_i^{\mu 2}}{2} \\
&- \frac{u_i^\beta + N_i}{2} \left(\frac{u_i^\beta u_i^\mu + N_i \overline{x}_i}{u_i^\beta + N_i} \right)^2
\end{aligned} \tag{2.29}
$$

$$
\begin{aligned}
=& u_i^b + \frac{N_i}{2}\mathbf{S}_i - \frac{1}{2} \left(N_i {\overline{x}_i}^2 - 2N_i {\overline{x}_i}^2 + 2N_i \mu_i \overline{x}_i - N_i {\mu_i}^2 \right) - \frac{N_i}{2}{\mu_i}^2 \\
&+ N_i \overline{x}_i \mu_i + \frac{u_i^\beta u_i^{\mu 2}}{2} - \frac{\left(u_i^\beta u_i^\mu \right)^2 + 2u_i^\beta u_i^\mu N_i \overline{x}_i + \left(N_i \overline{x}_i \right)^2}{2\left(u_i^\beta + N_i \right)}
\end{aligned} \tag{2.30}
$$

$$
= u_i^b + \frac{N_i}{2}\mathbf{S}_i + \frac{N_i}{2}{\overline{x}_i}^2 + \frac{u_i^\beta u_i^{\mu 2}}{2} - \frac{\left(u_i^\beta u_i^\mu \right)^2 + 2u_i^\beta u_i^\mu N_i \overline{x}_i + \left(N_i \overline{x}_i \right)^2}{2\left(u_i^\beta + N_i \right)} \tag{2.31}
$$

$$
\begin{aligned}
=& u_i^b + \frac{N_i}{2}\mathbf{S}_i + \frac{1}{2\left(u_i^\beta + N_i \right)} \left\{ N_i \left(u_i^\beta + N_i \right) {\overline{x}_i}^2 + \left(u_i^\beta + N_i \right) u_i^\beta u_i^{\mu 2} \right. \\
&\left. - \left(u_i^\beta u_i^\mu \right)^2 - 2u_i^\beta u_i^\mu N_i \overline{x}_i - \left(N_i \overline{x}_i \right)^2 \right\}
\end{aligned} \tag{2.32}
$$

$$
= u_i^b + \frac{N_i}{2}\mathbf{S}_i + \frac{N_i u_i^\beta {\overline{x}_i}^2 - 2u_i^\beta u_i^\mu N_i \overline{x}_i + N_i u_i^\beta u_i^{\mu 2}}{2\left(u_i^\beta + N_i \right)} \tag{2.33}
$$

$$
= u_i^b + \frac{N_i}{2}\mathbf{S}_i + \frac{u_i^\beta N_i}{2\left(u_i^\beta + N_i \right)} \left({\overline{x}_i}^2 - 2u_i^\mu \overline{x}_i + u_i^{\mu 2} \right) \tag{2.34}
$$

$$
= u_i^b + \frac{N_i}{2}\mathbf{S}_i + \frac{u_i^\beta N_i}{2\left(u_i^\beta + N_i \right)} \left(\overline{x}_i - u_i^\mu \right)^2 \tag{2.35}
$$

ただし、

$$\mathbf{S}_i = \frac{1}{N_i}\sum_{m=1}^{N}\overline{z_{mi}}(x_m - \overline{x}_i)^2 \tag{2.36}$$

これら μ_{0i}, β_i^μ, a_i^λ, b_i^λ を使って Gauss-Gamma 分布 (**??**) として、

$$q^*(\mu_i, \lambda_i) = \mathcal{N}\left(\mu_i|\mu_{0i}, (\beta_i^\mu\lambda_i)^{-1}\right)\mathrm{Gam}(\lambda_i|a_i^\lambda, b_i^\lambda) \tag{2.37}$$

$$= \sqrt{\frac{\beta_i^\mu\lambda_i}{2\pi}}\exp\left\{-\frac{\beta_i^\mu\lambda_i}{2}(\mu_i - \mu_{0i})^2\right\}\frac{1}{\Gamma(a_i^\lambda)}{b_i^\lambda}^{a_i^\lambda}{\lambda_i}^{a_i^\lambda-1}e^{-b_i^\lambda\lambda_i} \tag{2.38}$$

を得る。計算により、

$$\overline{\mu_i} = \mu_{0i} \tag{2.39}$$

$$\overline{\lambda_i} = \frac{a_i^\lambda}{b_i^\lambda} \tag{2.40}$$

$$\overline{\ln\lambda_i} = \psi(a_i^\lambda) - \ln b_i^\lambda \tag{2.41}$$

$$\overline{\lambda_i(x-\mu_i)^2} = \frac{1}{\beta_i^\mu} + \frac{a_i^\lambda}{b_i^\lambda}(x - \mu_{0i})^2 \tag{2.42}$$

が得られる。

2.1.4 変分下限 $\mathcal{L}(q)$ の算出

変分下限 $\mathcal{L}(q)$ を計算するため、式 (1.75) から、

$$\mathcal{L}(q) = \mathbb{E}\big[\ln p(\boldsymbol{\pi})\big] + \mathbb{E}\big[\ln p(\mathbf{A})\big] + \mathbb{E}\big[\ln p(\boldsymbol{\mu}, \boldsymbol{\lambda})\big]$$

$$- \mathbb{E}\big[\ln q(\boldsymbol{\pi})\big] - \mathbb{E}\big[\ln q(\mathbf{A})\big] - \mathbb{E}\big[\ln q(\boldsymbol{\mu}, \boldsymbol{\lambda})\big] + \prod_n^N \tilde{c}_n \tag{2.43}$$

を得る。

$\mathbb{E}\big[\ln p(\boldsymbol{\pi})\big]$, $\mathbb{E}\big[\ln p(\mathbf{A})\big]$, $\mathbb{E}\big[\ln q(\boldsymbol{\pi})\big]$, $\mathbb{E}\big[\ln q(\mathbf{A})\big]$ に関しては、式 (1.79), (1.83), (1.86), (1.90) で、$\tilde{c}_n$ に関しては E-step で、それぞれ得られているので、以下ではそれ以外の項について計算する。

$\mathbb{E}\big[\ln p(\boldsymbol{\mu}, \boldsymbol{\lambda})\big]$ **:**

$$\mathbb{E}\big[\ln p(\boldsymbol{\mu}, \boldsymbol{\lambda})\big] = \mathbb{E}\left[\ln \prod_i^K \mathcal{N}\left(\mu_i | u_i^\mu, (u_i^\beta \lambda_i)^{-1}\right) \mathrm{Gam}(\lambda_i | u_i^a, u_i^b)\right] \tag{2.44}$$

$$= \mathbb{E}\left[\ln \prod_i^K \sqrt{\frac{u_i^\beta \lambda_i}{2\pi}} \exp\left\{-\frac{u_i^\beta \lambda_i}{2}(\mu_i - u_i^\mu)^2\right\} \times \frac{1}{\Gamma(u_i^a)} u_i^{b\,u_i^a} \lambda_i^{u_i^a - 1} e^{-u_i^b \lambda_i}\right] \tag{2.45}$$

$$= \mathbb{E}\left[\sum_i^K \left\{\frac{1}{2}\ln u_i^\beta + \frac{1}{2}\ln \lambda_i - \frac{1}{2}\ln 2\pi - \frac{u_i^\beta}{2}\lambda_i(\mu_i - u_i^\mu)^2 - \ln \Gamma(u_i^a) + u_i^a \ln u_i^b + (u_i^a - 1)\ln \lambda_i - u_i^b \lambda_i\right\}\right] \tag{2.46}$$

$$= -\frac{K}{2}\ln 2\pi + \sum_i^K \left\{\frac{1}{2}\ln u_i^\beta - \frac{u_i^\beta}{2}\overline{\lambda_i(\mu_i - u_i^\mu)^2} - \ln \Gamma(u_i^a) + u_i^a \ln u_i^b + \left(u_i^a - \frac{1}{2}\right)\overline{\ln \lambda_i} - u_i^b \overline{\lambda_i}\right\} \tag{2.47}$$

$$= -\frac{K}{2}\ln 2\pi + \sum_i^K \left\{\frac{1}{2}\ln u_i^\beta - \frac{u_i^\beta}{2}\left[\frac{1}{\beta_i^\mu} + \frac{a_i^\lambda}{b_i^\lambda}(\mu_{0i} - u_i^\mu)^2\right] - \ln \Gamma(u_i^a) + u_i^a \ln u_i^b + \left(u_i^a - \frac{1}{2}\right)\overline{\ln \lambda_i} - u_i^b \overline{\lambda_i}\right\} \tag{2.48}$$

$\mathbb{E}\big[\ln q(\boldsymbol{\mu}, \boldsymbol{\lambda})\big]$ **:**

$$\mathbb{E}\big[\ln q(\boldsymbol{\mu}, \boldsymbol{\lambda})\big] = \mathbb{E}\left[\ln \prod_i^K \mathcal{N}\left(\mu_i | \mu_{0i}, (\beta_i^\mu \lambda_i)^{-1}\right) \mathrm{Gam}(\lambda_i | a_i^\lambda, b_i^\lambda)\right] \tag{2.49}$$

$$= \mathbb{E}\left[\ln \prod_i^K \sqrt{\frac{\beta_i^\mu \lambda_i}{2\pi}} \exp\left\{-\frac{\beta_i^\mu \lambda_i}{2}(\mu_i - \mu_{0i})^2\right\} \times \frac{1}{\Gamma(a_i^\lambda)} b_i^{\lambda\,a_i^\lambda} \lambda_i^{a_i^\lambda - 1} e^{-b_i^\lambda \lambda_i}\right] \tag{2.50}$$

$$= \mathbb{E}\left[\sum_i^K \left\{\frac{1}{2}\ln \beta_i^\mu + \frac{1}{2}\ln \lambda_i - \frac{1}{2}\ln 2\pi - \frac{\beta_i^\mu}{2}\lambda_i(\mu_i - \mu_{0i})^2 - \ln \Gamma(a_i^\lambda) + a_i^\lambda \ln b_i^\lambda + (a_i^\lambda - 1)\ln \lambda_i - b_i^\lambda \lambda_i\right\}\right] \tag{2.51}$$

$$
\begin{aligned}
=& -\frac{K}{2}\ln 2\pi + \sum_{i}^{K}\Bigg\{\frac{1}{2}\ln\beta_i^{\mu} - \frac{\beta_i^{\mu}}{2}\overline{\lambda_i(\mu_i-\mu_{0i})^2} \\
& -\ln\Gamma(a_i^{\lambda}) + a_i^{\lambda}\ln b_i^{\lambda} + \left(a_i^{\lambda}-\frac{1}{2}\right)\overline{\ln\lambda_i} - b_i^{\lambda}\overline{\lambda_i}\Bigg\} \qquad (2.52)\\
=& -\frac{K}{2}\ln 2\pi + \sum_{i}^{K}\Bigg\{\frac{1}{2}\ln\beta_i^{\mu} - \frac{\beta_i^{\mu}}{2}\left[\frac{1}{\beta_i^{\mu}} + \frac{a_i^{\lambda}}{b_i^{\lambda}}(\mu_{0i}-\mu_{0i})^2\right] \\
& -\ln\Gamma(a_i^{\lambda}) + a_i^{\lambda}\ln b_i^{\lambda} + \left(a_i^{\lambda}-\frac{1}{2}\right)\overline{\ln\lambda_i} - b_i^{\lambda}\overline{\lambda_i}\Bigg\} \qquad (2.53)\\
=& -\frac{K}{2}\ln 2\pi - \frac{K}{2} + \sum_{i}^{K}\Bigg\{\frac{1}{2}\ln\beta_i^{\mu} - \ln\Gamma(a_i^{\lambda}) + a_i^{\lambda}\ln b_i^{\lambda} \\
& + \left(a_i^{\lambda}-\frac{1}{2}\right)\overline{\ln\lambda_i} - b_i^{\lambda}\overline{\lambda_i}\Bigg\} \qquad (2.54)\\
=& -\frac{K}{2}\ln 2\pi - \frac{K}{2} + \sum_{i}^{K}\Bigg\{\frac{1}{2}\ln\beta_i^{\mu} - \ln\Gamma(a_i^{\lambda}) + a_i^{\lambda}\ln b_i^{\lambda} \\
& + \left(a_i^{\lambda}-\frac{1}{2}\right)\overline{\ln\lambda_i} - a_i^{\lambda}\Bigg\} \qquad (2.55)
\end{aligned}
$$

2.1.5 global analysis 対応

M-step と lower bounds の計算で、N_i, M_i, N_{ij}, $\overline{x}_i$, $\mathbf{S}_i$ をそれぞれ

$$N_i^R = \sum_{r=1}^{R}\sum_{m}^{N}\overline{z_{mi}^R} \qquad (2.56)$$

$$M_i^R = \sum_{j}^{K} N_{ij}^R = \sum_{r=1}^{R}\sum_{n=2}^{N^{(r)}}\sum_{j}^{K}\overline{z_{n-1,i}^{(r)}z_{nj}^{(r)}} \qquad (2.57)$$

$$\overline{x}_i^R = \frac{1}{N_i^R}\sum_{r=1}^{R}\sum_{m}^{N}\overline{z_{mi}^R}\cdot x_m^R \qquad (2.58)$$

$$\mathbf{S}_i^R = \frac{1}{N_i^R}\sum_{r=1}^{R}\sum_{m=1}^{N}\overline{z_{mi}^R}(x_m^R - \overline{x}_i^R)^2 \qquad (2.59)$$

に置き換えて計算すればよい。

2.2 Gaussian １分子蛍光強度信号

蛍光１分子の蛍光強度が平均 μ、分散 σ^2 (精度は $\lambda \equiv \sigma^{-2}$) の Gauss 分布

$$p(x_m|\mu,\lambda) = \mathcal{N}(x_m|\mu,\lambda) \tag{2.60}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_m-\mu)^2}{2\sigma^2}\right) \tag{2.61}$$

$$= \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda}{2}(x_m-\mu)^2\right) \tag{2.62}$$

で得られるとする。このとき、i 量体 (本節では虚数単位の i は用いない) の強度は

$$p(x_m|\mu,\lambda) = \mathcal{N}(x_m|i\mu,\lambda/i) \tag{2.63}$$

$$= \sqrt{\frac{\lambda}{2\pi i}} \exp\left(-\frac{\lambda}{2i}(x_m-i\mu)^2\right) \tag{2.64}$$

になるので、emission probability は

$$p(x_m|\mathbf{z}_m,\mu,\lambda) = \prod_{i=1}^{K}\left\{\sqrt{\frac{\lambda}{2\pi i}} \exp\left(-\frac{\lambda}{2i}(x_m-i\mu)^2\right)\right\}^{z_{mi}} \tag{2.65}$$

で与えられる。

状態遷移確率を遷移確率行列 $\mathbf{A}$ で与えることにすると、$\mathbf{X},\mathbf{Z},\theta$ の同時分布は、

$$p(\mathbf{X},\mathbf{Z},\theta) = p(\theta) \times \prod_{i}^{K} \pi^{z_{1i}} \times \prod_{n=2}^{N}\prod_{i}^{K}\prod_{j}^{K} {A_{ij}}^{z_{n-1,i}z_{nj}}$$

$$\times \prod_{m=1}^{N}\prod_{i=1}^{K}\left\{\sqrt{\frac{\lambda}{2\pi i}} \exp\left(-\frac{\lambda}{2i}(x_m-i\mu)^2\right)\right\}^{z_{mi}} \tag{2.66}$$

となる。

2.2.1 E-step：

$$\ln p(x_m|\mathbf{z}_m,\mu,\lambda) = \sum_{i=1}^{K} \ln\left\{\sqrt{\frac{\lambda}{2\pi i}} \exp\left(-\frac{\lambda}{2i}(x_m-i\mu)^2\right)\right\}^{z_{mi}} \tag{2.67}$$

$$= \sum_{i=1}^{K} \frac{z_{mi}}{2}\left\{\ln\lambda - \ln 2\pi i - \frac{\lambda}{i}(x_m-i\mu)^2\right\} \tag{2.68}$$

$$\overline{\ln p(x_m|\mathbf{z}_m,\mu,\lambda)} = \sum_{i=1}^{K} \frac{z_{mi}}{2}\left(\overline{\ln\lambda} - \ln 2\pi i - \frac{1}{i}\overline{\lambda(x_m-i\mu)^2}\right) \tag{2.69}$$

式 (1.55) に式 (2.65) を適用し、

$$\tilde{p}(x_m|\mathbf{z}_m, \mu, \lambda) = \prod_i^K \left\{ \frac{1}{\sqrt{2\pi i}} \exp\left(\frac{1}{2} \left[\overline{\ln \lambda} - \frac{1}{i} \overline{\lambda \left(x_m - i\mu \right)^2} \right] \right) \right\}^{z_{mi}} \tag{2.70}$$

後で得られる μ_0, β^μ, a^λ, b^λ を使って

$$= \prod_i^K \left\{ \frac{1}{\sqrt{2\pi i}} \exp\left(\frac{1}{2} \left[\overline{\ln \lambda} - \frac{i}{\beta^\mu} - \frac{a^\lambda}{b^\lambda} \frac{\left(x_m - i\mu_0 \right)^2}{i} \right] \right) \right\}^{z_{mi}} \tag{2.71}$$

を得る。

これを、式 (1.53), (1.54) とともに用いて forward-backward アルゴリズムを実行し、γ, ξ 分布を得る。

2.2.2 M-step：

$\boldsymbol{\pi}$, $\mathbf{A}$ に関しては、1.3 節と同様に式 (1.64–1.65, 1.71–1.72) を用いて期待値を求める。

μ, λ に関しては独立でなく、各 i に関しても独立ではないことに注意。Gauss-Gamma 事前分布

$$p(\mu, \lambda) = \mathcal{N}\left(\mu | u^\mu, (u^\beta \lambda)^{-1} \right) \mathrm{Gam}(\lambda | u^a, u^b) \tag{2.72}$$

を与えて (ただし、u^μ, u^β, u^a, u^b はハイパーパラメータ)、

$$\begin{aligned} \ln q^*(\mu, \lambda) &= \ln \mathcal{N}\left(\mu | u^\mu, (u^\beta \lambda)^{-1} \right) + \ln \mathrm{Gam}(\lambda | u^a, u^b) \\ &\quad + \mathbb{E}_{\mathbf{Z}} \left[\sum_{i=1}^K \sum_{m=1}^N z_{mi} \cdot \ln \mathcal{N}(x_m | i\mu, i/\lambda) \right] + \text{const.} \end{aligned} \tag{2.73}$$

μ, λ に依存する項だけ取り出すと、

$$\begin{aligned} =& \frac{1}{2} \ln \lambda - \frac{u^\beta \lambda}{2} (\mu - u^\mu)^2 + (u^a - 1) \ln \lambda - u^b \lambda \\ &+ \mathbb{E}_{\mathbf{Z}} \left[\sum_{i=1}^K \sum_{m=1}^N z_{mi} \cdot \frac{1}{2} \left\{ \ln \lambda - \frac{\lambda}{i} \left(x_m - i\mu \right)^2 \right\} \right] + \text{const.} \end{aligned} \tag{2.74}$$

$\ln q^*(\mu, \lambda)$ は $\ln q^*(\mu|\lambda) + \ln q^*(\lambda)$ と書けるので、μ に依存する項だけ考えると、

$$
\begin{aligned}
\ln q^*(\mu|\lambda) =& -\frac{u^\beta \lambda}{2}(\mu^2 - 2u^\mu \mu) \\
& - \mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^{K}\sum_{m=1}^{N} z_{mi} \cdot \frac{\lambda}{2i}\left\{(i\mu)^2 - 2ix_m\mu\right\}\right] + \text{const.} \quad (2.75)\\
=& -\frac{\lambda}{2}\left(u^\beta + \mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^{K}\sum_{m=1}^{N} iz_{mi}\right]\right)\mu^2 \\
& + \lambda\left(u^\beta u^\mu + \mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^{K}\sum_{m=1}^{N} z_{mi}x_m\right]\right)\mu + \text{const.} \quad (2.76)\\
=& -\frac{\lambda}{2}\left(u^\beta + \sum_{i=1}^{K} i \sum_{m=1}^{N} \overline{z_{mi}}\right)\mu^2 \\
& + \lambda\left(u^\beta u^\mu + \sum_{i=1}^{K}\sum_{m=1}^{N} \overline{z_{mi}}x_m\right)\mu + \text{const.} \quad (2.77)\\
=& -\frac{\lambda}{2}\left(u^\beta + N_0^i\right)\mu^2 + \lambda\left(u^\beta u^\mu + N_x\right)\mu + \text{const.} \quad (2.78)
\end{aligned}
$$

ただし、

$$N_i = \sum_{m=1}^{N} \overline{z_{mi}} \quad (2.79)$$

$$N_0^i = \sum_{i=1}^{K} iN_i \quad (2.80)$$

$$N_x = \sum_{i=1}^{K}\sum_{m=1}^{N} \overline{z_{mi}}x_m \quad (2.81)$$

$\ln q^*(\mu|\lambda)$ が μ_i の二乗に依存しているので、$q^*(\mu|\lambda)$ は Gauss 分布。平方完成から、

$$\beta^\mu = u^\beta + N_0^i \quad (2.82)$$

$$\mu_0 = \frac{1}{\beta^\mu}\left(u^\beta u^\mu + N_x\right) \quad (2.83)$$

を使って、

$$q^*(\mu|\lambda) = \mathcal{N}\left(\mu|\mu_0, (\beta^\mu\lambda)^{-1}\right) \quad (2.84)$$

を得る。

$q^*(\lambda)$ に関しては、

$$\ln q^*(\lambda) = \ln q^*(\mu, \lambda) - \ln q^*(\mu|\lambda) \tag{2.85}$$

の関係から、式 (2.74) と式 (2.20) を使って、λ に依存する項だけを抜き出して、

$$\begin{aligned}
\ln q^*(\lambda) =& \frac{1}{2}\ln\lambda - \frac{u^\beta\lambda}{2}(\mu - u^\mu)^2 + (u^a - 1)\ln\lambda - u^b\lambda \\
&+ \mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^{K}\sum_{m=1}^{N} z_{mi}\cdot\frac{1}{2}\left\{\ln\lambda - \frac{\lambda}{i}(x_m - i\mu)^2\right\}\right] \\
&- \left\{\frac{1}{2}\ln\lambda - \frac{\beta^\mu\lambda}{2}(\mu - \mu_0)^2\right\} + \text{const.} &(2.86) \\
=& \left(u^a - 1 + \frac{1}{2}\mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^{K}\sum_{m=1}^{N} z_{mi}\right]\right)\ln\lambda \\
&- \left\{\frac{u^\beta}{2}(\mu - u^\mu)^2 + u^b + \frac{1}{2}\mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^{K}\sum_{m=1}^{N} z_{mi}\frac{(x_m - i\mu)^2}{i}\right]\right. \\
&\left. - \frac{\beta^\mu}{2}(\mu - \mu_0)^2\right\}\lambda + \text{const.} &(2.87) \\
=& (a^\lambda - 1)\ln\lambda - b^\lambda\lambda + \text{const.} &(2.88)
\end{aligned}$$

ただし、

$$\begin{aligned}
a^\lambda =& u^a + \frac{1}{2}\mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^{K}\sum_{m=1}^{N} z_{mi}\right] &(2.89) \\
=& u^a + \frac{N}{2} &(2.90) \\
b^\lambda =& u^b + \frac{1}{2}\mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^{K}\sum_{m=1}^{N} z_{mi}\frac{(x_m - i\mu)^2}{i}\right] + \frac{u^\beta}{2}(\mu - u^\mu)^2 \\
&- \frac{\beta^\mu}{2}(\mu - \mu_0)^2 &(2.91) \\
=& u^b + \frac{1}{2}\left\{\sum_{i=1}^{K}\sum_{m=1}^{N} \overline{z_{mi}}\frac{{x_m}^2 - 2ix_m\mu + (i\mu)^2}{i}\right\} + \frac{u^\beta}{2}(\mu - u^\mu)^2 \\
&- \frac{u^\beta + N_0^i}{2}(\mu - \mu_0)^2 &(2.92) \\
=& u^b + \frac{1}{2}\left\{\sum_{i=1}^{K}\sum_{m=1}^{N} \frac{\overline{z_{mi}}{x_m}^2}{i} - 2\mu\sum_{i=1}^{K}\sum_{m=1}^{N}\overline{z_{mi}}x_m + \mu^2\sum_{i=1}^{K} i\sum_{m=1}^{N}\overline{z_{mi}}\right\} \\
&- \frac{N_0^i}{2}\mu^2 + \left\{(u^\beta + N_0^i)\mu_0 - u^\beta u^\mu\right\}\mu + \frac{u^\beta}{2}u^{\mu 2} - \frac{u^\beta + N_0^i}{2}{\mu_0}^2 &(2.93)
\end{aligned}$$

$$
\begin{aligned}
&= u^b + \frac{1}{2}\left\{\sum_{i=1}^{K}\sum_{m=1}^{N}\frac{\overline{z_{mi}}{x_m}^2}{i} - 2N_x\mu + N_0^i\mu^2\right\} \\
&\quad - \frac{N_0^i}{2}\mu^2 + (\beta^\mu\mu_0 - u^\beta u^\mu)\mu + \frac{u^\beta}{2}u^{\mu 2} - \frac{\beta^\mu}{2}{\mu_0}^2 \qquad (2.94) \\
&= u^b + \frac{1}{2}\sum_{i=1}^{K}\sum_{m=1}^{N}\frac{\overline{z_{mi}}{x_m}^2}{i} + \Big\{(u^\beta u^\mu + N_x) - u^\beta u^\mu - N_x\Big\}\mu \\
&\quad + \frac{u^\beta}{2}u^{\mu 2} - \frac{\beta^\mu}{2}{\mu_0}^2 \qquad (2.95) \\
&= u^b + \frac{1}{2}\sum_{i=1}^{K}\sum_{m=1}^{N}\frac{\overline{z_{mi}}{x_m}^2}{i} + \frac{u^\beta}{2}u^{\mu 2} - \frac{\beta^\mu}{2}{\mu_0}^2 \qquad (2.96)
\end{aligned}
$$

これら μ_0, β^μ, a^λ, b^λ を使って Gauss-Gamma 分布 (**??**) として、

$$
\begin{aligned}
q^*(\mu,\lambda) &= \mathcal{N}\left(\mu|\mu_0,(\beta^\mu\lambda)^{-1}\right)\mathrm{Gam}(\lambda|a^\lambda,b^\lambda) \qquad (2.97) \\
&= \sqrt{\frac{\beta^\mu\lambda}{2\pi}}\exp\left\{-\frac{\beta^\mu\lambda}{2}(\mu-\mu_0)^2\right\}\frac{1}{\Gamma(a^\lambda)}{b^\lambda}^{a^\lambda}\lambda^{a^\lambda-1}e^{-b^\lambda\lambda} \qquad (2.98)
\end{aligned}
$$

を得る。計算により、

$$
\begin{aligned}
\overline{\mu} &= \mu_0 \qquad (2.99) \\
\overline{\lambda} &= \frac{a^\lambda}{b^\lambda} \qquad (2.100) \\
\overline{\ln\lambda} &= \psi(a^\lambda) - \ln b^\lambda \qquad (2.101) \\
\overline{\lambda(x-\mu)^2} &= \frac{1}{\beta^\mu} + \frac{a^\lambda}{b^\lambda}(x-\mu_0)^2 \qquad (2.102) \\
\overline{\lambda(x-i\mu)^2} &= \frac{i^2}{\beta^\mu} + \frac{a^\lambda}{b^\lambda}(x-i\mu_0)^2 \qquad (2.103)
\end{aligned}
$$

が得られる。

2.2.3 変分下限 $\mathcal{L}(q)$ の算出

変分下限 $\mathcal{L}(q)$ を計算するため、式 (1.75) から、

$$
\begin{aligned}
\mathcal{L}(q) &= \mathbb{E}\big[\ln p(\boldsymbol{\pi})\big] + \mathbb{E}\big[\ln p(\mathbf{A})\big] + \mathbb{E}\big[\ln p(\mu,\lambda)\big] \\
&\quad - \mathbb{E}\big[\ln q(\boldsymbol{\pi})\big] - \mathbb{E}\big[\ln q(\mathbf{A})\big] - \mathbb{E}\big[\ln q(\mu,\lambda)\big] + \prod_n^N \tilde{c}_n \qquad (2.104)
\end{aligned}
$$

を得る。

$\mathbb{E}\big[\ln p(\boldsymbol{\pi})\big]$, $\mathbb{E}\big[\ln p(\mathbf{A})\big]$, $\mathbb{E}\big[\ln q(\boldsymbol{\pi})\big]$, $\mathbb{E}\big[\ln q(\mathbf{A})\big]$ に関しては、式 (1.79), (1.83), (1.86), (1.90) で、$\tilde{c}_n$ に関しては E-step で、それぞれ得られているので、以下ではそれ以外の項について計算する。

$\mathbb{E}\big[\ln p(\mu,\lambda)\big]$ **:**

$$\begin{aligned}
\mathbb{E}\big[\ln p(\mu,\lambda)\big] =&\mathbb{E}\left[\ln\mathcal{N}\left(\mu|u^{\mu},(u^{\beta}\lambda)^{-1}\right)\mathrm{Gam}(\lambda|u^{a},u^{b})\right] && (2.105)\\
=&\mathbb{E}\Bigg[\ln\sqrt{\frac{u^{\beta}\lambda}{2\pi}}\exp\left\{-\frac{u^{\beta}\lambda}{2}(\mu-u^{\mu})^2\right\}\\
&\times\frac{1}{\Gamma(u^{a})}u^{b^{u^{a}}}\lambda^{u^{a}-1}e^{-u^{b}\lambda}\Bigg] && (2.106)\\
=&\mathbb{E}\Bigg[\frac{1}{2}\ln u^{\beta}+\frac{1}{2}\ln\lambda-\frac{1}{2}\ln 2\pi-\frac{u^{\beta}}{2}\lambda(\mu-u^{\mu})^2\\
&-\ln\Gamma(u^{a})+u^{a}\ln u^{b}+(u^{a}-1)\ln\lambda-u^{b}\lambda\Bigg] && (2.107)\\
=&\frac{1}{2}\ln u^{\beta}-\frac{1}{2}\ln 2\pi-\frac{u^{\beta}}{2}\overline{\lambda(\mu-u^{\mu})^2}\\
&-\ln\Gamma(u^{a})+u^{a}\ln u^{b}+\left(u^{a}-\frac{1}{2}\right)\overline{\ln\lambda}-u^{b}\overline{\lambda} && (2.108)\\
=&\frac{1}{2}\ln u^{\beta}-\frac{1}{2}\ln 2\pi-\frac{u^{\beta}}{2}\left[\frac{1}{\beta^{\mu}}+\frac{a^{\lambda}}{b^{\lambda}}(\mu_0-u^{\mu})^2\right]\\
&-\ln\Gamma(u^{a})+u^{a}\ln u^{b}+\left(u^{a}-\frac{1}{2}\right)\overline{\ln\lambda}-u^{b}\overline{\lambda} && (2.109)
\end{aligned}$$

$\mathbb{E}\big[\ln q(\mu,\lambda)\big]$ **:**

$$\begin{aligned}
\mathbb{E}\big[\ln q(\mu,\lambda)\big] =&\mathbb{E}\left[\ln\mathcal{N}\left(\mu|\mu_0,(\beta^{\mu}\lambda)^{-1}\right)\mathrm{Gam}(\lambda|a^{\lambda},b^{\lambda})\right] && (2.110)\\
=&\mathbb{E}\Bigg[\ln\Bigg\{\sqrt{\frac{\beta^{\mu}\lambda}{2\pi}}\exp\left(-\frac{\beta^{\mu}\lambda}{2}(\mu-\mu_0)^2\right)\\
&\times\frac{1}{\Gamma(a^{\lambda})}b^{\lambda^{a^{\lambda}}}\lambda^{a^{\lambda}-1}e^{-b^{\lambda}\lambda}\Bigg\}\Bigg] && (2.111)\\
=&\mathbb{E}\Bigg[\Bigg\{\frac{1}{2}\ln\beta^{\mu}+\frac{1}{2}\ln\lambda-\frac{1}{2}\ln 2\pi-\frac{\beta^{\mu}}{2}\lambda(\mu-\mu_0)^2\\
&-\ln\Gamma(a^{\lambda})+a^{\lambda}\ln b^{\lambda}+(a^{\lambda}-1)\ln\lambda-b^{\lambda}\lambda\Bigg\}\Bigg] && (2.112)\\
=&\frac{1}{2}\ln\beta^{\mu}-\frac{1}{2}\ln 2\pi-\frac{\beta^{\mu}}{2}\overline{\lambda(\mu-\mu_0)^2}\\
&-\ln\Gamma(a^{\lambda})+a^{\lambda}\ln b^{\lambda}+\left(a^{\lambda}-\frac{1}{2}\right)\overline{\ln\lambda}-b^{\lambda}\overline{\lambda} && (2.113)
\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{2}\ln\beta^{\mu}-\frac{1}{2}\ln 2\pi-\frac{\beta^{\mu}}{2}\left[\frac{1}{\beta^{\mu}}+\frac{a^{\lambda}}{b^{\lambda}}\left(\mu_0-\mu_0\right)^2\right]\\
&-\ln\Gamma(a^{\lambda})+a^{\lambda}\ln b^{\lambda}+\left(a^{\lambda}-\frac{1}{2}\right)\overline{\ln\lambda}-b^{\lambda}\overline{\lambda} \qquad (2.114)\\
=&\frac{1}{2}\ln\beta^{\mu}-\frac{1}{2}\ln 2\pi-\frac{1}{2}-\ln\Gamma(a^{\lambda})+a^{\lambda}\ln b^{\lambda}\\
&+\left(a^{\lambda}-\frac{1}{2}\right)\overline{\ln\lambda}-a^{\lambda} \qquad (2.115)
\end{aligned}
$$

2.3　２次元ブラウン運動の動径分布

２次元平面上で、拡散係数 D で自由拡散する粒子の、時刻 t 後の起点からの変位 x は、

$$
\begin{aligned}
p(x|D,t)=&\frac{x}{2Dt}\exp\left(-\frac{x^2}{4Dt}\right) \qquad (2.116)\\
=&\frac{\delta x}{2}\exp\left(-\frac{\delta x^2}{4}\right) \qquad (2.117)
\end{aligned}
$$

にしたがう。ただし、$\delta\equiv(Dt)^{-1}$ とする。

したがって emission probability は

$$
p(x_m|\mathbf{z}_m,\boldsymbol{\delta})=\left\{\prod_{i=1}^{K}\frac{\delta_i x_m}{2}\exp\left(-\frac{\delta_i {x_m}^2}{4}\right)\right\}^{z_{mi}} \qquad (2.118)
$$

となる。

状態遷移確率を、遷移確率行列 $\mathbf{A}$ で与えることにすると、$\mathbf{X},\mathbf{Z},\theta$ の同時分布は、

$$
\begin{aligned}
p(\mathbf{X},\mathbf{Z},\theta)=&p(\theta)\cdot\prod_{i}^{K}\pi^{z_{1i}}\cdot\prod_{n=2}^{N}\prod_{i}^{K}\prod_{j}^{K}{A_{ij}}^{z_{n-1,i}z_{nj}}\\
&\times\prod_{m=1}^{N}\prod_{i=1}^{K}\left\{\frac{\delta_i x_m}{2}\exp\left(-\frac{\delta_i {x_m}^2}{4}\right)\right\}^{z_{mi}} \qquad (2.119)
\end{aligned}
$$

となる。

2.3.1 E-step：

$$\ln p(x_m|\mathbf{z}_m,\boldsymbol{\delta}) = \sum_{i=1}^{K} z_{mi} \ln\left\{\frac{\delta_i x_m}{2}\exp\left(-\frac{\delta_i {x_m}^2}{4}\right)\right\} \tag{2.120}$$

$$= \sum_{i=1}^{K} z_{mi}\left\{\ln\delta_i + \ln x_m - \ln 2 - \frac{\delta_i {x_m}^2}{4}\right\} \tag{2.121}$$

$$\overline{\ln p(x_m|\mathbf{z}_m,\boldsymbol{\delta})} = \sum_{i=1}^{K} z_{mi}\left\{\overline{\ln\delta_i} - \ln x_m - \ln 2 - \frac{1}{4}\overline{\delta_i}{x_m}^2\right\} \tag{2.122}$$

式 (1.55) に式 (2.118) を適用し、

$$\tilde{p}(x_m|\mathbf{z}_m,\boldsymbol{\delta}) = \prod_{i}^{K}\exp\left(\overline{\ln\delta_i} - \ln x_m - \ln 2 - \frac{1}{4}\overline{\delta_i}{x_m}^2\right)^{z_{mi}} \tag{2.123}$$

を得る。

これを、式 (1.53), (1.54) とともに用いて forward-backward アルゴリズムを実行し、γ,ξ 分布を得る。

2.3.2 M-step：

$\boldsymbol{\pi}$, $\mathbf{A}$ に関しては、1.3 節と同様に式 (1.64–1.65, 1.71–1.72) を用いて期待値を求める。

各 i に関しては独立として、$\boldsymbol{\delta}$ に Gamma 事前分布

$$p(\delta_i) = \mathrm{Gam}(\delta_i|u_i^a, u_i^b) \tag{2.124}$$

を与えて (ただし、u_i^a, u_i^b はハイパーパラメータ)、

$$\ln q^*(\delta_i) = \ln\mathrm{Gam}(\delta_i|u_i^a, u_i^b) + \mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^{N} z_{mi}\cdot\ln\left\{\frac{\delta_i x_m}{2}\exp\left(-\frac{\delta_i {x_m}^2}{4}\right)\right\}\right] + \mathrm{const.} \tag{2.125}$$

δ_i に依存する項だけ取り出すと、

$$
\begin{aligned}
=&(u_i^a - 1)\ln\delta_i - u_i^b\delta_i + \mathbb{E}_{\mathbf{Z}}\left[\sum_{m=1}^{N} z_{mi}\left\{\ln\delta_i - \frac{\delta_i {x_m}^2}{4}\right\}\right] \\
&+ \text{const.} \qquad (2.126)\\
=&(u_i^a - 1)\ln\delta_i - u_i^b\delta_i + \sum_{m=1}^{N}\overline{z_{mi}}\left\{\ln\delta_i - \frac{\delta_i {x_m}^2}{4}\right\} + \text{const.} \qquad (2.127)\\
=&(u_i^a - 1)\ln\delta_i - u_i^b\delta_i + N_i\ln\delta_i - \frac{R_i}{4}\delta_i + \text{const.} \qquad (2.128)
\end{aligned}
$$

ただし、$N_i = \sum_m^N \overline{z_{mi}}$、$R_i = \sum_m^N \overline{z_{mi}}\cdot {x_m}^2$。

$$
\begin{aligned}
=&(u_i^a + N_i - 1)\ln\delta_i - \left(u_i^b + \frac{R_i}{4}\right)\delta_i + \text{const.} \qquad (2.129)\\
=&(a_i^\delta - 1)\ln\delta_i - b_i^\delta\delta_i + \text{const.} \qquad (2.130)
\end{aligned}
$$

ただし、

$$
\begin{aligned}
a_i^\delta =& u_i^a + N_i \qquad (2.131)\\
b_i^\delta =& u_i^b + \frac{R_i}{4} \qquad (2.132)
\end{aligned}
$$

これら a_i^δ, b_i^δ を使って Gamma 分布 (**??**) として、

$$
\begin{aligned}
q^*(\delta_i) =& \mathrm{Gam}(\delta_i | a_i^\delta, b_i^\delta) \qquad (2.133)\\
=& \frac{1}{\Gamma(a_i^\delta)} {b_i^\delta}^{a_i^\delta} {\delta_i}^{a_i^\delta - 1} e^{-b_i^\delta\delta_i} \qquad (2.134)
\end{aligned}
$$

を得る。Gamma 分布の性質と計算により、

$$
\begin{aligned}
\overline{\delta_i} =& \frac{a_i^\delta}{b_i^\delta} \qquad (2.135)\\
\overline{\ln\delta_i} =& \psi(a_i^\delta) - \ln b_i^\delta \qquad (2.136)
\end{aligned}
$$

が得られる。

2.3.3 変分下限 $\mathcal{L}(q)$ の算出

変分下限 $\mathcal{L}(q)$ を計算するため、式 (1.75) から、

$$
\begin{aligned}
\mathcal{L}(q) =& \mathbb{E}\big[\ln p(\boldsymbol{\pi})\big] + \mathbb{E}\big[\ln p(\mathbf{A})\big] + \mathbb{E}\big[\ln p(\boldsymbol{\mu}, \boldsymbol{\lambda})\big] \\
&- \mathbb{E}\big[\ln q(\boldsymbol{\pi})\big] - \mathbb{E}\big[\ln q(\mathbf{A})\big] - \mathbb{E}\big[\ln q(\boldsymbol{\mu}, \boldsymbol{\lambda})\big] + \prod_n^N \tilde{c}_n \qquad (2.137)
\end{aligned}
$$

を得る。

$\mathbb{E}\bigl[\ln p(\boldsymbol{\pi})\bigr]$, $\mathbb{E}\bigl[\ln p(\mathbf{A})\bigr]$, $\mathbb{E}\bigl[\ln q(\boldsymbol{\pi})\bigr]$, $\mathbb{E}\bigl[\ln q(\mathbf{A})\bigr]$ に関しては、式 (1.79), (1.83), (1.86), (1.90) で、$\tilde{c}_n$ に関しては E-step で、それぞれ得られているので、以下ではそれ以外の項について計算する。

$\mathbb{E}\bigl[\ln p(\boldsymbol{\delta})\bigr]$ **:**

$$\mathbb{E}\bigl[\ln p(\boldsymbol{\delta})\bigr] = \mathbb{E}\left[\ln \prod_i^K \mathrm{Gam}(\delta_i | u_i^a, u_i^b)\right] \tag{2.138}$$

$$= \mathbb{E}\left[\ln \prod_i^K \left\{\frac{1}{\Gamma(u_i^a)} {u_i^b}^{u_i^a} {\delta_i}^{u_i^a - 1} e^{-u_i^b \delta_i}\right\}\right] \tag{2.139}$$

$$= \mathbb{E}\left[\sum_i^K \left\{-\ln \Gamma(u_i^a) + u_i^a \ln u_i^b + (u_i^a - 1)\ln \delta_i - u_i^b \delta_i\right\}\right] \tag{2.140}$$

$$= \sum_i^K \left\{-\ln \Gamma(u_i^a) + u_i^a \ln u_i^b + (u_i^a - 1)\overline{\ln \delta_i} - u_i^b \overline{\delta_i}\right\} \tag{2.141}$$

$\mathbb{E}\bigl[\ln q(\boldsymbol{\delta})\bigr]$ **:**

$$\mathbb{E}\bigl[\ln q(\boldsymbol{\delta})\bigr] = \mathbb{E}\left[\ln \prod_i^K \mathrm{Gam}(\delta_i | a_i^\delta, b_i^\delta)\right] \tag{2.142}$$

$$= \mathbb{E}\left[\ln \prod_i^K \frac{1}{\Gamma(a_i^\delta)} {b_i^\delta}^{a_i^\delta} {\delta_i}^{a_i^\delta - 1} e^{-b_i^\delta \delta_i}\right] \tag{2.143}$$

$$= \mathbb{E}\left[\sum_i^K \left\{-\ln \Gamma(a_i^\delta) + a_i^\delta \ln b_i^\delta + (a_i^\delta - 1)\ln \delta_i - b_i^\delta \delta_i\right\}\right] \tag{2.144}$$

$$= \sum_i^K \left\{-\ln \Gamma(a_i^\delta) + a_i^\delta \ln b_i^\delta + (a_i^\delta - 1)\overline{\ln \delta_i} - b_i^\delta \overline{\delta_i}\right\} \tag{2.145}$$

$$= \sum_i^K \left\{-\ln \Gamma(a_i^\delta) + a_i^\delta \ln b_i^\delta + (a_i^\delta - 1)\overline{\ln \delta_i} - a_i^\delta\right\} \tag{2.146}$$