変分ベイズ法とは

正田 備也

masada@rikkyo.ac.jp

Contents

ベイズ的モデリングにおける変分法

- 観測データを表す確率変数を $\mathcal{X} \equiv \{x_1, \ldots, x_N\}$ とする
- データモデルのパラメータを $\mathbf{\Theta}$ とする
- ベイズ的なモデリングでは、$\mathcal{X}$ だけでなく $\mathbf{\Theta}$ も確率変数
- ベイズ的なモデリングで知りたいのは、事後分布 $p(\mathbf{\Theta}|\mathcal{X})$

$$p(\mathbf{\Theta}|\mathcal{X}) = \frac{p(\mathcal{X}|\mathbf{\Theta})p(\mathbf{\Theta})}{p(\mathcal{X})} \tag{1}$$

- 変分ベイズ推論は $p(\mathbf{\Theta}|\mathcal{X})$ を近似する分布 $q(\mathbf{\Theta})$ を求める
  - $q(\mathbf{\Theta})$ を変分法 (variational methods) で求める（後述）
  - $q(\mathbf{\Theta})$ を変分事後分布 (variational posterior distribution) と呼ぶ

前回のEMアルゴリズムでの議論のパターン

- 潜在変数 $\mathcal{Z} = \{z_1, \ldots, z_N\}$ を含むモデリングを行いたい
- 確率モデルを指定することで同時分布 $p(\mathcal{X}, \mathcal{Z}) = p(\mathcal{Z})p(\mathcal{X}|\mathcal{Z}) = \prod_{i=1}^{N} p(z_i)p(x_i|z_i)$ が得られる
- 潜在変数 $\mathcal{Z}$ の周辺化 $\sum_{\mathcal{Z}} p(\mathcal{X}, \mathcal{Z})$ により観測データの尤度 $p(\mathcal{X})$ は得られるのだが、大抵この尤度は計算できない
- Jensen の不等式を使い、対数尤度 $\ln p(\mathcal{X})$ の下界を得る

$$\ln p(\mathcal{X}) \geq \sum_{i=1}^{N} \sum_{z_i=1}^{K} q_{i,z_i} \ln \frac{p(z_i)p(x_i|z_i)}{q_{i,z_i}}$$

- この下界を最大化することで、様々な未知量を推定する

この議論のパターンを事後分布の推論へ適用

- 潜在変数 $\mathbf{\Theta}$ を含むモデリングを行いたい
- 確率モデルを指定することで観測データと潜在変数の同時分布 $p(\mathcal{X}, \mathbf{\Theta}) = p(\mathbf{\Theta})p(\mathcal{X}|\mathbf{\Theta}) = p(\mathbf{\Theta}) \prod_{i=1}^{N} p(x_i|\mathbf{\Theta})$ が得られる
- 潜在変数 $\mathbf{\Theta}$ の周辺化 $\int p(\mathcal{X}, \mathbf{\Theta}) d\mathbf{\Theta}$ により観測データの周辺尤度 $p(\mathcal{X})$ は得られるのだが、大抵この尤度は計算できない
- Jensen の不等式を使い、対数周辺尤度 $\ln p(\mathcal{X})$ の下界を得る

$$\ln p(\mathcal{X}) \geq \int q(\mathbf{\Theta}) \ln \frac{p(\mathbf{\Theta})p(\mathcal{X}|\mathbf{\Theta})}{q(\mathbf{\Theta})} d\mathbf{\Theta}$$

- この下界を最大化することで、様々な未知量を推定する
  - この下界を ELBO(Evidence Lower BOund; 変分下限) と呼ぶ

変分ベイズ法(variational Bayesian methods)とは

- Jensen の不等式を適用することで、ELBO を次のように得た
$$\ln p(\mathcal{X}) \geq \int q(\mathbf{\Theta}) \ln \frac{p(\mathbf{\Theta})p(\mathcal{X}|\mathbf{\Theta})}{q(\mathbf{\Theta})} d\mathbf{\Theta}$$
- 実は、ELBO を大きくすればするほど、$\mathbf{\Theta}$ が従う確率分布である $q(\mathbf{\Theta})$ が、事後分布 $p(\mathbf{\Theta}|\mathcal{X})$ に近くなっていく
- つまり、この $q(\mathbf{\Theta})$ は、事後分布を近似する分布とみなせるような分布になっている
- $q(\mathbf{\Theta})$ は変分法 (variational method) で求められるので、変分事後分布 (variational posterior) と呼ばれる

「変分 (variational)」の意味

- ELBO の最大化は、$q(\boldsymbol{\Theta})$ を変化させることでおこなう
- このとき、$q(\boldsymbol{\Theta})$ の密度関数のかたち自体を変化させる
- 逆に言うと、$q(\boldsymbol{\Theta})$ の密度関数が特定のかたちを持つと仮定した上で、その関数のパラメータを動かすのではない
  - パラメータについて微分することで最大化問題を解くのではなく、いわば "関数について微分する" ことで最大化問題を解いている
- とても直感的に言うと、関数のかたちを決めてそのパラメータを動かすのではなく、関数のかたち自体を動かすことで問題を解く方法を、変分法と呼ぶ（cf. 汎関数微分）

ELBOを最大化する根拠

- Jensenの不等式の左辺から右辺を引いたものを求めてみる

$$
\begin{aligned}
&\ln p(\mathcal{X}) - \int q(\boldsymbol{\Theta}) \ln \frac{p(\boldsymbol{\Theta}|\mathcal{X})p(\mathcal{X})}{q(\boldsymbol{\Theta})} d\boldsymbol{\Theta} \\
&= \ln p(\mathcal{X}) - \int q(\boldsymbol{\Theta}) \ln \frac{p(\boldsymbol{\Theta}|\mathcal{X})}{q(\boldsymbol{\Theta})} d\boldsymbol{\Theta} - \int q(\boldsymbol{\Theta}) \ln p(\mathcal{X}) d\boldsymbol{\Theta} \\
&= \ln p(\mathcal{X}) - \int q(\boldsymbol{\Theta}) \ln \frac{p(\boldsymbol{\Theta}|\mathcal{X})}{q(\boldsymbol{\Theta})} d\boldsymbol{\Theta} - \ln p(\mathcal{X}) \int q(\boldsymbol{\Theta}) d\boldsymbol{\Theta} \\
&= \int q(\boldsymbol{\Theta}) \ln \frac{q(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta}|\mathcal{X})} d\boldsymbol{\Theta} = D_{\mathsf{KL}}(q(\boldsymbol{\Theta}) \parallel p(\boldsymbol{\Theta}|\mathcal{X}))
\end{aligned} \tag{2}
$$

$\therefore$ ELBOを$\ln p(\mathcal{X})$に近づける$\Leftrightarrow$ $q(\boldsymbol{\Theta})$を$p(\boldsymbol{\Theta}|\mathcal{X})$に近づける

変分事後分布に関する factorization の仮定

- モデルパラメータ $\mathbf{\Theta}$ は、多数のパラメータからなっている
- $\mathbf{\Theta}$ について、排他的なグループ分け $\mathbf{\Theta} = \mathbf{\Theta}_1 \cup \ldots \cup \mathbf{\Theta}_m$ を行い、これに対応して変分事後分布が $q(\mathbf{\Theta}) = q(\mathbf{\Theta}_1) \cdots q(\mathbf{\Theta}_m)$ という積に分解されると仮定することが、よくある
  - 「posterior factorizes」でググってみよう
- 最も極端な場合、個々のパラメータが従う確率分布の積へ分解されると仮定することも、わりとある
  - つまり、$\mathbf{\Theta}$ が d 個のパラメータ $\theta_1, \ldots, \theta_d$ からなるとすると、$q(\mathbf{\Theta}) = q(\theta_1) \cdots q(\theta_d)$ という積へ分解されると仮定する
  - これを平均場近似 (mean field approximation) と呼ぶ

より実際的な変分ベイズ法

- 上述の factorization の仮定をおくと、それだけで、変分事後分布の密度関数のかたちが決まってしまうこともある
  - この後示す例が、そうなっている
- しかし実際には、$q(\boldsymbol{\Theta})$ の密度関数が特定のかたちを持つと仮定してしまった上で、その関数のパラメータを動かすことによって、ELBO を最大化することも多い
  - 例えば、$q(\boldsymbol{\Theta})$ が多変量正規分布だと仮定して、ELBO を最大化するような平均パラメータと共分散行列パラメータを求める、など
  - 変分オートエンコーダでは、$q(\boldsymbol{\Theta})$ が多変量正規分布だと仮定し、さらにその共分散行列が対角行列だと仮定する

Contents

変分ベイズ法によるデータ分析の手順

- データモデル $p(\mathcal{X}|\mathbf{\Theta})$ とモデルパラメータの事前分布 $p(\mathbf{\Theta};\mathbf{\Xi})$ を指定する
  - $\mathbf{\Xi}$ は事前分布のパラメータ、つまり、ハイパーパラメータ
- 同時分布 $p(\mathcal{X},\mathbf{\Theta};\mathbf{\Xi})$ の式を書き下す
- ELBO の式を書き下す
- 変分事後分布 $q(\mathbf{\Theta};\mathbf{\Psi})$ について何らかの仮定を行う
- その仮定を利用して、変分事後分布のパラメータ $\mathbf{\Psi}$ を求めるための式（多くの場合、更新式）を得る
- この式を実装して計算機で動かす

例：メッセージ受信数の変化点の検知

- この授業の最初に採り上げた例

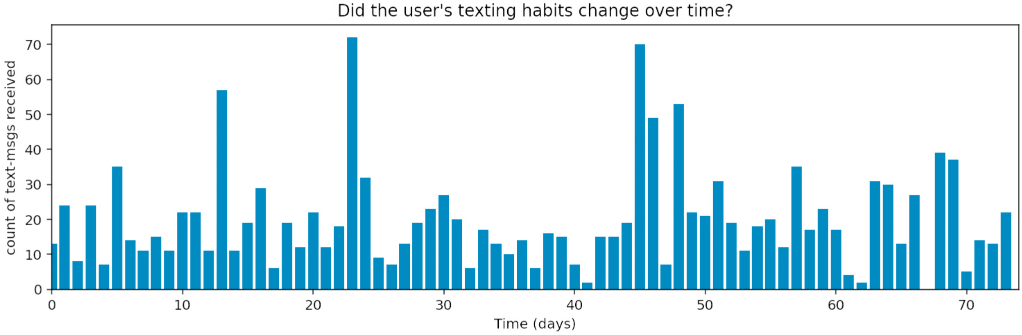

Figure: メッセージの受信数

モデルを指定する

- c_n が n 日目の受信数、τ が受信数の変化点、λ_1, λ_2 がそれぞれ $n < \tau, n \geq \tau$ の場合のポアソン分布のパラメータとする

$$
\begin{aligned}
\tau &\sim \texttt{Uniform}(1, N) \\
\lambda_1 &\sim \texttt{Gam}(a, b) \\
\lambda_2 &\sim \texttt{Gam}(a, b) \\
c_n &\sim \texttt{Poi}(\lambda_1) \quad \text{for } n < \tau \\
c_n &\sim \texttt{Poi}(\lambda_2) \quad \text{for } n \geq \tau
\end{aligned}
$$

- 参考: http://machine-learning.hatenablog.com/entry/2017/08/19/200841

同時分布を書き下す

同時分布は、観測データを $\boldsymbol{c} = \{c_1, \ldots, c_N\}$ とすると

$$
\begin{aligned}
p(\boldsymbol{c}, \lambda_1, \lambda_2, \tau) &= p(\boldsymbol{c}|\lambda 1, \lambda_2, \tau)p(\lambda_1; a, b)p(\lambda_2; a, b)p(\tau) \\
&= p(\lambda_1; a, b)p(\lambda_2; a, b)p(\tau) \prod_{n=1}^{N} p(c_n|\lambda_1)^{\delta(n<\tau)} p(c_n|\lambda_2)^{\delta(n\geq\tau)} \qquad (3)
\end{aligned}
$$

- $p(\lambda_i; a, b) \equiv \frac{b^a}{\Gamma(a)} \lambda_i^{a-1} e^{-b\lambda_i}$ for $i = 1, 2$
- $p(\tau) \equiv \frac{1}{N}$
- $\delta(\cdot)$ は、カッコ内の命題が真ならば 1、偽ならば 0
- $p(c_n|\lambda_i) \equiv \frac{\lambda_i^{c_n} e^{-\lambda_i}}{c_n!}$ for $i = 1, 2$

ELBOを書き下す

ELBO は

$$
\begin{aligned}
\ln p(\boldsymbol{c}) &= \ln \int \sum_{\tau} p(\boldsymbol{c}, \lambda_1, \lambda_2, \tau) d\lambda_1 d\lambda_2 \\
&\geq \int \sum_{\tau} q(\lambda_1, \lambda_2, \tau) \ln \frac{p(\boldsymbol{c}, \lambda_1, \lambda_2, \tau)}{q(\lambda_1, \lambda_2, \tau)} d\lambda_1 d\lambda_2 \qquad (4)
\end{aligned}
$$

- このままではこれ以上議論を進められないので、変分事後分布 $q(\lambda_1, \lambda_2, \tau)$ について、それを単純化するような、何らかの仮定を行う

平均場近似の仮定

ここでは、変分事後分布 $q(\lambda_1, \lambda_2, \tau)$ について、
$q(\lambda_1, \lambda_2, \tau) = q(\lambda_1)q(\lambda_2)q(\tau)$ と分解できることを仮定する

$$
\begin{aligned}
\ln p(\boldsymbol{c}) &\geq \int \sum_{\tau} q(\lambda_1, \lambda_2, \tau) \ln \frac{p(\boldsymbol{c}, \lambda_1, \lambda_2, \tau)}{q(\lambda_1, \lambda_2, \tau)} d\lambda_1 d\lambda_2 \\
&= \int \sum_{\tau} q(\lambda_1)q(\lambda_2)q(\tau) \ln \frac{p(\boldsymbol{c}, \lambda_1, \lambda_2, \tau)}{q(\lambda_1)q(\lambda_2)q(\tau)} d\lambda_1 d\lambda_2 \qquad (5)
\end{aligned}
$$

- 同時分布の式 (3) を使って、ELBO をさらに詳しく書き下す

ハイパーパラメータ a, b は省略する。

$$
\begin{aligned}
\ln p(\boldsymbol{c}) &\geq \int \sum_{\tau} q(\lambda_1)q(\lambda_2)q(\tau) \ln \frac{p(\boldsymbol{c}, \lambda_1, \lambda_2, \tau)}{q(\lambda_1)q(\lambda_2)q(\tau)} d\lambda_1 d\lambda_2 \\
&= \int \sum_{\tau} q(\lambda_1)q(\lambda_2)q(\tau) \ln \frac{p(\lambda_1)p(\lambda_2)p(\tau) \prod_{n=1}^{N} p(c_n|\lambda_1)^{\delta(n<\tau)} p(c_n|\lambda_2)^{\delta(n\geq\tau)}}{q(\lambda_1)q(\lambda_2)q(\tau)} d\lambda_1 d\lambda_2 \\
&= \int q(\lambda_1) \ln \frac{p(\lambda_1)}{q(\lambda_1)} d\lambda_1 + \int q(\lambda_2) \ln \frac{p(\lambda_2)}{q(\lambda_2)} d\lambda_2 + \sum_{\tau} q(\tau) \ln \frac{p(\tau)}{q(\tau)} \\
&\quad + \sum_{\tau} \sum_{n=1}^{N} \delta(n<\tau) \int q(\lambda_1) \ln p(c_n|\lambda_1) d\lambda_1 + \sum_{\tau} \sum_{n=1}^{N} \delta(n \geq \tau) \int q(\lambda_2) \ln p(c_n|\lambda_2) d\lambda_2 \\
&= -D_{\mathsf{KL}}(q(\lambda_1) \parallel p(\lambda_1)) - D_{\mathsf{KL}}(q(\lambda_2) \parallel p(\lambda_2)) - D_{\mathsf{KL}}(q(\tau) \parallel p(\tau)) \\
&\quad + \sum_{\tau} \sum_{n=1}^{N} \delta(n<\tau) \int q(\lambda_1) \ln p(c_n|\lambda_1) d\lambda_1 + \sum_{\tau} \sum_{n=1}^{N} \delta(n \geq \tau) \int q(\lambda_2) \ln p(c_n|\lambda_2) d\lambda_2 \quad (6)
\end{aligned}
$$

- 上記の ELBO の値を求めるコードも書いて、変分事後分布のパラメータ更新によって ELBO が徐々に大きくなっているかを、チェックする。

変分事後分布を求める

- この例の場合、平均場近似の仮定を置くと、変分事後分布の密度関数の式のかたちが決まってしまう
  - 結論を先取りすると、$q(\lambda_1)$ と $q(\lambda_2)$ の密度関数の式はガンマ分布の密度関数の式のかたちに一致し、$q(\tau)$ の質量関数の式はカテゴリカル分布の質量関数の式のかたちに一致する
- 具体的には、$q(\lambda_1)$ と $q(\lambda_2)$ と $q(\tau)$ のうち 2 つを固定し、残りの 1 つについて、変分事後分布の事後分布に対する KL 情報量 $D_{\mathrm{KL}}(q \parallel p)$ がゼロになる条件を明らかにする
- すると、その変分事後分布の密度関数のかたちがおのずと決まってくる

$q(\lambda_1)$ の密度関数のかたちを求める

$q(\lambda_2)$ と $q(\tau)$ を固定し、$D_{\mathsf{KL}}(q(\lambda_1)q(\lambda_2)q(\tau) \parallel p(\lambda_1, \lambda_2, \tau|\boldsymbol{c}))$ を最小にする $q(\lambda_1)$ を求める。

$$
\begin{aligned}
&D_{\mathsf{KL}}(q(\lambda_1)q(\lambda_2)q(\tau) \parallel p(\lambda_1, \lambda_2, \tau|\boldsymbol{c})) \\
&= \int \sum_{\tau} q(\lambda_1)q(\lambda_2)q(\tau) \ln \frac{q(\lambda_1)q(\lambda_2)q(\tau)}{p(\lambda_1, \lambda_2, \tau|\boldsymbol{c})} d\lambda_1 d\lambda_2 \\
&= \int q(\lambda_1)\Big\{ \ln q(\lambda_1) - \int \sum_{\tau} q(\lambda_2)q(\tau) \ln p(\lambda_1, \lambda_2, \tau|\boldsymbol{c}) d\lambda_2 \Big\} d\lambda_1 + const. \\
&= \int q(\lambda_1) \ln \frac{q(\lambda_1)}{\exp \int \sum_{\tau} q(\lambda_2)q(\tau) \ln p(\lambda_1, \lambda_2, \tau|\boldsymbol{c}) d\lambda_2} d\lambda_1 + const. \\
&= D_{\mathsf{KL}}(q(\lambda_1) \parallel \frac{1}{Z} \exp \int q(\lambda_2)q(\tau) \ln p(\lambda_1, \lambda_2, \tau|\boldsymbol{c}) d\lambda_2 d\tau) + const.
\end{aligned}
\tag{7}
$$

$q(\lambda_1) = \frac{1}{Z} \exp \int \sum_{\tau} q(\lambda_2)q(\tau) \ln p(\lambda_1, \lambda_2, \tau|\boldsymbol{c}) d\lambda_2$ のとき、上の KL 情報量は最小。つまり、
$\ln q(\lambda_1) = \int \sum_{\tau} q(\lambda_2)q(\tau) \ln p(\lambda_1, \lambda_2, \tau|\boldsymbol{c}) d\lambda_2 - \ln Z$ のとき、上の KL 情報量は最小。

$$\ln q(\lambda_1) = \int \sum_{\tau} q(\lambda_2) q(\tau) \ln \left\{ p(\lambda_1; a, b) p(\lambda_2; a, b) p(\tau) \prod_{n=1}^{N} p(c_n|\lambda_1)^{\delta(n<\tau)} p(c_n|\lambda_2)^{\delta(n\geq\tau)} \right\} d\lambda_2 - \ln Z$$

$$= \ln p(\lambda_1; a, b) + \int q(\lambda_2) \ln p(\lambda_2; a, b) d\lambda_2 + \sum_{\tau} q(\tau) \ln p(\tau)$$

$$+ \sum_{n=1}^{N} \sum_{\tau} q(\tau) \delta(n < \tau) \ln p(c_n|\lambda_1) + \sum_{n=1}^{N} \int \sum_{\tau} q(\lambda_2) q(\tau) \delta(n \geq \tau) \ln p(c_n|\lambda_2) d\lambda_2 - \ln Z$$

$$= \ln \frac{b^a}{\Gamma(a)} \lambda_1^{a-1} e^{-b\lambda_1} + \sum_{n=1}^{N} \Big(\sum_{\tau} q(\tau) \delta(n < \tau) \Big) \ln \frac{\lambda_1^{c_n} e^{-\lambda_1}}{c_n!} + const.$$

$$= \left(a - 1 + \sum_{n=1}^{N} \Big(\sum_{\tau} q(\tau) \delta(n < \tau) \Big) c_n \right) \ln \lambda_1 - \left(b + \sum_{n=1}^{N} \Big(\sum_{\tau} q(\tau) \delta(n < \tau) \Big) \right) \lambda_1 + const.$$

よって、$q(\lambda_1)$ は、shape パラメータが $a + \sum_{n=1}^{N} \Big(\sum_{\tau} q(\tau) \delta(n < \tau) \Big) c_n$ で、rate パラメータが $b + \sum_{n=1}^{N} \Big(\sum_{\tau} q(\tau) \delta(n < \tau) \Big)$ のガンマ分布となる。

$q(\lambda_2)$ についても同様に計算すると、やはりガンマ分布であることが分かる。

$q(\tau)$ のかたちを求める

$q(\lambda_1)$ と $q(\lambda_2)$ を固定し、$D_{\mathsf{KL}}(q(\lambda_1)q(\lambda_2)q(\tau) \parallel p(\lambda_1, \lambda_2, \tau | \boldsymbol{c}))$ を最小にする $q(\tau)$ を求める。

$$\begin{aligned}
&D_{\mathsf{KL}}(q(\lambda_1)q(\lambda_2)q(\tau) \parallel p(\lambda_1, \lambda_2, \tau | \boldsymbol{c})) \\
&= \int \sum_{\tau} q(\lambda_1)q(\lambda_2)q(\tau) \ln \frac{q(\lambda_1)q(\lambda_2)q(\tau)}{p(\lambda_1, \lambda_2, \tau | \boldsymbol{c})} d\lambda_1 d\lambda_2 \\
&= \sum_{\tau} q(\tau) \Big\{ \ln q(\tau) - \int q(\lambda_1)q(\lambda_2) \ln p(\lambda_1, \lambda_2, \tau | \boldsymbol{c}) d\lambda_1 d\lambda_2 \Big\} + const. \\
&= \sum_{\tau} q(\tau) \ln \frac{q(\tau)}{\exp \int q(\lambda_1)q(\lambda_2) \ln p(\lambda_1, \lambda_2, \tau | \boldsymbol{c}) d\lambda_1 d\lambda_2} + const. \\
&= D_{\mathsf{KL}}(q(\tau) \parallel \frac{1}{Z} \exp \int q(\lambda_1)q(\lambda_2) \ln p(\lambda_1, \lambda_2, \tau | \boldsymbol{c}) d\lambda_1 d\lambda_2) + const. \qquad (8)
\end{aligned}$$

$q(\tau) = \frac{1}{Z} \exp \int q(\lambda_1)q(\lambda_2) \ln p(\lambda_1, \lambda_2, \tau | \boldsymbol{c}) d\lambda_1 d\lambda_2$ のとき、上の KL 情報量は最小。つまり、
$\ln q(\tau) = \int q(\lambda_1)q(\lambda_2) \ln p(\lambda_1, \lambda_2, \tau | \boldsymbol{c}) d\lambda_1 d\lambda_2 - \ln Z$ のとき、上の KL 情報量は最小。

$$
\begin{aligned}
\ln q(\tau) &= \int q(\lambda_1)q(\lambda_2)\ln\left\{p(\lambda_1;a,b)p(\lambda_2;a,b)p(\tau)\prod_{n=1}^{N}p(c_n|\lambda_1)^{\delta(n<\tau)}p(c_n|\lambda_2)^{\delta(n\geq\tau)}\right\}d\lambda_1 d\lambda_2 - lnZ \\
&= \int q(\lambda_1)\ln p(\lambda_1;a,b)d\lambda_1 + \int q(\lambda_2)\ln p(\lambda_2;a,b)d\lambda_2 + \ln p(\tau) \\
&\quad + \sum_{n=1}^{N}\delta(n<\tau)\int q(\lambda_1)\ln p(c_n|\lambda_1)d\lambda_1 + \sum_{n=1}^{N}\delta(n\geq\tau)\int q(\lambda_2)\ln p(c_n|\lambda_2)d\lambda_2 + const.
\end{aligned}
\tag{9}
$$

よって、$q(\tau)$ はカテゴリカル分布であり、

$$
q(\tau) \propto \exp\left[\sum_{n=1}^{N}\delta(n<\tau)\int q(\lambda_1)\ln p(c_n|\lambda_1)d\lambda_1 + \sum_{n=1}^{N}\delta(n\geq\tau)\int q(\lambda_2)\ln p(c_n|\lambda_2)d\lambda_2\right]
$$

となる。ただし、$\sum_{\tau=1}^{N} q(\tau) = 1$ を満たす。

まとめ

- メッセージ受信数の変化点を検知するため、ベイズ的なモデルを立てた
- 事後分布を近似するために、変分ベイズ推論を行った
- その際、変分事後分布 $q(\lambda_1, \lambda_2, \tau)$ について、$q(\lambda_1, \lambda_2, \tau) = q(\lambda_1)q(\lambda_2)q(\tau)$ と分解できることを仮定した
- このように仮定すると、$q(\lambda_1)$ と $q(\lambda_2)$ はガンマ分布となり、$q(\tau)$ はカテゴリカル分布となった

課題9

- メッセージ受信数の変化点検知の例を考える。
- λ_1 の値が従う変分事後分布 $q(\lambda_1)$ は、ガンマ分布であることが分かった。
- そこで、$q(\lambda_1)$ の shape パラメータを α_1 とし、rate パラメータを β_1 とする。
- このとき、$\int q(\lambda_1) \ln p(\lambda_1; a, b) d\lambda_1$ を計算せよ。
  - これは ELBO の算出に必要な計算。