混合分布モデルのベイズ化

正田 備也

masada@rikkyo.ac.jp

Contents

混合分布

- 観測データを、いくつかのまとまりに分けられそうな場合、混合分布を観測データのモデルとして用いる
  - そのまとまりのことを、以下、「コンポーネント」と呼ぶ
- 各々のインスタンス $\boldsymbol{x}_i$ が、どのコンポーネントに属するか、すでに分かっている場合は、教師あり学習をおこなう
  - これは、分類 (classification)
- 各々のインスタンス $\boldsymbol{x}_i$ が、どのコンポーネントに属するか、観測されていない場合は、教師なし学習をおこなう
  - これは、クラスタリング (clustering) $\leftarrow$ 今回の設定はこちら

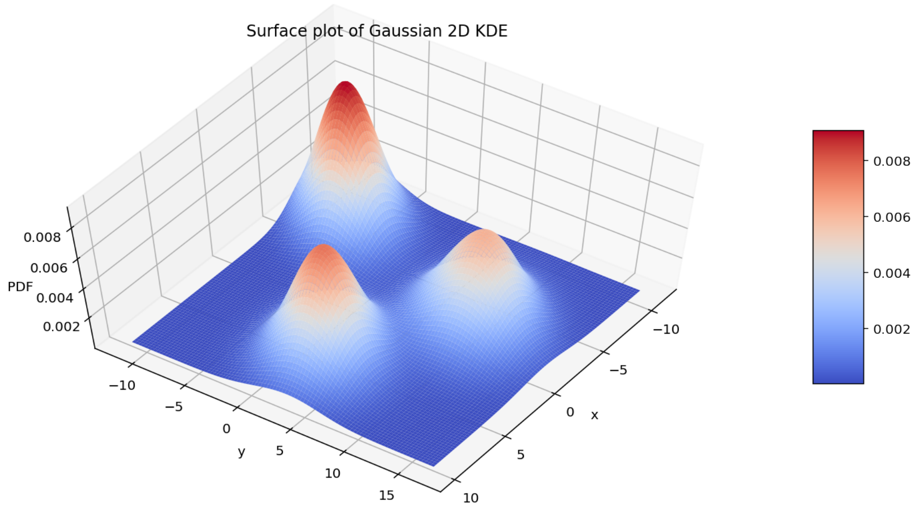

Figure: 2 次元ベクトルの集合をモデリングする混合分布の例

混合分布モデルを指定する

- 以下の二つを決めることにより混合分布モデルを指定できる

1. コンポーネントの個数 K を決める
   - k 番目のコンポーネントが選ばれる確率を θ_k とする
     - $(\theta_1, \ldots, \theta_K)$ を $\boldsymbol{\theta}$ と書く。$\sum_{k=1}^{K} \theta_k = 1$ を満たす。
2. 各コンポーネントを表す分布を決める
   - 全てのコンポーネントは同じ確率分布（例：正規分布）で表され、パラメータの値が違うだけ、とすることが多い
   - k 番目のコンポーネントを表す確率分布のパラメータを ϕ_k とする

     例: 単変量正規分布の場合は $\phi_k = (\mu_k, \sigma_k)$
   - $\phi_1, \ldots, \phi_K$ をまとめて $\boldsymbol{\Phi}$ と書く。

混合分布による観測データのモデリング

- 混合分布を使ったモデリングでは、N 個のインスタンス $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ が以下のように生成されると仮定する

1. カテゴリカル分布 $p(z; \boldsymbol{\theta})$ から、確率変数 z_i の値を draw
2. z_i 番目のコンポーネントを表す確率分布 $p(\boldsymbol{x}|z_i; \boldsymbol{\phi}_{z_i})$ から、$\boldsymbol{x}_i$ の値を draw

注: 確率モデルで観測データをモデリングすることとは、観測データがどのように生成されるかを様々な確率分布を組み合わせて記述することである

混合分布モデルの同時分布

- 前のスライドで書いた観測データの生成を、そのまま式にすると、同時分布の式になる
- $\boldsymbol{x}_i$ が属するコンポーネントを表す確率変数を z_i とする
  - $z_i = k$ は、$\boldsymbol{x}_i$ が k 番目のコンポーネントに属するという意味
- すると同時分布 $p(\mathcal{X}, \mathcal{Z})$ は

$$p(\mathcal{X}, \mathcal{Z}) = \prod_{i=1}^{N} p(z_i; \boldsymbol{\theta}) p(\boldsymbol{x}_i | z_i; \boldsymbol{\phi}_{z_i}) \tag{1}$$

  - ただし、$\mathcal{X} \equiv \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, $\mathcal{Z} \equiv \{z_1, \ldots, z_N\}$ と定義した
  - 教師なしの設定では、各 z_i は潜在変数

Contents

混合分布モデルをベイズ化する

- 事前分布を指定することにより混合分布をベイズ化できる
  - （下記の一方だけベイズ化するのもあり）

1. コンポーネント選択のカテゴリカル分布に用いる事前分布
   - z の値が従うカテゴリカル分布 $p(z|\boldsymbol{\theta})$ には、ディリクレ分布 $p(\boldsymbol{\theta};\boldsymbol{\alpha})$ を事前分布として使う
2. 各コンポーネントを表す確率分布に用いる事前分布
   - 各コンポーネントを表す分布 $p(\boldsymbol{x}|z,\boldsymbol{\phi}_z)$ は K 個ある
   - これら全てに対して、同じ一つの事前分布 $p(\boldsymbol{\phi};\boldsymbol{\beta})$ を使う

- 事後分布 $p(\boldsymbol{\theta},\boldsymbol{\Phi}|\mathcal{X})$ を求めることが課題

復習：ディリクレ分布

- $\sum_{k=1}^{K}\theta_k = 1$ を満たす K 個の非負実数 $(\theta_1, \ldots, \theta_K)$ の組の集合の上に定義される、連続確率分布
  - カテゴリカル分布や多項分布の事前分布として使える
- ディリクレ分布のパラメータは K 個の正実数 $(\alpha_1, \ldots, \alpha_K)$
- 密度関数の式は

$$p(\boldsymbol{\theta}; \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \tag{2}$$

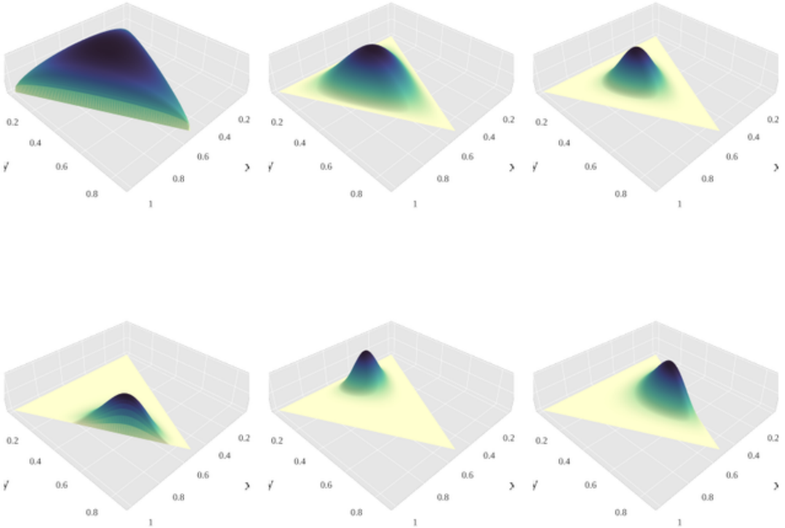

Figure: ディリクレ分布の確率密度関数の例 $(K = 3)$

ベイズ化された混合分布によるモデリング

- ベイズ化された混合分布を使ったモデリングでは、N 個のインスタンス $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ が以下のように生成されると仮定

1. ディリクレ事前分布 $p(\boldsymbol{\theta}; \boldsymbol{\alpha})$ から $\boldsymbol{\theta}$ の値を draw
2. 事前分布 $p(\boldsymbol{\phi}; \boldsymbol{\beta})$ からコンポーネント K 個分のパラメータ $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$ の値を draw
3. 各インスタンス x_i for $i = 1, \ldots, N$ を以下のように draw
   - 3.1 カテゴリカル分布 $p(z|\boldsymbol{\theta})$ から確率変数 z_i の値を draw
   - 3.2 第 z_i コンポーネントの確率分布 $p(\boldsymbol{x}|z_i, \boldsymbol{\phi}_{z_i})$ から $\boldsymbol{x}_i$ の値を draw

同時分布

- 前のスライドで書いた観測データの生成を、そのまま式にすると、同時分布の式になる

$$
\begin{aligned}
&p(\mathcal{X}, \mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\Phi}; \boldsymbol{\alpha}, \boldsymbol{\beta}) \\
&= p(\boldsymbol{\theta}; \boldsymbol{\alpha}) \times \prod_{k=1}^{K} p(\boldsymbol{\phi}_k; \boldsymbol{\beta}) \times \prod_{i=1}^{N} p(z_i | \boldsymbol{\theta}) p(\boldsymbol{x}_i | z_i, \boldsymbol{\phi}_{z_i})
\end{aligned} \tag{3}
$$

  - ベイズ化することで、$\boldsymbol{\theta}$ も $\boldsymbol{\Phi}$ も、いまや単なる自由パラメータではなく、確率変数になっている

ベイズ化された混合分布モデルの教師なし学習

- 潜在変数 $\mathcal{Z} = \{z_1, \dots, z_N\}$ を周辺化によって消去し…
- さらにモデルパラメータ $\boldsymbol{\theta}, \boldsymbol{\Phi}$ も周辺化によって消去し…
- 観測データの周辺尤度 $p(\mathcal{X}; \boldsymbol{\alpha}, \boldsymbol{\beta})$ を下のように得る

$$
\begin{aligned}
&p(\mathcal{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}) \\
&= \int \sum_{\mathcal{Z}} p(\boldsymbol{\theta}; \boldsymbol{\alpha}) \prod_{k=1}^{K} p(\boldsymbol{\phi}_k; \boldsymbol{\beta}) \prod_{i=1}^{N} p(z_i | \boldsymbol{\theta}) p(\boldsymbol{x}_i | z_i, \boldsymbol{\phi}_{z_i}) d\boldsymbol{\theta} d\boldsymbol{\Phi}
\end{aligned}
$$

- そして周辺尤度 $p(\mathcal{X}; \boldsymbol{\alpha}, \boldsymbol{\beta})$ を最大化する、という問題を解く
  - 対数周辺尤度 $\ln p(\mathcal{X}; \boldsymbol{\alpha}, \boldsymbol{\beta})$ を最大化する
- この最大化問題を解くことで事後分布を近似的に求める

Contents

混合ポアソン分布モデル

- 混合ポアソン分布を使ったモデリングでは、N 個の非負整数からなる観測データ $\{x_1, \ldots, x_N\}$ が以下のように生成されると仮定する

1. カテゴリカル分布 $p(z; \boldsymbol{\theta})$ から、確率変数 z_i の値を draw
2. z_i 番目のコンポーネントを表すポアソン分布 $p(x|z_i; \lambda_{z_i})$ から、x_i の値を draw

ベイズ化された混合ポアソン分布モデル

- ベイズ化された混合ポアソン分布を使ったモデリングでは、N 個の非負整数からなる観測データ $\{x_1, \ldots, x_N\}$ が以下のように生成されると仮定する

1. ディリクレ事前分布 $p(\boldsymbol{\theta}; \boldsymbol{\alpha})$ から $\boldsymbol{\theta}$ の値を draw
2. ガンマ事前分布 $p(\lambda; a, b)$ からコンポーネント K 個分のパラメータ $\lambda_1, \ldots, \lambda_K$ の値を draw
3. 各インスタンス x_i for $i = 1, \ldots, N$ を以下のように draw
   3.1 カテゴリカル分布 $p(z|\boldsymbol{\theta})$ から確率変数 z_i の値を draw
   3.2 第 z_i コンポーネントのポアソン分布 $p(x|z_i, \lambda_{z_i})$ から x_i の値を draw

同時分布

- 前のスライドで書いた観測データの生成を、そのまま式にすると、同時分布の式になる

$$
\begin{aligned}
&p(\mathcal{X}, \mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\alpha}, a, b) \\
&= p(\boldsymbol{\theta}; \boldsymbol{\alpha}) \times \prod_{k=1}^{K} p(\lambda_k; a, b) \times \prod_{i=1}^{N} p(z_i | \boldsymbol{\theta}) p(x_i | z_i, \lambda_{z_i}) \\
&= \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \times \prod_{k=1}^{K} \frac{b^a}{\Gamma(a)} \lambda_k^{a-1} e^{-b\lambda_k} \\
&\quad \times \prod_{i=1}^{N} \prod_{k=1}^{K} \left(\theta_k \times \frac{\lambda_k^{x_i} e^{-\lambda_k}}{x_i!} \right)^{\delta(z_i = k)}
\end{aligned} \tag{4}
$$

ELBO

Jensen の不等式を適用し、対数周辺尤度の下界、つまり ELBO(evidence lower bound) を得る。

$$
\begin{aligned}
&\ln p(\mathcal{X}; \boldsymbol{\alpha}, a, b) \\
&= \ln \int \sum_{\mathcal{Z}} p(\boldsymbol{\theta}; \boldsymbol{\alpha}) \prod_{k=1}^{K} p(\lambda_k; a, b) \prod_{i=1}^{N} p(z_i|\boldsymbol{\theta}) p(x_i|z_i, \lambda_{z_i}) d\boldsymbol{\theta} d\boldsymbol{\lambda} \\
&= \ln \int \sum_{\mathcal{Z}} q(\mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\lambda}) \frac{p(\boldsymbol{\theta}; \boldsymbol{\alpha}) \prod_{k=1}^{K} p(\lambda_k; a, b) \prod_{i=1}^{N} p(z_i|\boldsymbol{\theta}) p(x_i|z_i, \lambda_{z_i})}{q(\mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\lambda})} d\boldsymbol{\theta} d\boldsymbol{\lambda} \\
&\geq \int \sum_{\mathcal{Z}} q(\mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\lambda}) \ln \frac{p(\boldsymbol{\theta}; \boldsymbol{\alpha}) \prod_{k=1}^{K} p(\lambda_k; a, b) \prod_{i=1}^{N} p(z_i|\boldsymbol{\theta}) p(x_i|z_i, \lambda_{z_i})}{q(\mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\lambda})} d\boldsymbol{\theta} d\boldsymbol{\lambda} \qquad (5)
\end{aligned}
$$

以下、変分事後分布 $q(\mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\lambda})$ が $q(\mathcal{Z})q(\boldsymbol{\theta}, \boldsymbol{\lambda})$ と factorize する、と仮定する。

$q(\mathcal{Z})$ を求める

分布 $q(\boldsymbol{\theta}, \boldsymbol{\lambda})$ を固定する。

$$
\begin{aligned}
\ln p(\mathcal{X}; \boldsymbol{\alpha}, a, b) &\geq \int \sum_{\mathcal{Z}} q(\mathcal{Z}) q(\boldsymbol{\theta}, \boldsymbol{\lambda}) \ln \frac{p(\boldsymbol{\theta}; \boldsymbol{\alpha}) \prod_{k=1}^{K} p(\lambda_k; a, b) \prod_{i=1}^{N} p(z_i|\boldsymbol{\theta}) p(x_i|z_i, \lambda_{z_i})}{q(\mathcal{Z}) q(\boldsymbol{\theta}, \boldsymbol{\lambda})} d\boldsymbol{\theta} d\boldsymbol{\lambda} \\
&= \int \sum_{\mathcal{Z}} q(\mathcal{Z}) q(\boldsymbol{\theta}, \boldsymbol{\lambda}) \ln \left\{ \prod_{i=1}^{N} p(z_i|\boldsymbol{\theta}) p(x_i|z_i, \lambda_{z_i}) \right\} d\boldsymbol{\theta} d\boldsymbol{\lambda} - \sum_{\mathcal{Z}} q(\mathcal{Z}) \ln q(\mathcal{Z}) + const. \\
&= \sum_{i=1}^{N} \int \sum_{\mathcal{Z}} q(\mathcal{Z}) q(\boldsymbol{\theta}, \boldsymbol{\lambda}) \ln \left\{ p(z_i|\boldsymbol{\theta}) p(x_i|z_i, \lambda_{z_i}) \right\} d\boldsymbol{\theta} d\boldsymbol{\lambda} - \sum_{\mathcal{Z}} q(\mathcal{Z}) \ln q(\mathcal{Z}) + const. \qquad (6)
\end{aligned}
$$

最初の項の $\sum_{\mathcal{Z}}$ という和の内側には、各 z_i が単独でしか含まれない。
つまり、個々の z_i について $\sum_{z_i=1}^{K}$ という和を求めればよいだけである。
このことは、$q(\mathcal{Z})$ が $\prod_{i=1}^{N} q(z_i)$ と factorize することを意味する。

$$
\begin{aligned}
\ln p(\mathcal{X};\boldsymbol{\alpha},a,b) &\geq \sum_{i=1}^{N}\int\sum_{\mathcal{Z}} q(\mathcal{Z})q(\boldsymbol{\theta},\boldsymbol{\lambda})\ln\left\{p(z_i|\boldsymbol{\theta})p(x_i|z_i,\lambda_{z_i})\right\}d\boldsymbol{\theta}d\boldsymbol{\lambda} - \sum_{\mathcal{Z}} q(\mathcal{Z})\ln q(\mathcal{Z}) + const. \\
&= \sum_{i=1}^{N}\sum_{z_i=1}^{K} q(z_i)\int q(\boldsymbol{\theta},\boldsymbol{\lambda})\ln\left\{p(z_i|\boldsymbol{\theta})p(x_i|z_i,\lambda_{z_i})\right\}d\boldsymbol{\theta}d\boldsymbol{\lambda} - \sum_{i=1}^{N}\sum_{z_i=1}^{K} q(z_i)\ln q(z_i) + const. \\
&= -\sum_{i=1}^{N} D_{\mathsf{KL}}(q(z_i) \parallel \frac{1}{Z}\exp\int q(\boldsymbol{\theta},\boldsymbol{\lambda})\ln\left\{p(z_i|\boldsymbol{\theta})p(x_i|z_i,\lambda_{z_i})\right\}d\boldsymbol{\theta}d\boldsymbol{\lambda}) + const. \qquad (7)
\end{aligned}
$$

Z は、$\exp\int q(\boldsymbol{\theta},\boldsymbol{\lambda})\ln\left\{p(z_i|\boldsymbol{\theta})p(x_i|z_i,\lambda_{z_i})\right\}d\boldsymbol{\theta}d\boldsymbol{\lambda}$ を z_i の分布の質量関数にするための、規格化定数。そして、$q(z_i) = \frac{1}{Z}\exp\int q(\boldsymbol{\theta},\boldsymbol{\lambda})\ln\left\{p(z_i|\boldsymbol{\theta})p(x_i|z_i,\lambda_{z_i})\right\}d\boldsymbol{\theta}d\boldsymbol{\lambda}$ のとき、ELBO は最大。

$$
\begin{aligned}
\int q(\boldsymbol{\theta},\boldsymbol{\lambda})\ln\left\{p(z_i=k|\boldsymbol{\theta})p(x_i|z_i=k,\lambda_k)\right\}d\boldsymbol{\theta}d\boldsymbol{\lambda} &= \int q(\boldsymbol{\theta},\boldsymbol{\lambda})\ln\left(\theta_k\frac{\lambda_k^{x_i}e^{-\lambda_k}}{x_i!}\right)d\boldsymbol{\theta}d\boldsymbol{\lambda} \\
&= \int q(\boldsymbol{\theta},\boldsymbol{\lambda})(\ln\theta_k + x_i\ln\lambda_k - \lambda_k)d\boldsymbol{\theta}d\boldsymbol{\lambda} + const. \qquad (8)
\end{aligned}
$$

よって、$q(z_i=k) = \frac{\exp\int q(\boldsymbol{\theta},\boldsymbol{\lambda})(\ln\theta_k + x_i\ln\lambda_k - \lambda_k)d\boldsymbol{\theta}d\boldsymbol{\lambda}}{\sum_l \exp\int q(\boldsymbol{\theta},\boldsymbol{\lambda})(\ln\theta_l + x_i\ln\lambda_l - \lambda_l)d\boldsymbol{\theta}d\boldsymbol{\lambda}}$ を得る。

$q(\boldsymbol{\theta}, \boldsymbol{\lambda})$ を求める

次に、$q(\mathcal{Z})$ を固定する。

$$
\begin{aligned}
\ln p(\mathcal{X}; \boldsymbol{\alpha}, a, b) &\geq \int \sum_{\mathcal{Z}} q(\mathcal{Z}) q(\boldsymbol{\theta}, \boldsymbol{\lambda}) \ln \frac{p(\boldsymbol{\theta}; \boldsymbol{\alpha}) \prod_{k=1}^{K} p(\lambda_k; a, b) \prod_{i=1}^{N} p(z_i|\boldsymbol{\theta}) p(x_i|z_i, \lambda_{z_i})}{q(\mathcal{Z}) q(\boldsymbol{\theta}, \boldsymbol{\lambda})} d\boldsymbol{\theta} d\boldsymbol{\lambda} \\
&= \int q(\boldsymbol{\theta}, \boldsymbol{\lambda}) \ln p(\boldsymbol{\theta}; \boldsymbol{\alpha}) d\boldsymbol{\theta} d\boldsymbol{\lambda} + \sum_{i=1}^{N} \int q(\boldsymbol{\theta}, \boldsymbol{\lambda}) \sum_{z_i=1}^{K} q(z_i) \ln p(z_i|\boldsymbol{\theta}) d\boldsymbol{\theta} d\boldsymbol{\lambda} \\
&\quad + \sum_{k=1}^{K} \int q(\boldsymbol{\theta}, \boldsymbol{\lambda}) \ln p(\lambda_k; a, b) d\boldsymbol{\theta} d\boldsymbol{\lambda} + \sum_{i=1}^{N} \int q(\boldsymbol{\theta}, \boldsymbol{\lambda}) \sum_{z_i=1}^{K} q(z_i) \ln p(x_i|z_i, \lambda_{z_i}) d\boldsymbol{\theta} d\boldsymbol{\lambda} \\
&\quad - \int q(\boldsymbol{\theta}, \boldsymbol{\lambda}) \ln q(\boldsymbol{\theta}, \boldsymbol{\lambda}) d\boldsymbol{\theta} d\boldsymbol{\lambda} + const. \qquad (9)
\end{aligned}
$$

$q(\boldsymbol{\theta}, \boldsymbol{\lambda})$ と掛けられている関数を、$\boldsymbol{\theta}$ だけを含む関数と $\boldsymbol{\lambda}$ だけを含む関数とに、分けられる。
よって、$q(\boldsymbol{\theta}, \boldsymbol{\lambda})$ を $q(\boldsymbol{\theta}) p(\boldsymbol{\lambda})$ と factorize できる。

$$
\begin{aligned}
\ln p(\mathcal{X};\boldsymbol{\alpha},a,b) \geq \int \sum_{\mathcal{Z}} q(\mathcal{Z})q(\boldsymbol{\theta})q(\boldsymbol{\lambda}) \ln \frac{p(\boldsymbol{\theta};\boldsymbol{\alpha})\prod_{k=1}^{K} p(\lambda_k;a,b)\prod_{i=1}^{N} p(z_i|\boldsymbol{\theta})p(x_i|z_i,\lambda_{z_i})}{q(\mathcal{Z})q(\boldsymbol{\theta})q(\boldsymbol{\lambda})} d\boldsymbol{\theta} d\boldsymbol{\lambda} \\
= \int q(\boldsymbol{\theta}) \ln p(\boldsymbol{\theta};\boldsymbol{\alpha}) d\boldsymbol{\theta} + \sum_{i=1}^{N} \int q(\boldsymbol{\theta}) \sum_{z_i=1}^{K} q(z_i) \ln p(z_i|\boldsymbol{\theta}) d\boldsymbol{\theta} \\
+ \sum_{k=1}^{K} \int q(\boldsymbol{\lambda}) \ln p(\lambda_k;a,b) d\boldsymbol{\lambda} + \sum_{i=1}^{N} \int q(\boldsymbol{\lambda}) \sum_{z_i=1}^{K} q(z_i) \ln p(x_i|z_i,\lambda_{z_i}) d\boldsymbol{\lambda} \\
- \int q(\boldsymbol{\theta}) \ln q(\boldsymbol{\theta}) d\boldsymbol{\theta} - \int q(\boldsymbol{\lambda}) \ln q(\boldsymbol{\lambda}) d\boldsymbol{\lambda} + const. \\
= \int q(\boldsymbol{\theta}) \sum_{k=1}^{K} (\alpha_k - 1) \ln \theta_k d\boldsymbol{\theta} + \int q(\boldsymbol{\theta}) \sum_{k=1}^{K} \Big(\sum_{i=1}^{N} q(z_i = k) \Big) \ln \theta_k d\boldsymbol{\theta} \\
+ \sum_{k=1}^{K} \int q(\boldsymbol{\lambda}) \{(a-1)\ln \lambda_k - b\lambda_k\} d\boldsymbol{\lambda} + \sum_{k=1}^{K} \int q(\boldsymbol{\lambda}) \sum_{i=1}^{N} q(z_i = k)(x_i \ln \lambda_k - \lambda_k) d\boldsymbol{\lambda} \\
- \int q(\boldsymbol{\theta}) \ln q(\boldsymbol{\theta}) d\boldsymbol{\theta} - \int q(\boldsymbol{\lambda}) \ln q(\boldsymbol{\lambda}) d\boldsymbol{\lambda} + const. \qquad (10)
\end{aligned}
$$

$q(\boldsymbol{\theta})$ を求める

$q(\mathcal{Z})$ と $q(\boldsymbol{\lambda})$ を、固定する。

$$\ln p(\mathcal{X};\boldsymbol{\alpha},a,b) \geq \int \sum_{\mathcal{Z}} q(\mathcal{Z})q(\boldsymbol{\theta})q(\boldsymbol{\lambda}) \ln \frac{p(\boldsymbol{\theta};\boldsymbol{\alpha})\prod_{k=1}^{K} p(\lambda_k;a,b)\prod_{i=1}^{N} p(z_i|\boldsymbol{\theta})p(x_i|z_i,\lambda_{z_i})}{q(\mathcal{Z})q(\boldsymbol{\theta})q(\boldsymbol{\lambda})} d\boldsymbol{\theta} d\boldsymbol{\lambda}$$

$$= \int q(\boldsymbol{\theta}) \sum_{k=1}^{K} \left(\alpha_k + \sum_{i=1}^{N} q(z_i = k) - 1\right) \ln \theta_k d\boldsymbol{\theta} - \int q(\boldsymbol{\theta}) \ln q(\boldsymbol{\theta}) d\boldsymbol{\theta} + const.$$

$$= -D_{\mathsf{KL}}(q(\boldsymbol{\theta}) \parallel \frac{1}{Z} \exp\left[\sum_{k=1}^{K} \left(\alpha_k + \sum_{i=1}^{N} q(z_i = k) - 1\right) \ln \theta_k\right]) + const. \tag{11}$$

Z は $\exp\left[\sum_{k=1}^{K}\left(\alpha_k + \sum_{i=1}^{N} q(z_i = k) - 1\right)\ln \theta_k\right]$ を $\boldsymbol{\theta}$ の分布の密度関数にする規格化定数。
そして、$q(\boldsymbol{\theta}) = \frac{1}{Z}\exp\left[\sum_{k=1}^{K}\left(\alpha_k + \sum_{i=1}^{N} q(z_i = k) - 1\right)\ln \theta_k\right]$ のとき、ELBO は最大。
このとき、$q(\boldsymbol{\theta}) \propto \prod_{k=1}^{K} \theta_k^{\alpha_k + \sum_{i=1}^{N} q(z_i=k) - 1}$ となり、$q(\boldsymbol{\theta})$ はディリクレ分布だと分かる。

$q(\boldsymbol{\lambda})$ を求める

$q(\mathcal{Z})$ と $q(\boldsymbol{\theta})$ を、固定する。

$$\ln p(\mathcal{X};\boldsymbol{\alpha},a,b) \geq \int \sum_{\mathcal{Z}} q(\mathcal{Z})q(\boldsymbol{\theta})q(\boldsymbol{\lambda}) \ln \frac{p(\boldsymbol{\theta};\boldsymbol{\alpha}) \prod_{k=1}^{K} p(\lambda_k;a,b) \prod_{i=1}^{N} p(z_i|\boldsymbol{\theta})p(x_i|z_i,\lambda_{z_i})}{q(\mathcal{Z})q(\boldsymbol{\theta})q(\boldsymbol{\lambda})} d\boldsymbol{\theta} d\boldsymbol{\lambda}$$

$$= \sum_{k=1}^{K} \int q(\boldsymbol{\lambda})\{(a-1)\ln \lambda_k - b\lambda_k\} d\boldsymbol{\lambda} + \sum_{k=1}^{K} \int q(\boldsymbol{\lambda}) \sum_{i=1}^{N} q(z_i = k)(x_i \ln \lambda_k - \lambda_k) d\boldsymbol{\lambda}$$

$$- \int q(\boldsymbol{\lambda}) \ln q(\boldsymbol{\lambda}) d\boldsymbol{\lambda} + const.$$

$$= -D_{\mathsf{KL}}\left(q(\boldsymbol{\lambda}) \parallel \frac{1}{Z} \exp \sum_{k=1}^{K} \left[\left\{a + \sum_{i=1}^{N} q(z_i = k)x_i - 1\right\} \ln \lambda_k - \left\{b + \sum_{i=1}^{N} q(z_i = k)\right\}\lambda_k\right]\right) + const.$$

$q(\boldsymbol{\lambda}) \propto \exp \sum_{k=1}^{K} \left[\left\{a + \sum_{i=1}^{N} q(z_i = k)x_i - 1\right\} \ln \lambda_k - \left\{b + \sum_{i=1}^{N} q(z_i = k)\right\}\lambda_k\right]$ の時、ELBO は最大。このとき、$q(\lambda_k) \propto \lambda_k^{a+\sum_{i=1}^{N} q(z_i=k)x_i - 1} e^{-\{b+\sum_{i=1}^{N} q(z_i=k)\}\lambda_k}$ となり、各 $q(\lambda_k)$ は、ガンマ分布であることが分かる。

$q(z_i)$ に戻る

ディリクレ変分事後分布 $q(\boldsymbol{\theta})$ のパラメータを ζ とする。

ガンマ変分事後分布 $q(\lambda_k)$ のパラメータを a_k, b_k とする。

$$q(z_i = k) \propto \exp \int q(\boldsymbol{\theta}; \boldsymbol{\zeta}) q(\lambda_k; a_k, b_k)(\ln \theta_k + x_i \ln \lambda_k - \lambda_k) d\boldsymbol{\theta} d\lambda_k$$

$$= \left(\exp \int q(\boldsymbol{\theta}; \boldsymbol{\zeta}) \ln \theta_k d\boldsymbol{\theta} \right) \left(\exp \int q(\lambda_k; a_k, b_k)(x_i \ln \lambda_k - \lambda_k) d\lambda_k \right)$$

$$= \exp \left(\psi(\zeta_k) - \psi(\sum_k \zeta_k) \right) \exp \left(x_i \psi(a_k) - x_i \ln b_k - \frac{a_k}{b_k} \right) \tag{12}$$

$$\therefore q(z_i = k) = \frac{\exp \left(\psi(\zeta_k) + x_i \{\psi(a_k) - \ln b_k\} - a_k / b_k \right)}{\sum_l \exp \left(\psi(\zeta_l) + x_i \{\psi(a_l) - \ln b_l\} - a_l / b_l \right)} \tag{13}$$

変分事後分布の更新式のまとめ

$$q(z_i = k) \leftarrow \frac{\exp\left(\psi(\zeta_k) + x_i\{\psi(a_k) - \ln b_k\} - a_k/b_k\right)}{\sum_l \exp\left(\psi(\zeta_l) + x_i\{\psi(a_l) - \ln b_l\} - a_l/b_l\right)} \tag{14}$$

$$\zeta_k \leftarrow \alpha_k + \sum_{i=1}^{N} q(z_i = k) \tag{15}$$

$$a_k \leftarrow a + \sum_{i=1}^{N} q(z_i = k) x_i \tag{16}$$

$$b_k \leftarrow b + \sum_{i=1}^{N} q(z_i = k) \tag{17}$$

変分事後分布を使ってELBOを書き下す

変分事後分布 $q(\mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\lambda})$ が、$\prod_{i=1}^{N} q(z_i) \times q(\boldsymbol{\theta}) \times \prod_{k=1}^{K} q(\lambda_k)$ と factorize することが分かった。よって、ELBO は以下のように書ける。

$$
\begin{aligned}
&\ln p(\mathcal{X}; \boldsymbol{\alpha}, a, b) \\
&\geq \int \sum_{\mathcal{Z}} q(\mathcal{Z}) q(\boldsymbol{\theta}) q(\boldsymbol{\lambda}) \ln \frac{p(\boldsymbol{\theta}; \boldsymbol{\alpha}) \prod_{k=1}^{K} p(\lambda_k; a, b) \prod_{i=1}^{N} p(z_i|\boldsymbol{\theta}) p(x_i|z_i, \lambda_{z_i})}{q(\mathcal{Z}) q(\boldsymbol{\theta}) q(\boldsymbol{\lambda})} d\boldsymbol{\theta} d\boldsymbol{\lambda} \\
&= \sum_{i=1}^{N} \int \sum_{z_i=1}^{K} q(z_i) q(\boldsymbol{\theta}) \ln p(z_i|\boldsymbol{\theta}) d\boldsymbol{\theta} + \sum_{i=1}^{N} \int \sum_{z_i=1}^{K} q(z_i) q(\lambda_{z_i}) \ln p(x_i|z_i, \lambda_{z_i}) d\lambda_{z_i} \\
&\quad - D_{\mathsf{KL}}(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta}; \boldsymbol{\alpha})) - \sum_{k=1}^{K} D_{\mathsf{KL}}(q(\lambda_k) \parallel p(\lambda_k; a, b)) - \sum_{i=1}^{N} \sum_{z_i=1}^{K} q(z_i) \ln q(z_i)
\end{aligned}
\tag{18}
$$

この ELBO の計算も実装し、変分ベイズ法によって変分事後分布のパラメータを更新することで、本当に ELBO が少しずつ大きくなっていっているかを、ちゃんとチェックする。

$q(z_i = k) \equiv \gamma_{i,k}$ と書くことにすると

$$\begin{aligned}\int \sum_{z_i=1}^{K} q(z_i) q(\boldsymbol{\theta}) \ln p(z_i|\boldsymbol{\theta}) d\boldsymbol{\theta} &= \sum_{k=1}^{K} \gamma_{i,k} \left(\int q(\boldsymbol{\theta}; \boldsymbol{\zeta}) \ln \theta_k d\boldsymbol{\theta} \right) \\ &= \sum_{k=1}^{K} \gamma_{i,k} \big(\psi(\zeta_k) - \psi(\zeta_0) \big) \quad \text{where } \zeta_0 \equiv \textstyle\sum_{k=1}^{K} \zeta_k \end{aligned} \tag{19}$$

$$\begin{aligned}\int \sum_{z_i=1}^{K} q(z_i) q(\lambda_{z_i}) \ln p(x_i|z_i, \lambda_{z_i}) d\lambda_{z_i} &= \sum_{k=1}^{K} \gamma_{i,k} \int q(\lambda_k) \ln p(x_i; \lambda_k) d\lambda_k \\ &= \sum_{k=1}^{K} \gamma_{i,k} \int q(\lambda_k) \ln \left(\frac{\lambda_k^{x_i} e^{-\lambda_k}}{x_i!} \right) d\lambda_k \\ &= \sum_{k=1}^{K} \gamma_{i,k} \left(x_i \int q(\lambda_k) \ln \lambda_k d\lambda_k - \int q(\lambda_k) \lambda_k d\lambda_k - \ln(x_i!) \right) \\ &= \sum_{k=1}^{K} \gamma_{i,k} \left(x_i \big(\psi(a_k) - \ln b_k \big) - \frac{a_k}{b_k} - \ln(x_i!) \right) \end{aligned} \tag{20}$$

$$-D_{\mathsf{KL}}(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta};\boldsymbol{\alpha})) = \int q(\boldsymbol{\theta}) \ln p(\boldsymbol{\theta};\boldsymbol{\alpha}) d\boldsymbol{\theta} - \int q(\boldsymbol{\theta}) \ln q(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{21}$$

$$\begin{aligned} \int q(\boldsymbol{\theta}) \ln p(\boldsymbol{\theta};\boldsymbol{\alpha}) d\boldsymbol{\theta} &= \int q(\boldsymbol{\theta}) \ln \left(\frac{\Gamma(\alpha_0)}{\prod_k \Gamma(\alpha_k)} \prod_k \theta_k^{\alpha_k - 1} \right) d\boldsymbol{\theta} \\ &= \ln \Gamma(\alpha_0) - \sum_{k=1}^{K} \ln \Gamma(\alpha_k) + \sum_{k=1}^{K} (\alpha_k - 1) \left(\int q(\boldsymbol{\theta}) \ln \theta_k d\boldsymbol{\theta} \right) \\ &= \ln \Gamma(\alpha_0) - \sum_{k=1}^{K} \ln \Gamma(\alpha_k) + \sum_{k=1}^{K} (\alpha_k - 1) \big(\psi(\zeta_k) - \psi(\zeta_0) \big) \end{aligned} \tag{22}$$

$$-\int q(\boldsymbol{\theta}) \ln q(\boldsymbol{\theta}) d\boldsymbol{\theta} = -\ln \Gamma(\zeta_0) + \sum_{k=1}^{K} \ln \Gamma(\zeta_k) - \sum_{k=1}^{K} (\zeta_k - 1) \big(\psi(\zeta_k) - \psi(\zeta_0) \big) \tag{23}$$

$$-D_{\mathsf{KL}}(q(\lambda_k) \parallel p(\lambda_k; a, b)) = \int q(\lambda_k) \ln p(\lambda_k; a, b) d\lambda_k - \int q(\lambda_k) \ln q(\lambda_k) d\lambda_k \tag{24}$$

$$\begin{aligned}
\int q(\lambda_k) \ln p(\lambda_k; a, b) d\lambda_k &= \int q(\lambda_k) \ln \left(\frac{b^a}{\Gamma(a)} \lambda_k^{a-1} e^{-b\lambda_k} \right) d\lambda_k \\
&= a \ln b - \ln \Gamma(a) + (a-1) \int q(\lambda_k) \ln \lambda_k d\lambda_k - b \int q(\lambda_k) \lambda_k d\lambda_k \\
&= a \ln b - \ln \Gamma(a) + (a-1)\big(\psi(a_k) - \ln b_k\big) - \frac{b a_k}{b_k}
\end{aligned} \tag{25}$$

$$-\int q(\lambda_k) \ln q(\lambda_k) d\lambda_k = -a_k \ln b_k + \ln \Gamma(a_k) - (a_k - 1)\big(\psi(a_k) - \ln b_k\big) + a_k \tag{26}$$

$$\sum_{z_i=1}^{K} q(z_i) \ln q(z_i) = \sum_{k=1}^{K} \gamma_{i,k} \ln \gamma_{i,k} \tag{27}$$