ベイズ本 4.3.4 節

ここでは，ポアソン混合分布に対する崩壊型ギブスサンプリングのアルゴリズムについてまとめる．

■Notation

以下の記号を用いる．

- $\boldsymbol{X} = \{x_1, \dots, x_N\}$: データ
- $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_K\}$: ポアソン分布のパラメータの集合
- $\boldsymbol{S} = \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_N\}$: 潜在変数の集合
  - $\boldsymbol{s}_n$: k 次元ベクトル
  - $s_{n,k} = 1 \Longleftrightarrow k$ 番目のクラスタが指定された
- $\boldsymbol{\pi}$: 混合比率
  - $\sum_{k=1}^{K} \pi_k = 1$

■モデル

まず，以下の確率分布を定義する．

- クラスタ k に対する観測モデル:

$$p(x_n|\lambda_k) = \mathrm{Poi}(x_n|\lambda_k), \quad k = 1, \dots, K. \tag{1}$$

- $\boldsymbol{s}_n$ をサンプルするための分布:

$$p(\boldsymbol{s}_n|\boldsymbol{\pi}) = \mathrm{Cat}(\boldsymbol{s}_n|\boldsymbol{\pi}). \tag{2}$$

- 混合分布における条件付き分布:

$$p(x_n|\boldsymbol{s}_n, \boldsymbol{\lambda}) = \sum_{k=1}^{K} \mathrm{Poi}(x_n|\lambda_k)^{s_{n,k}}. \tag{3}$$

- ポアソン分布のパラメータ $\boldsymbol{\lambda}$ に対する事前分布:

$$p(\lambda_k) = \mathrm{Gam}(\lambda_k|a, b), \quad k = 1, \dots, K. \tag{4}$$

  - a, b: ハイパーパラメータ．
- 混合比率 $\boldsymbol{\pi}$ に対する事前分布:

$$p(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}). \tag{5}$$

– $\boldsymbol{\alpha}$: ハイパーパラメータ．K 次元ベクトル．

これらを用いると，同時分布は

$$p(\boldsymbol{X}, \boldsymbol{S}, \boldsymbol{\lambda}, \boldsymbol{\pi}) = p(\boldsymbol{X}|\boldsymbol{S}, \boldsymbol{\lambda})p(\boldsymbol{S}|\boldsymbol{\pi})p(\boldsymbol{\lambda})p(\boldsymbol{\pi})$$
$$= \left(\prod_{n=1}^{N} p(x_n|\boldsymbol{s}_n, \boldsymbol{\lambda})p(\boldsymbol{s}_n|\boldsymbol{\pi})\right)\left(\prod_{k=1}^{K} p(\lambda_k)\right)p(\boldsymbol{\pi}) \tag{6}$$

と表される [1].

■崩壊型ギブスサンプリング

細かい導出は省略し，全体の流れをまとめる．

混合モデルにおける崩壊型ギブスサンプリングでは，同時分布からパラメータを周辺化除去することを考える．

$$p(\boldsymbol{X}, \boldsymbol{S}) = \iint p(\boldsymbol{X}, \boldsymbol{S}, \boldsymbol{\lambda}, \boldsymbol{\pi})d\boldsymbol{\lambda}d\boldsymbol{\pi} \tag{7}$$

周辺化後，$\boldsymbol{S}$ を $p(\boldsymbol{S}|\boldsymbol{X})$ からサンプルできれば良い．ここでは，周辺化されたモデルの事後分布 $p(\boldsymbol{S}|\boldsymbol{X})$ に対してギブスサンプリングを適用し，各 $\boldsymbol{s}_1, \dots, \boldsymbol{s}_N$ を別々にサンプリングする．つまり，あるサンプルしたい潜在変数 $\boldsymbol{s}_n$ 以外の全ての潜在変数の集合 $\boldsymbol{S}_{-n} \coloneqq \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_{n-1}, \boldsymbol{s}_{n+1}, \dots, \boldsymbol{s}_N\}$ に対して，条件付き分布 $p(\boldsymbol{s}_n|\boldsymbol{X}, \boldsymbol{S}_{-n})$ が十分簡単な確率分布として得られればよい，という考えの下の処理である．

ここで，$p(\boldsymbol{s}_n|\boldsymbol{X}, \boldsymbol{S}_{-n})$ は,

$$p(\boldsymbol{s}_n|\boldsymbol{X}, \boldsymbol{S}_{-n}) \propto p(x_n|\boldsymbol{X}_{-n}, \boldsymbol{s}_n, \boldsymbol{S}_{-n})p(\boldsymbol{s}_n|\boldsymbol{S}_{-n}) \tag{8}$$

のように分解できる．ただし，$p(\boldsymbol{s}_n|\boldsymbol{S}_{-n})$ は,

$$p(\boldsymbol{s}_n|\boldsymbol{S}_{-n}) = \mathrm{Cat}(\boldsymbol{s}_n|\boldsymbol{\eta}_{-n}), \tag{9}$$
$$\eta_{-n,k} \propto \hat{\alpha}_{-n,k} = \sum_{n' \neq n} s_{n',k} + \alpha_k \tag{10}$$

である．ここで，カテゴリ分布の定義は,

$$\mathrm{Cat}(\boldsymbol{s}|\boldsymbol{\eta}) = \prod_{k=1}^{K} \eta_k^{s_k} \tag{11}$$

である．また, $p(x_n|\boldsymbol{X}_{-n}, \boldsymbol{s}_n, \boldsymbol{S}_{-n})$ を求めるために, ある k に対して $s_{n,k} = 1$ となるときの $p(x_n|\boldsymbol{X}_{-n}, s_{n,k} = 1, \boldsymbol{S}_{-n})$ を考える．

$$p(x_n|\boldsymbol{X}_{-n}, s_{n,k} = 1, \boldsymbol{S}_{-n}) = \mathrm{NB}\left(x_n \,\middle|\, \hat{a}_{-n,k}, \frac{1}{\hat{b}_{-n,k} + 1}\right) \tag{12}$$

であるので,

$$p(x_n|\boldsymbol{X}_{-n}, \boldsymbol{s}_n, \boldsymbol{S}_{-n}) = \prod_{k=1}^{K} \mathrm{NB}\left(x_n \,\middle|\, \hat{a}_{-n,k}, \frac{1}{\hat{b}_{-n,k} + 1}\right)^{s_{n,k}} \tag{13}$$

である．これは，$s_{n,k'}=0$ である全ての k' に対して 1 が掛けられることから，$s_{n,k}=1$ となる k の負の二項分布の項のみが残るため，従う．ここで，

$$\hat{a}_{-n,k} = \sum_{n' \neq n} s_{n',k} x_{n'} + a \tag{14}$$

$$\hat{b}_{-n,k} = \sum_{n' \neq n} s_{n',k} + b \tag{15}$$

である．また，負の二項分布の定義は，

$$\mathrm{NB}(x|r,p) = \frac{\Gamma(x+r)}{x!\Gamma(r)}(1-p)^r p^x \tag{16}$$

である．ここで，

$$\frac{\Gamma(x+r)}{x!\Gamma(r)} = \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} = \frac{\Gamma(x+r)}{x\Gamma(x)\Gamma(r)} = \frac{1}{xB(x,r)} \tag{17}$$

である．

よって，

$$\begin{aligned} p(\boldsymbol{s}_n|\boldsymbol{X},\boldsymbol{S}_{-n}) &\propto \left(\prod_{k=1}^{K} \mathrm{NB}\left(x_n \,\middle|\, \hat{a}_{-n,k}, \frac{1}{\hat{b}_{-n,k}+1}\right)^{s_{n,k}}\right)\left(\prod_{k=1}^{K} \eta_{-n,k}^{s_{n,k}}\right) \quad (\because (8),(9),(11),(13)) \\ &= \prod_{k=1}^{K}\left(\mathrm{NB}\left(x_n \,\middle|\, \hat{a}_{-n,k}, \frac{1}{\hat{b}_{-n,k}+1}\right)\eta_{-n,k}\right)^{s_{n,k}} \quad \left(\because s_{n,k}\in\{0,1\},\ \sum_{k=1}^{K} s_{n,k}=1\right) \\ &= \mathrm{Cat}(\boldsymbol{s}_n|\boldsymbol{\xi}_{-n}) \end{aligned} \tag{18}$$

を得る．ただし，

$$\xi_{-n,k} \propto \mathrm{NB}\left(x_n \,\middle|\, \hat{a}_{-n,k}, \frac{1}{\hat{b}_{-n,k}+1}\right)\hat{\alpha}_{-n,k} \tag{19}$$

である．実装の際は，全ての k に対して (19) の右辺を計算し，その後 $\sum_{k=1}^{K}\xi_{-n,k}=1$ となるように正規化する．

崩壊型ギブスサンプリングでは，各 n 番目のデータ点ごとにサンプリングを行うので，毎回のサンプリングに関して $N-1$ 個のデータや潜在変数の足し合わせが不要である．つまり，i 番目のサンプルを得た直後に j 番目のサンプルを得たいとすると，

$$\hat{\alpha}_{-j,k} = \hat{\alpha}_{-i,k} + s_{i,k} - s_{j,k} \tag{20}$$

$$\hat{a}_{-j,k} = \hat{a}_{-i,k} + s_{i,k}x_i - s_{j,k}x_j \tag{21}$$

$$\hat{b}_{-j,k} = \hat{b}_{-i,k} + s_{i,k} - s_{j,k} \tag{22}$$

とすればよい．実装の際には，Algorithm 3 のようにすると良い．i 回目の繰り返しにおけるサンプルやパラメータには上付きの i を付けるものとする．

$$\hat{\alpha}_k = \sum_{n=1}^{N} s^0_{n,k} + \alpha_k \tag{23}$$

$$\hat{a}_k = \sum_{n=1}^{N} s^0_{n,k} x_n + a \tag{24}$$

$$\hat{b}_k = \sum_{n=1}^{N} s^0_{n,k} + b \tag{25}$$

とする.

Algorithm 1 Erase statistics related to nth data. Erase($\hat{\boldsymbol{\alpha}}, \hat{a}, \hat{b}, \boldsymbol{s}^{i-1}_n, x_n$).

Input: $\hat{\boldsymbol{\alpha}}$, $\hat{a}$, $\hat{b}$: parameters, $\boldsymbol{s}^{i-1}_n$: The sample when $i-1$th iteration, x_n: nth data.

Output: $\hat{\boldsymbol{\alpha}}^i_{-n}$, $\hat{a}^i_{-n}$, $\hat{b}^i_{-n}$: modified parameter.

$\hat{\boldsymbol{\alpha}}^i_{-n} \leftarrow \mathbf{0} \in \mathbb{R}^K$, $\hat{a}^i_{-n} \leftarrow \mathbf{0} \in \mathbb{R}^K$, $\hat{b}^i_{-n} \leftarrow \mathbf{0} \in \mathbb{R}^K$.
for $k = 1$ to K **do**
    Update $\hat{\alpha}^i_{-n,k}$, $\hat{a}^i_{-n,k}$, $\hat{b}^i_{-n,k}$ as

$$\hat{\alpha}^i_{-n,k} \leftarrow \hat{\alpha}_k - s^{i-1}_{n,k},$$
$$\hat{a}^i_{-n,k} \leftarrow \hat{a}_k - s^{i-1}_{n,k} x_n,$$
$$\hat{b}^i_{-n,k} \leftarrow \hat{b}_k - s^{i-1}_{n,k}.$$

end for
return $\hat{\boldsymbol{\alpha}}^i_{-n}$, $\hat{a}^i_{-n}$, $\hat{b}^i_{-n}$

Algorithm 2 Add statistics related to nth data. Add($\hat{\boldsymbol{\alpha}}, \hat{a}, \hat{b}, \boldsymbol{\alpha}^{\hat{i}}_{-n}, \hat{a}^i_{-n}, \hat{b}^i_{-n}, \boldsymbol{s}^i_n, x_n$).

Input: $\hat{\boldsymbol{\alpha}}$, $\hat{a}$, $\hat{b}$: parameters, $\boldsymbol{\alpha}^{\hat{i}}_{-n}$, $\hat{a}^i_{-n}$, $\hat{b}^i_{-n}$: modified parameters, $\boldsymbol{s}^i_n$: The sample when ith iteration, x_n: nth data.

Output: $\hat{\boldsymbol{\alpha}}$, $\hat{a}$, $\hat{b}$: updated parameter.

for $k = 1$ to K **do**
    Update $\hat{\alpha}_k$, $\hat{a}_k$, $\hat{b}_k$ as

$$\hat{\alpha}_k \leftarrow \hat{\alpha}^i_{-n,k} + s^i_{n,k},$$
$$\hat{a}_k \leftarrow \hat{a}^i_{-n,k} + s^i_{n,k} x_n,$$
$$\hat{b}_k \leftarrow \hat{b}^i_{-n,k} + s^i_{n,k}.$$

end for
return $\hat{\boldsymbol{\alpha}}$, $\hat{a}$, $\hat{b}$

Algorithm 3 Collapsed Gibbs sampling.

Input:

Input: MAXITER: the number of iteration, N: the number of data, K: the number of clusters.

Output: $\boldsymbol{S}_{\text{sample}}$: The set of the samples of hidden variables.

Initialisation :

Set the initial value $\boldsymbol{s}_1^0, \ldots, \boldsymbol{s}_N^0$, $\boldsymbol{S}^0 \leftarrow \{\boldsymbol{s}_1^0, \ldots, \boldsymbol{s}_N^0\}$, $\boldsymbol{S}_{\text{sample}} \leftarrow \{\boldsymbol{S}^0\}$.

Calculate $\hat{\boldsymbol{\alpha}}$, $\hat{a}$, and $\hat{b}$ using (23) – (25).

LOOP Process

for $i = 1$ to MAXITER **do**

  for $n = 1$ to N **do**

    $\hat{\boldsymbol{\alpha}}^i_{-n},\ \hat{a}^i_{-n},\ \hat{b}^i_{-n} \leftarrow \mathrm{Erase}(\hat{\boldsymbol{\alpha}}, \hat{a}, \hat{b}, \boldsymbol{s}_n^{i-1}, x_n)$.

    $\boldsymbol{\xi}^i_{-n} \leftarrow \mathbf{0} \in \mathbb{R}^K$.

    for $k = 1$ to K **do**

      Update $\xi^i_{-n,k}$ using the right hand side of (19).

    end for

    Normalize $\boldsymbol{\xi}^i_{-n}$.

    Sample $\boldsymbol{s}_n^i \sim \mathrm{Cat}(\boldsymbol{s}_n | \boldsymbol{\xi}^i_{-n})$.

    $\hat{\boldsymbol{\alpha}},\ \hat{a},\ \hat{b} \leftarrow \mathrm{Add}(\hat{\boldsymbol{\alpha}}, \hat{a}, \hat{b}, \boldsymbol{\alpha}^{\hat{i}}_{-n}, \hat{a}^i_{-n}, \hat{b}^i_{-n}, \boldsymbol{s}_n^i, x_n)$.

  end for

  $\boldsymbol{S}^i \leftarrow \{\boldsymbol{s}_1^i, \ldots, \boldsymbol{s}_N^i\}$, $\boldsymbol{S}_{\text{sample}} \leftarrow \boldsymbol{S}_{\text{sample}} \cup \{\boldsymbol{S}^{\boldsymbol{i}}\}$.

end for

return $\boldsymbol{S}_{\text{sample}}$.

参考文献

[1] A. Suyama, ベイズ推論による機械学習 入門, 講談社, 2017.