

ベイズ本 4.3.3 節

ここでは、例として、ポアソン混合分布に対する変分推論アルゴリズムについてまとめる．導出は手書きの資料に載せているので、ここでは省略する．

■Notation

以下の記号を用いる．

- $\mathbf{X} = \{x_1, \dots, x_N\}$: 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
- $\mathbf{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$: 潜在変数の集合
 - \mathbf{s}_n : k 次元ベクトル
 - $s_{n,k} = 1 \iff k$ 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:

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

- \mathbf{s}_n をサンプルするための分布:

$$p(\mathbf{s}_n | \boldsymbol{\pi}) = \text{Cat}(\mathbf{s}_n | \boldsymbol{\pi}). \quad (2)$$

- 混合分布における条件付き分布:

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

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

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

- a, b : ハイパーパラメータ.

- 混合比率 π に対する事前分布:

$$p(\pi) = \text{Dir}(\pi|\alpha). \quad (5)$$

– α : 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

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

と表される [1].

■変分推論

潜在変数とパラメータを分けることで, 事後分布を近似する. つまり,

$$p(\mathbf{S}, \boldsymbol{\lambda}, \pi|\mathbf{X}) \approx q(\mathbf{S})q(\boldsymbol{\lambda}, \pi) \quad (7)$$

として近似する.

まず, 潜在変数 \mathbf{S} について, 各点ごとの独立な分布 $q(\mathbf{s}_1), \dots, q(\mathbf{s}_N)$ に分解されることが示される. 各 \mathbf{s}_n はカテゴリ分布に従うことが示される.

$$q(\mathbf{s}_n) = \text{Cat}(\mathbf{s}_n|\boldsymbol{\eta}_n), \quad (8)$$

ただし,

$$\eta_{n,k} \propto \exp(x_n \langle \ln \lambda_k \rangle - \langle \lambda_k \rangle + \langle \ln \pi_k \rangle), \quad \text{s.t.} \quad \sum_{k=1}^K \eta_{n,k} = 1 \quad (9)$$

である. この更新式には $\boldsymbol{\lambda}$ や π に対する期待値計算が必要である.

次に, パラメータに対する近似分布の更新式を求める. ギブスサンプリングのときと同様に, $\boldsymbol{\lambda}$ と π が独立であることが示される. $\boldsymbol{\lambda}$ の分布は, 各 k に対して

$$\lambda_k \sim \text{Gam}(\lambda_k|\hat{a}_k, \hat{b}_k), \quad (10)$$

ただし,

$$\hat{a}_k = \sum_{n=1}^N \langle s_{n,k} \rangle x_n + a \quad (11)$$

$$\hat{b}_k = \sum_{n=1}^N \langle s_{n,k} \rangle + b \quad (12)$$

である [1].

また, π の分布は

$$\pi \sim \text{Dir}(\pi|\hat{\alpha}), \quad (13)$$

ただし,

$$\hat{\alpha}_k = \sum_{n=1}^N \langle s_{n,k} \rangle + \alpha_k \quad (14)$$

である.

(10), (13) の更新には期待値 $\langle s_{n,k} \rangle$ の計算が必要である. (8) より, $q(\mathbf{s}_n)$ はカテゴリ分布に従うので,

$$\langle s_{n,k} \rangle = \eta_{n,k} \quad (15)$$

である.

また, (8) の更新に必要な λ と π の期待値は, (10), (13) より,

$$\langle \lambda_k \rangle = \frac{\hat{\alpha}_k}{\hat{b}_k} \quad (16)$$

$$\langle \ln \lambda_k \rangle = \psi(\hat{\alpha}_k) - \ln \hat{b}_k \quad (17)$$

$$\langle \ln \pi_k \rangle = \psi(\hat{\alpha}_k) - \psi\left(\sum_{i=1}^K \hat{\alpha}_i\right) \quad (18)$$

である. ただし, $\psi(\cdot)$ は, ディガンマ関数である. 変分推論の更新式における期待値計算の部分サンプル値に置き換えるとギブスサンプリングが得られる.

以上を用いると, ポアソン混合モデルのための変分推論は Algorithm 1 で与えられる.

Algorithm 1 Gibbs Sampling for Poisson mixture model.

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

Output: $q(\mathbf{S}), q(\lambda), q(\pi)$: Approximate distribution.

Initialisation : Set initial $q(\lambda)$ and $q(\pi)$.

LOOP Process

```

1: for  $i = 1$  to MAXITER do
2:   for  $n = 1$  to  $N$  do
3:     Update  $q(\mathbf{s}_n)$  using (8).
4:   end for
5:   for  $k = 1$  to  $K$  do
6:     Update  $q(\lambda_k)$  using (10).
7:   end for
8:   Update  $q(\pi)$  using (13).
9: end for
10: return  $q(\mathbf{S}), q(\lambda), q(\pi)$ 

```

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.