Ch.4 混合モデルと近似推論.

§4.1 混合モデルと事後分布の推論

・現実世界の複雑なデータ生成過程を記述するのに，単純なモデルでは表現しきれない.

→ 混合モデル：複数のモデルを組みあわせて使う.

4.1.1 混合モデルを使う理由

・データのクラスタやトレンドが複数個あるとき，一つのモデルでは表現できない

・他の確率モデルと組みあわせやすい.

4.1.2 混合モデルのデータ生成過程

Kクラスタからなるデータ $X = \{x_1, \cdots, x_N\}$ の生成過程を，次のモデル(生成モデル)で記述してみる：

1. 各クラスタの混合比率

$$\pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_K \end{pmatrix} \qquad \pi_k \in (0,1), \quad \sum_{k=1}^{K} \pi_k = 1$$

が事前分布 $p(\pi)$ から生成.

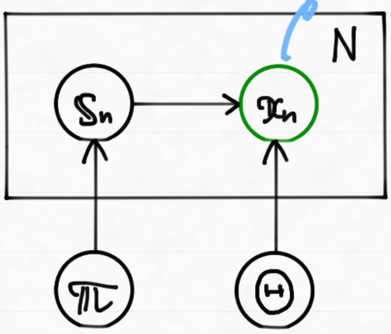
2. 各クラスタ $k=1,\dots,K$ に対する観測モデル $p(x|\theta_k)$ の
パラメータ θ_k が、事前分布 $p(\theta_k)$ から生成.

3. 各 $n=1,\dots,N$ に対して, x_n に対応するクラスタの割り当て
s_n (潜在変数) が 比率 π を利用して $p(s_n|\pi)$ から生成.
↳ この変数は直接観測することができない.

4. 各 $n=1,\dots,N$ に対して, s_n で選ばれた k番目の
モデル $p(x|\theta_k)$ から x_n が生成.

・グラフィカルモデル:

観測できるのは x_n だけ



$\Theta := \{\theta_1, \dots, \theta_K\}$.

* 上の生成モデルをより具体的に書いてみる.

◦ 4. について.

混合Gaussモデル

観測モデル $p(x|\theta_k)$ を Gauss分布にしてみる:

$$x_n \sim p(x_n|\theta_k) := \mathcal{N}(x_n|\mu_k, \Sigma_k) \quad (k=1,\dots,K).$$

$(\theta_k := \{\mu_k, \Sigma_k\}.)$

- 3. について.

s_n に 1 of K 表現を使う. これはカテゴリ分布からのサンプリングでOK:

$$s_n = \begin{pmatrix} s_{n,1} \\ \vdots \\ s_{n,K} \end{pmatrix} \sim p(s_n | \pi) := \mathrm{Cat}(s_n | \pi)$$

(π : 混合比率パラメータ)

→ 3と4をあわせて,

$$x_n \sim p(x_n | s_n, \Theta) = \prod_{k=1}^{K} p(x_n | \theta_k)^{s_{n,k}}$$

$$(\Theta = \{\theta_1, \cdots, \theta_K\})$$

と書ける.

- 2. について.

θ_k の事前分布 $p(\theta_k)$ を定める: $\theta_k \sim p(\theta_k)$.

4. で設定したモデル $p(x_n | \theta_k)$ の共役事前分布をとるのが良い.

ここでは, Gauss分布を設定しているので, Gauss-Wishart分布:

$$p(\theta_k) := \mathrm{NW}(\mu_k, \Sigma_k | m, \beta, \nu, W).$$

- 1 について.

π の事前分布 $p(\pi)$ を定める: $\pi \sim p(\pi)$.

π：カテゴリ分布のパラメータなので，この共役事前分布である

K次元Dirichlet分布に設定．

$$p(\pi) := \mathrm{Dir}(\pi|\alpha) \quad (\alpha：ハイパーパラメータ)$$

- 以上から，$S := \{s_1, \cdots, s_N\}$ として，モデルの全体像は次のとおり：

$$
\begin{aligned}
&p(X, S, \Theta, \pi) \\
&= p(X|S, \Theta, \pi)p(S|\Theta, \pi)p(\Theta|\pi)p(\pi) \\
&= p(X|S, \Theta)p(S|\pi)p(\Theta)p(\pi) \\
&= \left(\prod_{n=1}^{N} p(x_n|s_n, \Theta)p(s_n|\pi)\right)\left(\prod_{k=1}^{K} p(\theta_k)\right)p(\pi).
\end{aligned}
$$

独立性

4.1.3 混合モデルの事後分布．

- X：given．

S, Θ, π の事後分布

$$p(S, \Theta, \pi|X) = \frac{p(X, S, \Theta, \pi)}{p(X)},$$

あるいは，データ X の所属クラスタ S の推定を行うために

$$p(S|X) = \iint p(S, \Theta, \pi|X)\, d\Theta\, d\pi$$

を計算したい．

→ 混合モデルで，これらの計算は効率的にできない！

∵ 事後分布は S, Θ, π が複雑に絡みあった分布で，

$p(\mathcal{X})$ の計算が困難：

周辺化

$$p(\mathcal{X}) = \sum_{S} \iint p(\mathcal{X}, S, \Theta, \pi)\, d\Theta\, d\pi$$
$$= \sum_{S} p(\mathcal{X}, S).$$

$\sum_{S}$ は，$S = \{ \mathbb{s}_1, \cdots, \mathbb{s}_N \}$ としてとりうる組合せ全てについての和．

$\mathbb{s}_k$ のパターンは K 通り．N コあるので S の組合せは K^N 通り．

機械学習で用いるデータは N：大 であることが多いので

この計算は現実的な時間で終わらない．

§4.2 確率分布の近似手法.

- 近似推論 (approximate inference):

直接 $p(x)$ の計算をするのは諦めて, 簡単な計算で済ませる枠組み.

4.2.1 Gibbsサンプリング.

$\mathbf{z} = (z_1, \cdots, z_M)^T$. $p(\mathbf{z})$ に関して, 期待値計算などを行いたい.

→ $\mathbf{z}$ の実現値を複数個サンプルして計算する:

$$\mathbf{z}^{(i)} \sim p(\mathbf{z}) \quad (i = 1, 2, \cdots)$$

- 分布 $p(\mathbf{z})$ が複雑だと, 全変数を同時にサンプルするのは困難.

→ Gibbsサンプリング (MCMCの手法の一種)

i個目の変数 $\mathbf{z}^{(i)}$ の各成分 $z_m^{(i)}$ を次のように一つずつサンプリング:

for $i = 1, 2, \cdots, N_{max}$ do

$z_2^{(0)}, \cdots, z_M^{(0)}$ についてはランダムに与えておく.

$$z_1^{(i)} \sim p(z_1 \mid z_2^{(i-1)}, z_3^{(i-1)}, \cdots, z_M^{(i-1)});$$

$$z_2^{(i)} \sim p(z_2 \mid z_1^{(i)}, z_3^{(i-1)}, \cdots, z_M^{(i-1)});$$

$z_1^{(i-1)}$ ではない!

$$\vdots$$

$$z_M^{(i)} \sim p(z_M \mid z_1^{(i)}, z_2^{(i)}, \cdots, z_{M-1}^{(i)})$$

end for.

・Gibbsサンプリングのアイデア：

z_m をサンプルするために，既にサンプルされた値で分布を条件付けして，

よりサンプルしやすい分布を得る．

・Gibbsサンプリングによりサンプリングされた z_m は，サンプル数が十分なら

真の事後分布からサンプルされたものとみなせる（ことが知られている）．

・問題点：推論に必要なサンプル数が明確でない．

→サンプル点がどれだけうまく真の分布を表しているのかを知ることは

できない（可視化が一般にできないため）．

＊ Gibbsサンプリングの変種．

。ブロッキングGibbsサンプリング

$\mathbf{z} = (\mathbf{z}_1^T, \cdots, \mathbf{z}_K^T)^T$ とブロックに分かれているとする．

各 $\mathbf{z}_k$ の全成分は同時にサンプリングできるとする．

このとき，i個目の変数 $\mathbf{z}^{(i)}$ の各ブロック $\mathbf{z}_k^{(i)}$ を，

通常のGibbsサンプリングのように，一つずつサンプリングする．

擬似コードで表すと，次のとおり：

$$
\left[\begin{array}{l}
\text{for } i = 1, 2, \cdots, N_{max} \text{ do} \\
\quad \mathbf{z}_1^{(i)} \sim p(\mathbf{z}_1 \mid \mathbf{z}_2^{(i-1)}, \mathbf{z}_3^{(i-1)}, \cdots, \mathbf{z}_K^{(i-1)}); \\
\quad \mathbf{z}_2^{(i)} \sim p(\mathbf{z}_2 \mid \mathbf{z}_1^{(i)}, \mathbf{z}_3^{(i-1)}, \cdots, \mathbf{z}_K^{(i-1)}); \\
\qquad\qquad \vdots \\
\quad \mathbf{z}_K^{(i)} \sim p(\mathbf{z}_2 \mid \mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}, \cdots, \mathbf{z}_{K-1}^{(i)}) \\
\text{end for.}
\end{array}\right.
$$

✓ ブロックごとの変数の同時サンプリング

→ より真の事後分布に近いサンプルが得られる。

△ サンプリングに使う各条件付き分布が、計算コスト的に十分簡単にサンプリングできる必要がある。

- 崩壊型 Gibbsサンプリング

$\mathbf{z} = (\mathbf{z}_1^T, \mathbf{z}_2^T)^T$ として、周辺化によりモデルから $\mathbf{z}_2$ を滅ぼす：

$$
p(\mathbf{z}_1) = \int p(\mathbf{z})\, d\mathbf{z}_2. \qquad - (*)
$$

$p(\mathbf{z}_1)$ について、通常の Gibbsサンプリングを行う。

✓ サンプルする変数が少なく、高速にサンプリングできる。

△ (*) の計算が解析的にできなければならない。

△ 滅ぼした $\mathbf{z}_2$ のサンプリングが容易に行える必要がある。

4.2.2 変分推論

- **変分推論** (variational inference):

最適化問題を解いて、未知の確率分布の近似的な表現を得る.

- $p(\mathbf{z})$: 近似表現を得たい複雑な分布.

→ p を、より簡単な分布 $q(\mathbf{z})$ で近似する.

→ $q_{opt}(\mathbf{z}) := \underset{q}{\operatorname{argmin}} \underline{KL[q(\mathbf{z}) \| p(\mathbf{z})]}$ —(*)

「分布 q が p とどれだけ似ているか」

となるように $q_{opt}(\mathbf{z})$ を選ぶとよい?

- (*) を制約なしで解くと、$q_{opt} = p$ となり意味がない.

→ q の表現能力を限定し、このもとで最適化をする.

- **平均場近似** (mean-field approximation) に基づく変分推論.

q に独立性を仮定する, i.e., 次式で p を近似:

$$p(\mathbf{z}) \approx \prod_{m=1}^{M} q_m(z_m)$$

→ 各 q_m を、KL divergence が小さくなるように

一つ一つ更新していく.

- q_m $(m = 2, \cdots, M)$: given とする. このとき、最適な q_1 は,

$$q_1^*(z_1) := \underset{q_1}{\operatorname{argmin}} KL[q_1(z_1) \prod_{m \neq 1} q_m(z_m) \| p(\mathbf{z})]$$

で求まる. KL divergence を計算してみる.

$\mathbb{E}_{q_1,\cdots,q_M}[\cdot] =: \mathbb{E}_{1,\cdots,M}[\cdot]$ などと書く.

$$
\begin{aligned}
&\mathrm{KL}\left[q_1(z_1)\prod_{m\neq 1} q_m(z_m) \,\Big\|\, p(\mathbf{z})\right] \\
&= -\mathbb{E}_{1,\cdots,M}\left[\log \frac{p(\mathbf{z})}{q_1(z_1)\prod_{m\neq 1} q_m(z_m)}\right] \\
&\quad\text{期待値を分解. (定義を考えればあたりまえ)} \\
&= -\mathbb{E}_1\left[\mathbb{E}_{2,\cdots,M}\left[\log \frac{p(\mathbf{z})}{q_1(z_1)\prod_{m\neq 1} q_m(z_m)}\right]\right] \\
&\quad\text{log の計算} \\
&= -\mathbb{E}_1\left[\mathbb{E}_{2,\cdots,M}\left[\log p(\mathbf{z}) - \log q_1(z_1) - \underbrace{\sum_{m\neq 1}\log q_m(z_m)}_{\text{const.}}\right]\right] \\
&\quad\text{期待値の線形性} \\
&= -\mathbb{E}_1\left[\mathbb{E}_{2,\cdots,M}\left[\log p(\mathbf{z})\right] - \underbrace{\mathbb{E}_{2,\cdots,M}\left[\log q_1(z_1)\right]}_{q_1 \text{ は } q_2 \sim q_M \text{ に依らない}}\right] + \text{const.} \\
&= -\mathbb{E}_1\left[\mathbb{E}_{2,\cdots,M}\left[\log p(\mathbf{z})\right] - \log q_1(z_1)\right] + \text{const.} \\
&\quad \log \circ \exp = \mathrm{id}. \\
&= -\mathbb{E}_1\left[\log\left(\exp\left(\mathbb{E}_{2,\cdots,M}\left[\log p(\mathbf{z})\right]\right)\right) - \log q_1(z_1)\right] + \text{const.} \\
&= -\mathbb{E}_1\left[\log \frac{\exp\left(\mathbb{E}_{2,\cdots,M}\left[\log p(\mathbf{z})\right]\right)}{q_1(z_1)}\right] + \text{const.} \\
&= \mathrm{KL}\left[q_1(z_1) \,\|\, \exp\left(\mathbb{E}_{2,\cdots,M}\left[\log p(\mathbf{z})\right]\right)\right] + \text{const.}
\end{aligned}
$$

これは $q_1(z_1) = \exp\left(\mathbb{E}_{2,\cdots,M}\left[\log p(\mathbf{z})\right]\right)$ のとき最小.

$\therefore$ $q_1^*(z_1)$ は次式をみたす:

要るんか…?

$$
\log q_1^*(z_1) = \mathbb{E}_{2,\cdots,M}\left[\log p(\mathbf{z})\right] + \text{const.}
$$

- 残りの分布 $q^*_m(z_m)$ $(m=2,\cdots,M)$ についても同様.

$\mathbb{Z}_{\setminus m} := (z_1, \cdots, z_{m-1}, z_{m+1}, \cdots, z_M)^T$ $(m=1,\cdots,M)$ として,

$\mathbb{E}_{q(\mathbb{Z}_{\setminus m})}[\cdot] =: \mathbb{E}_{\setminus m}[\cdot]$ と書くことにする.

変分推論のアルゴリズムは次のとおり:

```
for m = 2, ..., M do
    initialize q_m(z_m)
end for;
for i = 1, ..., N_max do
    for m = 1, ..., M do
        log q_m(z_m) := E_\m[ log p(Z) ] + const.
    end for
end for.
```

$$
\left[
\begin{array}{l}
\text{for } m = 2,\cdots,M \text{ do} \\
\quad \text{initialize } q_m(z_m) \\
\text{end for;} \\
\text{for } i = 1,\cdots,N_{max} \text{ do} \\
\quad \text{for } m = 1,\cdots,M \text{ do} \\
\quad\quad \log q_m(z_m) := \mathbb{E}_{\setminus m}[\log p(\mathbb{Z})] + \text{const.} \\
\quad \text{end for} \\
\text{end for.}
\end{array}
\right.
$$

- 繰り返しの終了条件としては,上記のように N_{max} を設定する以外にも,計算時間を制限する,ELBO (evidence lower bound) などの評価値の収束をもって終了する,などの方法がある.

  - ELBOについては後述.

未観測変数

- 観測データ$\mathcal{D}$が与えられたときの，確率モデル$p(\mathcal{D}, \mathbf{z})$の事後分布$p(\mathbf{z}|\mathcal{D})$の近似は，先の計算で$p(\mathbf{z})$の部分を$p(\mathbf{z}|\mathcal{D})$とすればよく，

$$\log q_{\eta_m}(\mathbf{z}_m) = \mathbb{E}_{\backslash m}\left[\log p(\mathbf{z}|\mathcal{D})\right] + \text{const.}$$

$p(\mathcal{D})$はconst.に

$$= \mathbb{E}_{\backslash m}\left[\log p(\mathcal{D}, \mathbf{z})\right] + \text{const.}$$

✓ Gibbsサンプリングより高速．

△ 強い相関をもつ分布ではうまく近似できない．

- 構造化変分推論 (structured variational inference)

$$p(\mathbf{z}) \approx \prod_{k=1}^{K} q_{\eta_k}(\mathbf{z}_k)$$

のように，大きなまとまりで近似する．

✓ 変数間の相関をとらえることができ，よい近似になる．

△ メモリコスト，計算コストの増加．

cf. ELBOについて．

- 変分推論：周辺尤度の下限の最大化アルゴリズムとも考えることができる．

・観測データ$\mathcal{D}$，未観測変数 $\mathbf{z} = (z_1, \cdots, z_M)^T$ について，

確率モデル $p(\mathcal{D}, \mathbf{z})$ を考え，$\mathbf{z}$ についての分布 $q(\mathbf{z})$ を

仮定する．

Thm. (Jensenの不等式)

X : r.v. $X \sim p(x)$, $f : \mathbb{R} \to \mathbb{R}$: 凸関数.

このとき， $f(\mathbb{E}_p[X]) \leq \mathbb{E}_p[f(X)]$. □

pf. f : 凸関数 より. ${}^\forall x_0 \in \mathbb{R}$,

「接線よりもグラフが上にくる」

$$f(x) \geq f(x_0) + f'(x_0)(x - x_0), \quad {}^\forall x \in \mathbb{R}.$$

x を X として期待値をとると，

$$\mathbb{E}_p[f(X)] \geq f(x_0) + f'(x_0)(\mathbb{E}_p[X] - x_0).$$

$x_0 = \mathbb{E}_p[X]$ とすることで

$$f(\mathbb{E}_p[X]) \leq \mathbb{E}_p[f(X)].$$ ■

[Claim] $-\log$ は 凸関数.

pf. 分かる. ■

Cor. X : r.v. $X \sim p(x)$,

このとき， $\log \mathbb{E}_p[X] \geq \mathbb{E}_p[\log X]$. □

この不等式を用いると，モデルの周辺尤度 $p(\mathcal{D})$ の対数について，

下限を計算できる：

$$
\begin{aligned}
&\log p(\mathcal{D}) \\
&\overset{\text{周辺化}}{=} \log\left(\int p(\mathcal{D}, \mathbf{z})\,d\mathbf{z}\right) \\
&= \log\left(\int \frac{p(\mathcal{D}, \mathbf{z})}{q(\mathbf{z})}\,q(\mathbf{z})\,d\mathbf{z}\right) \\
&\overset{\text{期待値の定義}}{=} \log \mathbb{E}_q\left[\frac{p(\mathcal{D}, \mathbf{z})}{q(\mathbf{z})}\right] \\
&\overset{\text{先のCor.}}{\geq} \mathbb{E}_q\left[\log \frac{p(\mathcal{D}, \mathbf{z})}{q(\mathbf{z})}\right] \\
&\overset{\text{期待値の定義}}{=} \int q(\mathbf{z}) \log \frac{p(\mathcal{D}, \mathbf{z})}{q(\mathbf{z})}\,d\mathbf{z} \\
&=: \mathcal{L}[q(\mathbf{z})]
\end{aligned}
$$

<u>Def.</u> (ELBO)

対数周辺尤度の下限 $\mathcal{L}[q(\mathbf{z})]$ を分布 $q(\mathbf{z})$ に対する

ELBO という． □

→ $\mathcal{L}[q(\mathbf{z})]$ をなるべく大きくする $q(\mathbf{z})$ を求めることで，$\log p(\mathcal{D})$ を

近似的に計算できる．

$\log p(\mathcal{D})$が大きくなるqを選んでいるわけではない.
$\mathcal{L}[q(\mathbf{z})]$が$\log p(\mathcal{D})$を一番よく近似する$q$を選んでいる.

$\to$ $q^*(\mathbf{z}) := \underset{q}{\operatorname{argmax}}\, \mathcal{L}[q(\mathbf{z})]$ により q を選択する.

<u>Prop.</u> $\log p(\mathcal{D}) - \mathcal{L}[q(\mathbf{z})] = \mathrm{KL}[q(\mathbf{z}) \| p(\mathbf{z}|\mathcal{D})]$. $\square$

pf.

$$
\begin{aligned}
&\log p(\mathcal{D}) - \mathcal{L}[q(\mathbf{z})] \\
&\quad \downarrow \text{by def.} \\
&= \log p(\mathcal{D}) - \int q(\mathbf{z}) \log \frac{p(\mathcal{D}, \mathbf{z})}{q(\mathbf{z})} d\mathbf{z} \\
&\quad \downarrow p(\mathcal{D}, \mathbf{z}) = p(\mathbf{z}|\mathcal{D})p(\mathcal{D}) \\
&= \log p(\mathcal{D}) - \int q(\mathbf{z}) \log \frac{p(\mathbf{z}|\mathcal{D})p(\mathcal{D})}{q(\mathbf{z})} d\mathbf{z} \\
&\quad \downarrow \log\text{の計算} \\
&= \log p(\mathcal{D}) - \int q(\mathbf{z}) \left(\log \frac{p(\mathbf{z}|\mathcal{D})}{q(\mathbf{z})} + \log p(\mathcal{D}) \right) d\mathbf{z} \\
&= \cancel{\log p(\mathcal{D})} - \underbrace{\int q(\mathbf{z}) \log \frac{p(\mathbf{z}|\mathcal{D})}{q(\mathbf{z})} d\mathbf{z}}_{= \mathrm{KL}[q(\mathbf{z}) \| p(\mathbf{z}|\mathcal{D})]} - \cancel{\log p(\mathcal{D})} \underbrace{\int q(\mathbf{z}) d\mathbf{z}}_{=1} \\
&= \mathrm{KL}[q(\mathbf{z}) \| p(\mathbf{z}|\mathcal{D})]. \quad \blacksquare
\end{aligned}
$$

・ $\log p(\mathcal{D})$ は $\mathcal{D}$: given なら一定値.

以上より, $\mathcal{L}[q(\mathbf{z})]$ が最大となる q を選ぶのと,

$\mathrm{KL}[q(\mathbf{z}) \| p(\mathbf{z}|\mathcal{D})]$ が最小となる q を選ぶのは等価.

§4.3 Poisson混合モデルにおける推論.

4.3.1 Poisson混合モデル

- 多峰性の離散非負データの学習に利用できる.

- $\boldsymbol{\lambda} := \{\lambda_1, \cdots, \lambda_K\}$: パラメータ.

  クラスタkに対する観測モデル: $p(x_n | \lambda_k) := \mathrm{Poi}(x_n | \lambda_k)$.

  → 混合分布における条件付き分布は,

$$p(x_n | s_n, \boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathrm{Poi}(x_n | \lambda_k)^{s_{n,k}}$$

  s_nによって, K コのPoisson分布の一つが指定される.

- $\boldsymbol{\lambda}$に対して, 事前分布を設定する.

  Poisson分布の共役事前分布 → ガンマ分布.

  クラスターごとに a_k, b_k とパラメータを加えても良い.

$$p(\lambda_k) := \mathrm{Gam}(\lambda_k | a, b), \quad a, b: \text{ハイパーパラメータ}.$$

- Sに対する事前分布は, $p(s_n | \boldsymbol{\pi}) := \mathrm{Cat}(s_n | \boldsymbol{\pi})$,

  $\boldsymbol{\pi}$に対する事前分布は, $p(\boldsymbol{\pi}) := \mathrm{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha})$

  ($\boldsymbol{\alpha}$: ハイパーパラメータ)

  と設定することにする.

Remark モデルを構築した後は、データの表現能力や挙動について、思い違いやミスがないかを確認するため、いくつかのハイパーパラメータを使ってサンプルを得てみて、視覚的に確認してみるとよい。

問題によっては、事前知識を与えることができるかも。

4.3.2 Gibbsサンプリング

- $\mathcal{X} = \{x_1, \cdots, x_N\}$: given.　このとき、条件付き分布は、
  $p(\mathcal{S}, \pi, \lambda | \mathcal{X})$. この分布から $\mathcal{S}, \pi, \lambda$ をサンプルするアルゴリズムを、Gibbsサンプリングにより導出。

- 混合分布では、潜在変数と分布のパラメータを分けてサンプリングすると、十分に簡単な分布が得られる(らしい)。

→ 次のようにサンプリングすることにする：

$$\mathcal{S}^{(i)} \sim p(\mathcal{S} | \pi^{(i-1)}, \lambda^{(i-1)}, \mathcal{X}) \quad - (1)$$

$$\pi^{(i)}, \lambda^{(i)} \sim p(\pi, \lambda | \mathcal{S}^{(i)}, \mathcal{X}) \quad - (2)$$

- (1)について、$\mathcal{S}^{(i)} = \{ s_1^{(i)}, \cdots, s_N^{(i)} \}$ をサンプルするための条件付き分布 $p(\mathcal{S} | \underline{\pi^{(i-1)}, \lambda^{(i-1)}}, \mathcal{X})$ を求める。

given として扱うのがポイント。

$$p(\mathcal{S}|\pi^{(i-1)},\lambda^{(i-1)},\mathcal{X})$$

$$\propto p(\mathcal{S},\pi^{(i-1)},\lambda^{(i-1)},\mathcal{X})$$

乗法則

$$= p(\mathcal{X}|\mathcal{S},\pi^{(i-1)},\lambda^{(i-1)})\,p(\mathcal{S}|\pi^{(i-1)},\lambda^{(i-1)})\,\underbrace{p(\pi^{(i-1)},\lambda^{(i-1)})}_{\text{const.}}$$

独立性

$$\propto p(\mathcal{X}|\mathcal{S},\lambda^{(i-1)})\,p(\mathcal{S}|\pi^{(i-1)})$$

$$= \left(\prod_{n=1}^{N} p(x_n|\mathcal{S},\lambda^{(i-1)})\right)\left(\prod_{n=1}^{N} p(s_n|\pi^{(i-1)})\right)$$

x_n には s_n の情報しか使わない

$$= \left(\prod_{n=1}^{N} p(x_n|s_n,\lambda^{(i-1)})\right)\left(\prod_{n=1}^{N} p(s_n|\pi^{(i-1)})\right)$$

$$= \prod_{n=1}^{N} p(x_n|s_n,\lambda^{(i-1)})\,p(s_n|\pi^{(i-1)})$$

→ 各 $s_n^{(i)}$ が独立にサンプリングできる（条件付き独立性）.

- $s_n^{(i)}$ をサンプルするための確率分布を計算. 対数をとってやっていく.

$$p(x_n|s_n,\lambda^{(i-1)}) = \prod_{k=1}^{K} \mathrm{Poi}(x_n|\lambda_k^{(i-1)})^{s_{n,k}}$$ なので,

$$\log p(x_n|s_n,\lambda^{(i-1)})$$

$$= \sum_{k=1}^{K} s_{n,k}\log \mathrm{Poi}(x_n|\lambda_k^{(i-1)})$$

$$= \sum_{k=1}^{K} s_{n,k}\left(x_n\log\lambda_k^{(i-1)} - \lambda_k^{(i-1)}\right) + \text{const.}$$

また,

$$\log p(s_n|\pi^{(i-1)}) = \log \mathrm{Cat}(s_n|\pi^{(i-1)}) = \sum_{k=1}^{K} s_{n,k}\log\pi_k .$$

二つあわせて,

$$\log p(x_n|s_n,\lambda^{(i-1)})p(s_n|\pi^{(i-1)})$$

$$= \log p(x_n|s_n,\lambda^{(i-1)}) + \log p(s_n|\pi^{(i-1)})$$

$$= \sum_{k=1}^{K} s_{n,k}\left(x_n \log \lambda_k^{(i-1)} - \lambda_k^{(i-1)}\right) + \sum_{k=1}^{K} s_{n,k} \log \pi_k^{(i-1)} + \text{const.}$$

$$= \sum_{k=1}^{K} s_{n,k}\left(x_n \log \lambda_k^{(i-1)} - \lambda_k^{(i-1)} + \log \pi_k^{(i-1)}\right) + \text{const.}$$

制約 $\sum_{k=1}^{K} s_{n,k} = 1$ を考慮すると，これは s_n に関する

カテゴリ分布（の定数倍）の対数をとったもの．

$$\therefore\ s_n^{(i)} \sim \mathrm{Cat}(s_n|\eta_n^{(i)}),$$

$$\eta_{n,k}^{(i)} \propto \exp\left(x_n \log \lambda_k^{(i-1)} - \lambda_k^{(i-1)} + \log \pi_k^{(i-1)}\right),\ \sum_{k=1}^{K} \eta_{n,k}^{(i)} = 1.$$

- (2)について，$\pi^{(i)}, \lambda^{(i)} = \{\lambda_1^{(i)}, \cdots, \lambda_K^{(i)}\}$ をサンプルするための

条件付き分布 $p(\pi, \lambda | \underline{S^{(i)}}, X)$ を求める．

↳ given とみる．

先の式変形と同様にして，

$$p(\pi, \lambda | S^{(i)}, X)$$

$$\propto p(\pi, \lambda, S^{(i)}, X)$$

乗法則

$$= p(X|\pi,\lambda,S^{(i)})p(S^{(i)}|\pi,\lambda)p(\pi|\lambda)p(\lambda)$$

独立性

$$= p(X|S^{(i)},\lambda)p(S^{(i)}|\pi)p(\pi)p(\lambda)$$

$$= \underbrace{p(X|S^{(i)},\lambda)p(\lambda)}_{\lambda\text{に関する項}}\underbrace{p(S^{(i)}|\pi)p(\pi)}_{\pi\text{に関する項}}.$$

$\rightarrow \mathcal{S}^{(i)}, \mathcal{X}$: given のもとで, π と λ は条件付き独立.

別々にサンプリングを行える.

・λ の分布 $p(\lambda | \mathcal{S}^{(i)}, \mathcal{X}) \propto p(\mathcal{X} | \mathcal{S}^{(i)}, \lambda) p(\lambda)$ について.

対数をとって計算していく.

$$
\begin{aligned}
&\log p(\lambda | \mathcal{S}^{(i)}, \mathcal{X}) \\
&= \log p(\mathcal{X} | \mathcal{S}^{(i)}, \lambda) p(\lambda) + \mathrm{const.} \\
&= \log p(\mathcal{X} | \mathcal{S}^{(i)}, \lambda) + \log p(\lambda) + \mathrm{const.} \qquad p(\lambda) = \prod_{k=1}^{K} p(\lambda_k). \\
&= \sum_{n=1}^{N} \log p(x_n | s_n^{(i)}, \lambda) + \sum_{k=1}^{K} \log p(\lambda_k) + \mathrm{const.} \\
&= \sum_{n=1}^{N} \log \left(\prod_{k=1}^{K} \mathrm{Poi}(x_n | \lambda_k)^{s_{n,k}^{(i)}} \right) + \sum_{k=1}^{K} \log \mathrm{Gam}(\lambda_k | a, b) + \mathrm{const.} \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} s_{n,k}^{(i)} \log \mathrm{Poi}(x_n | \lambda_k) + \sum_{k=1}^{K} \log \mathrm{Gam}(\lambda_k | a, b) + \mathrm{const.} \\
&= \sum_{k=1}^{K} \left(\sum_{n=1}^{N} s_{n,k}^{(i)} (x_n \log \lambda_k - \lambda_k) + (a-1) \log \lambda_k - b\lambda_k \right) + \mathrm{const.} \\
&= \sum_{k=1}^{K} \underbrace{\left(\left(\sum_{n=1}^{N} s_{n,k}^{(i)} x_n + a - 1 \right) \log \lambda_k - \left(\sum_{n=1}^{N} s_{n,k}^{(i)} + b \right) \lambda_k \right)}_{\text{ガンマ分布のpdfの対数をとった形.}} + \mathrm{const.}
\end{aligned}
$$

$$
\therefore\ p(\lambda | \mathcal{S}^{(i)}, \mathcal{X}) = \prod_{k=1}^{K} \mathrm{Gam}(\lambda_k | \hat{a}_k, \hat{b}_k), \qquad \text{各}\lambda_k\text{は独立.}
$$

$\hat{a}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} x_n + a, \quad \hat{b}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + b$ と書ける.

つまり，各$k=1,\cdots,K$に対して，

$$\lambda_k^{(i)} \sim \mathrm{Gam}(\lambda_k \mid \hat{a}_k, \hat{b}_k),$$

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

とサンプリングすればよい．

・πの分布$p(\pi \mid S^{(i)}, X) \propto p(S^{(i)} \mid \pi)p(\pi)$について．

これも，対数をとって計算していく．

$$\begin{aligned}
&\log p(\pi \mid S^{(i)}, X) \\
&= \log p(S^{(i)} \mid \pi)p(\pi) + \mathrm{const.} \\
&= \log p(S^{(i)} \mid \pi) + \log p(\pi) + \mathrm{const.} \\
&= \log\left(\prod_{n=1}^{N} p(s_n^{(i)} \mid \pi)\right) + \log p(\pi) + \mathrm{const.} \\
&= \sum_{n=1}^{N} \log \mathrm{Cat}(s_n^{(i)} \mid \pi) + \log \mathrm{Dir}(\pi \mid \alpha) + \mathrm{const.} \\
&= \sum_{n=1}^{N} \log\left(\prod_{k=1}^{K} \pi_k^{s_{n,k}^{(i)}}\right) + \sum_{k=1}^{K}(\alpha_k - 1)\log \pi_k + \mathrm{const.} \\
&= \sum_{k=1}^{K} \underbrace{\left(\sum_{n=1}^{N} s_{n,k}^{(i)} + \alpha_k - 1\right)\log \pi_k}_{\text{Dirichlet分布のpdfの対数をとった形.}} + \mathrm{const.}
\end{aligned}$$

$\therefore \pi^{(i)}$は，以下のDirichlet分布からサンプリングできる：

$$\pi^{(i)} \sim \mathrm{Dir}(\pi \mid \hat{\alpha}),\ \hat{\alpha}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + \alpha_k \quad (k=1,\cdots,K).$$

- 以上より，Gibbsサンプリングのアルゴリズムは次のとおり：

$\lambda^{(0)}, \pi^{(0)}$ をランダムに生成；

for $i = 1, \cdots, N_{max}$ do

  for $n = 1, \cdots, N$ do

$$\eta_{n,k} \propto \exp\left(x_n \log \lambda_k^{(i-1)} - \lambda_k^{(i-1)} + \log \pi_k^{(i-1)}\right)$$

$(k = 1, \cdots, K)$，$\sum_{k=1}^{K} \eta_{n,k} = 1$ となる η_n をつくる；

$$s_n^{(i)} \sim \mathrm{Cat}(s_n \mid \eta_n)$$

  end for;

  for $k = 1, \cdots, K$ do

$$\hat{a}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} x_n + a; \quad \hat{b}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + b;$$

$$\lambda_k^{(i)} \sim \mathrm{Gam}(\lambda_k \mid \hat{a}_k, \hat{b}_k)$$

  end for;

  $\hat{\alpha}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + \alpha_k$ $(k = 1, \cdots, K)$ となる $\hat{\alpha}$ をつくる；

$$\pi^{(i)} \sim \mathrm{Dir}(\pi \mid \hat{\alpha})$$

end for.

4.3.3 変分推論

- $p(S,\pi,\lambda|X) \approx q(S)q(\pi,\lambda)$ と，

潜在変数と分布のパラメータを分けて近似

（変分 Bayes EM アルゴリズム）

- $q(S)$ について．

4.2.2の公式を利用すると．

$$
\begin{aligned}
&\log q(S) \\
&= \mathbb{E}_{q(\lambda,\pi)}[\log p(S,\pi,\lambda,X)] + \text{const.} \\
&= \mathbb{E}_{q(\lambda,\pi)}[\log p(X|S,\pi,\lambda)] + \mathbb{E}_{q(\lambda,\pi)}[\log p(S|\pi,\lambda)] \\
&\quad + \underbrace{\mathbb{E}_{q(\lambda,\pi)}[\log p(\pi,\lambda)]}_{\text{const.}} + \text{const.}
\end{aligned}
$$

独立性

$$
\begin{aligned}
&= \mathbb{E}_{q(\lambda,\pi)}[\log p(X|S,\lambda)] + \mathbb{E}_{q(\lambda,\pi)}[\log p(S|\pi)] + \text{const.}
\end{aligned}
$$

積分中，πはここしか現れない． $\int q(\lambda,\pi)d\pi = q(\lambda)$.

$$
= \int q(\lambda,\pi)\log p(X|S,\lambda)d\lambda d\pi
$$

積分中，λはここしか現れない． $\int q(\lambda,\pi)d\lambda = q(\pi)$.

$$
\begin{aligned}
&\quad + \int q(\lambda,\pi)\log p(S|\pi)d\lambda d\pi + \text{const.} \\
&= \int q(\lambda)\log p(X|S,\lambda)d\lambda + \int q(\pi)\log p(S|\pi)d\pi + \text{const.} \\
&= \mathbb{E}_{q(\lambda)}[\log p(X|S,\lambda)] + \mathbb{E}_{q(\pi)}[\log p(S|\pi)] + \text{const.}
\end{aligned}
$$

$$= \mathbb{E}_{q(\lambda)}\left[\sum_{n=1}^{N} \log p(x_n|s_n, \lambda)\right]$$

$$+ \mathbb{E}_{q(\pi)}\left[\sum_{n=1}^{N} \log p(s_n|\pi)\right] + \text{const.}$$

期待値の線形性

$$= \sum_{n=1}^{N}\left(\mathbb{E}_{q(\lambda)}[\log p(x_n|s_n, \lambda)] + \mathbb{E}_{q(\pi)}[\log p(s_n|\pi)]\right) + \text{const}$$

→ $\log q(S)$ は s_n に関する項の和で書ける.

$\therefore\ q(S) = \prod_{n=1}^{N} q(s_n)$ と書ける.

- s_n に関する項の計算をすすめていく.

$$\mathbb{E}_{q(\lambda)}[\log p(x_n|s_n, \lambda)]$$

$$= \mathbb{E}_{q(\lambda)}\left[\log\left(\prod_{k=1}^{K} \mathrm{Poi}(x_n|\lambda_k)^{s_{n,k}}\right)\right]$$

$$= \mathbb{E}_{q(\lambda)}\left[\sum_{k=1}^{K} s_{n,k} \log \mathrm{Poi}(x_n|\lambda_k)\right]$$

期待値の線形性

$$= \sum_{k=1}^{K} s_{n,k}\, \mathbb{E}_{q(\lambda)}\left[\log \mathrm{Poi}(x_n|\lambda_k)\right]$$

λ_k のみ出てくる.

$$= \sum_{k=1}^{K} s_{n,k}\, \mathbb{E}_{q(\lambda_k)}\left[x_n \log \lambda_k - \lambda_k\right] + \text{const.}$$

期待値の線形性

$$= \sum_{k=1}^{K} s_{n,k}\left(x_n \mathbb{E}_{q(\lambda_k)}[\log \lambda_k] - \mathbb{E}_{q(\lambda_k)}[\lambda_k]\right) + \text{const.}$$

また，第2項についても，

$$\mathbb{E}_{q(\pi)}[\log p(s_n|\pi)]$$

$$= \mathbb{E}_{q(\pi)}[\log \mathrm{Cat}(s_n|\pi)]$$

$$= \mathbb{E}_{q(\pi)}\left[\log\left(\prod_{k=1}^{K} \pi_k^{s_{n,k}}\right)\right]$$

$$= \mathbb{E}_{q(\pi)}\left[\sum_{k=1}^{K} s_{n,k} \log \pi_k\right]$$

期待値の線形性

$$= \sum_{k=1}^{K} s_{n,k}\, \mathbb{E}_{q(\pi)}[\log \pi_k].$$

$$\therefore \mathbb{E}_{q(\lambda)}[\log p(x_n|s_n, \lambda)] + \mathbb{E}_{q(\pi)}[\log p(s_n|\pi)]$$

$$= \underbrace{\sum_{k=1}^{K} s_{n,k}\left(x_n \mathbb{E}_{q(\lambda_k)}[\log \lambda_k] - \mathbb{E}_{q(\lambda_k)}[\lambda_k] + \mathbb{E}_{q(\pi)}[\log \pi_k]\right)}_{\text{カテゴリ分布のpdfの対数をとった形.}} + \text{const.}$$

$$\therefore q(s_n) = \mathrm{Cat}(s_n|\eta_n),$$

$$\eta_{n,k} \propto \exp\left(x_n \mathbb{E}_{q(\lambda_k)}[\log \lambda_k] - \mathbb{E}_{q(\lambda_k)}[\lambda_k] + \mathbb{E}_{q(\pi)}[\log \pi_k]\right)$$

$(k = 1, \cdots, K)$ かつ $\sum_{k=1}^{K} \eta_{n,k} = 1.$

含まれている期待値計算を行うには, $q(\pi)$, $q(\lambda_k)$ の形を明らかにする必要があるので, ここは後回し.

- $q(\pi, \lambda)$ について.

$$\log q(\pi, \lambda)$$

$$= \mathbb{E}_{q(S)}[\log p(S, \pi, \lambda, X)] + \text{const.}$$

$$= \mathbb{E}_{q(S)}[\log p(\mathcal{X}|S,\pi,\lambda)] + \mathbb{E}_{q(S)}[\log p(S|\pi,\lambda)]$$

$$+ \mathbb{E}_{q(S)}[\log p(\pi|\lambda)] + \mathbb{E}_{q(S)}[\log p(\lambda)] + \text{const.}$$

独立性

$$= \mathbb{E}_{q(S)}[\log p(\mathcal{X}|S,\lambda)] + \mathbb{E}_{q(S)}[\log p(S|\pi)]$$

$$+ \mathbb{E}_{q(S)}[\underbrace{\log p(\pi)}_{S\text{によらない}}] + \mathbb{E}_{q(S)}[\underbrace{\log p(\lambda)}_{S\text{によらない}}] + \text{const.}$$

$$= \mathbb{E}_{q(S)}[\log p(\mathcal{X}|S,\lambda)] + \log p(\lambda)$$

$$+ \mathbb{E}_{q(S)}[\log p(S|\pi)] + \log p(\pi) + \text{const.}$$

→ λに関する項とπに関する項が完全に分離している.

$\therefore\ q(\pi,\lambda) = q(\lambda)q(\pi)$: 独立.

・λに関する項について.

$$\log q(\lambda)$$

$$= \mathbb{E}_{q(S)}[\log p(\mathcal{X}|S,\lambda)] + \log p(\lambda) + \text{const.}$$

$$= \mathbb{E}_{q(S)}\left[\sum_{n=1}^{N} \log p(x_n|s_n,\lambda)\right] + \log p(\lambda) + \text{const.}$$

$$= \sum_{n=1}^{N} \mathbb{E}_{q(S)}[\log p(x_n|s_n,\lambda)] + \log p(\lambda) + \text{const.}$$

期待値の線形性

$$= \sum_{n=1}^{N} \mathbb{E}_{q(S)}[\log p(x_n|s_n,\lambda)] + \log p(\lambda) + \text{const.}$$

期待値をとるものは, s_n以外に依らない.

$$= \sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[\log p(x_n|s_n,\lambda)] + \log p(\lambda) + \text{const.}$$

$$= \sum_{n=1}^{N} \mathbb{E}_{q(s_n)}\left[\log\left(\prod_{k=1}^{K} \mathrm{Poi}(x_n|\lambda_k)^{s_{n,k}}\right)\right] + \sum_{k=1}^{K} \log p(\lambda_k) + \mathrm{const.}$$

$$= \sum_{n=1}^{N} \mathbb{E}_{q(s_n)}\left[\sum_{k=1}^{K} s_{n,k} \log \mathrm{Poi}(x_n|\lambda_k)\right] + \sum_{k=1}^{K} \log p(\lambda_k) + \mathrm{const.}$$

期待値の線形性

$$= \sum_{k=1}^{K}\left(\sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[s_{n,k}] \log \mathrm{Poi}(x_n|\lambda_k) + \log \mathrm{Gam}(\lambda_k|a,b)\right) + \mathrm{const.}$$

$$= \sum_{k=1}^{K}\left(\sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[s_{n,k}](x_n \log\lambda_k - \lambda_k) + (a-1)\log\lambda_k - b\lambda_k\right) + \mathrm{const.}$$

$$= \sum_{k=1}^{K}\underbrace{\left(\left(\sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[s_{n,k}]x_n + a - 1\right)\log\lambda_k - \left(\sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[s_{n,k}] + b\right)\lambda_k\right)}_{\text{ガンマ分布のpdfの対数をとった形}} + \mathrm{const.}$$

$$\therefore\ q(\lambda) = \prod_{k=1}^{K} \mathrm{Gam}(\lambda_k|\hat{a}_k, \hat{b}_k),$$

各λ_kは独立.

$$\hat{a}_k := \sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[s_{n,k}]x_n + a, \quad \hat{b}_k := \sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[s_{n,k}] + b$$

と書ける.

- πに関する項について.

$$\log q(\pi)$$

$$= \mathbb{E}_{q(S)}[\log p(S|\pi)] + \log p(\pi) + \mathrm{const.}$$

$$= \mathbb{E}_{q(S)}\left[\sum_{n=1}^{N} \log p(s_n|\pi)\right] + \log p(\pi) + \mathrm{const.}$$

期待値の線形性

$$= \sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[\log \mathrm{Cat}(s_n|\pi)] + \log \mathrm{Dir}(\pi|\alpha) + \mathrm{const.}$$

$$= \sum_{n=1}^{N} \mathbb{E}_{q(s_n)}\left[\log\left(\prod_{k=1}^{K} \pi_k^{s_{n,k}}\right)\right] + \sum_{k=1}^{K}(\alpha_k - 1)\log \pi_k + \text{const.}$$

$$= \sum_{n=1}^{N} \mathbb{E}_{q(s_n)}\left[\sum_{k=1}^{K} s_{n,k} \log \pi_k\right] + \sum_{k=1}^{K}(\alpha_k - 1)\log \pi_k + \text{const.}$$

期待値の線形性

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{E}_{q(s_n)}[s_{n,k}] \log \pi_k + \sum_{k=1}^{K}(\alpha_k - 1)\log \pi_k + \text{const.}$$

$$= \underbrace{\sum_{k=1}^{K}\left(\sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[s_{n,k}] + \alpha_k - 1\right)\log \pi_k + \text{const.}}_{\text{Dirichlet分布のpdfの対数をとった形}}$$

$$\therefore q(\pi) = \mathrm{Dir}(\pi \mid \hat{\alpha}),$$

$$\hat{\alpha}_k := \sum_{n=1}^{N} \mathbb{E}_{q(s_n)}[s_{n,k}] + \alpha_k \quad (k = 1, \cdots, K).$$

→ $q(\lambda_k)$, $q(\pi)$ とも，それぞれ事前分布と同じ種類の分布になった．

- 各種期待値について．

- $q(\lambda_k) = \mathrm{Gam}(\lambda_k \mid \hat{a}_k, \hat{b}_k)$ なので，

$$\mathbb{E}_{q(\lambda_k)}[\lambda_k] = \frac{\hat{a}_k}{\hat{b}_k}, \quad \mathbb{E}_{q(\lambda_k)}[\log \lambda_k] = \psi(\hat{a}_k) - \log \hat{b}_k.$$

- $q(\pi) = \mathrm{Dir}(\pi \mid \hat{\alpha})$ なので，

$$\mathbb{E}_{q(\pi)}[\log \pi_k] = \psi(\hat{\alpha}_k) - \psi\left(\sum_{i=1}^{K} \hat{\alpha}_i\right).$$

- $q(s_n) = \mathrm{Cat}(s_n \mid \eta_n)$ なので，

$$\mathbb{E}_{q(s_n)}[s_{n,k}] = \eta_{n,k}.$$

- 結局，分布は以下のとおり：

- $q(s_n) = \mathrm{Cat}(s_n \mid \eta_n)$,

$$\eta_{n,k} \propto \exp\left(x_n \left(\psi(\hat{a}_k) - \log \hat{b}_k \right) - \frac{\hat{a}_k}{\hat{b}_k} + \psi(\hat{\alpha}_k) - \psi\left(\sum_{i=1}^{K} \hat{\alpha}_i \right) \right) \quad (k = 1, \cdots, K)$$

かつ $\sum_{k=1}^{K} \eta_{n,k} = 1$.

- $q(\lambda_k) = \mathrm{Gam}(\lambda_k \mid \hat{a}_k, \hat{b}_k)$

$$\hat{a}_k := \sum_{n=1}^{N} \eta_{n,k} x_n + a, \quad \hat{b}_k := \sum_{n=1}^{N} \eta_{n,k} + b.$$

- $q(\pi) = \mathrm{Dir}(\pi \mid \hat{\alpha})$,

$$\hat{\alpha}_k := \sum_{n=1}^{N} \eta_{n,k} + \alpha_k \quad (k = 1, \cdots, K).$$

- 以上より，変分推論のアルゴリズムは次のとおり：

$q(\lambda_k) = \mathrm{Gam}(\lambda_k \mid \hat{a}_k, \hat{b}_k)$, $q(\pi) = \mathrm{Dir}(\pi \mid \hat{\alpha})$ と initialize;

for $i = 1, \cdots, N_{max}$ do ← 適当に上限を設けている

for $n = 1, \cdots, N$ do

$$\eta_{n,k} \propto \exp\left(x_n \left(\psi(\hat{a}_k) - \log \hat{b}_k \right) - \hat{a}_k / \hat{b}_k + \psi(\hat{\alpha}_k) - \psi\left(\sum_{i=1}^{K} \hat{\alpha}_i \right) \right) \quad (k = 1, \cdots, K),$$

$\sum_{k=1}^{K} \eta_{n,k} = 1$ となる η_n をつくる

end for;

for $k = 1, \cdots, K$ do

$$\hat{a}_k := \sum_{n=1}^{N} \eta_{n,k} x_n + a;$$

$$\hat{b}_k := \sum_{n=1}^{N} \eta_{n,k} + b$$

end for;

$\hat{\alpha}_k := \sum_{n=1}^{N} \eta_{n,k} + \alpha_k \quad (k = 1, \cdots, K)$ となる $\hat{\alpha}$ をつくる;

$\pi \sim \mathrm{Dir}(\pi \mid \hat{\alpha})$

end for.

・ここでも，終了条件にELBOを採用することもできる．

ELBOを計算してみる．

$$\mathcal{L}[q]$$

$$= \mathbb{E}_{q(S)q(\pi,\lambda)}\left[\log \frac{p(S,\pi,\lambda,X)}{q(S)q(\pi,\lambda)}\right]$$

乗法則

$$= \mathbb{E}_{q(S)q(\pi,\lambda)}\left[\log \frac{p(X|S,\pi,\lambda)p(S|\pi,\lambda)p(\pi|\lambda)p(\lambda)}{q(S)q(\pi,\lambda)}\right]$$

独立性

$$= \mathbb{E}_{q(S)q(\pi,\lambda)}\left[\log \frac{p(X|S,\lambda)p(S|\pi)p(\pi)p(\lambda)}{q(S)q(\pi,\lambda)}\right]$$

$q(\pi,\lambda)=q(\pi)q(\lambda)$

$$= \mathbb{E}_{q(S)q(\pi)q(\lambda)}\left[\log \frac{p(X|S,\lambda)p(S|\pi)p(\pi)p(\lambda)}{q(S)q(\pi)q(\lambda)}\right]$$

logの性質と
期待値の線形性

$$= \mathbb{E}_{q(S)q(\pi)q(\lambda)}\left[\log p(X|S,\lambda)\right] + \mathbb{E}_{q(S)q(\pi)q(\lambda)}\left[\log p(S|\pi)\right]$$

πによらない　　λによらない

$$- \mathbb{E}_{q(S)q(\pi)q(\lambda)}\left[\log q(S)\right] + \mathbb{E}_{q(S)q(\pi)q(\lambda)}\left[\log \frac{p(\lambda)}{q(\lambda)}\right]$$

π, λによらない　　π, Sによらない

$$+ \mathbb{E}_{q(S)q(\pi)q(\lambda)}\left[\log \frac{p(\pi)}{q(\pi)}\right]$$

λ, Sによらない

$$= \mathbb{E}_{q(S)q(\lambda)}\left[\log p(X|S,\lambda)\right] + \mathbb{E}_{q(S)q(\pi)}\left[\log p(S|\pi)\right]$$

$$- \mathbb{E}_{q(S)}\left[\log q(S)\right] + \mathbb{E}_{q(\lambda)}\left[\log \frac{p(\lambda)}{q(\lambda)}\right] + \mathbb{E}_{q(\pi)}\left[\log \frac{p(\pi)}{q(\pi)}\right]$$

$$
\begin{aligned}
&= \mathbb{E}_{q(S)q(\lambda)}\left[\sum_{n=1}^{N}\log p(x_n|s_n,\lambda)\right] + \mathbb{E}_{q(S)q(\pi)}\left[\sum_{n=1}^{N}\log p(s_n|\pi)\right] \\
&\quad - \mathbb{E}_{q(S)}\left[\sum_{n=1}^{N}\log q(s_n)\right] - \mathrm{KL}[q(\lambda)\|p(\lambda)] \\
&\quad - \mathrm{KL}[q(\pi)\|p(\pi)] \\
&= \sum_{n=1}^{N}\mathbb{E}_{q(s_n)q(\lambda)}\left[\log p(x_n|s_n,\lambda)\right] + \sum_{n=1}^{N}\mathbb{E}_{q(s_n)q(\pi)}\left[\log p(s_n|\pi)\right] \\
&\quad - \sum_{n=1}^{N}\mathbb{E}_{q(s_n)}\left[\log q(s_n)\right] - \mathrm{KL}[q(\lambda)\|p(\lambda)] \\
&\quad - \mathrm{KL}[q(\pi)\|p(\pi)].
\end{aligned}
$$

- 各項の計算は，分布の形が分かっているので，やればできる．
- $\mathcal{L}[q]$ の最初の3項：データに対する当てはまり具合や，分布 $q(S)$ の望ましさを表す．
- $\mathcal{L}[q]$ の最後の2項：$q(\lambda), q(\pi)$ が事前分布 $p(\lambda), p(\pi)$ から大きく離れるのを防ぐ"正則化項"．

→ 変分推論アルゴリズムは，自然に過適合を防ぐようになっている！

4.3.4 崩壊型Gibbsサンプリング

・まずは，同時分布 $p(\mathcal{S}, \pi, \lambda, \mathcal{X})$ から，分布のパラメータ π, λ を周辺化により除去する：

$$p(\mathcal{S}, \mathcal{X}) = \int p(\mathcal{S}, \pi, \lambda, \mathcal{X})\, d\lambda\, d\pi.$$

→ $\mathcal{S}$ は $p(\mathcal{S}|\mathcal{X})$ からサンプリングすればよい．

パラメータの事後分布などは，サンプリングされた $\mathcal{S}$ から計算可能．

・周辺化後は，各 s_n が互いに依存関係をもってしまう．

（各 s_n は π：given のもとでは条件付き独立であった）

→ 全ての s_n を同時にサンプリングするのは計算量的にキツい：

$\mathcal{S} = \{s_1, \cdots, s_N\} \subseteq \{0,1\}^K$ のとりうる組合せ数：K^N 通り．

→ 周辺化後のモデルの事後分布 $p(\mathcal{S}|\mathcal{X})$ に対し，Gibbs サンプリングを適用して，各 s_n をばらばらにサンプルする．

。$\mathcal{S}_{\backslash n}^{(i)} := \{s_1^{(i)}, \cdots, s_{n-1}^{(i)}, s_{n+1}^{(i-1)}, \cdots, s_N^{(i-1)}\}$，

$\mathcal{X}_{\backslash n} := \{x_1, \cdots, x_{n-1}, x_{n+1}, \cdots, x_N\}$ とする．

$s_n^{(i)} \sim p(s_n|\mathcal{X}, \mathcal{S}_{\backslash n}^{(i)})$ としたい．

→ $p(s_n|\mathcal{X}, \mathcal{S}_{\backslash n}^{(i)})$ を計算していく．

- $$p(s_n \mid \mathcal{X}, \mathcal{S}_{\backslash n}^{(i)})$$

Bayes' Thm.

$$\propto p(s_n, \mathcal{X}, \mathcal{S}_{\backslash n}^{(i)})$$

$\mathcal{X} = \mathcal{X}_{\backslash n} \cup \{x_n\}$

$$= p(x_n, \mathcal{X}_{\backslash n}, s_n, \mathcal{S}_{\backslash n}^{(i)})$$

乗法則

グラフィカルモデルを考えると s_n は $\mathcal{X}_{\backslash n}$ の各 x と条件付き独立だと分かる。

$$= p(x_n \mid \mathcal{X}_{\backslash n}, s_n, \mathcal{S}_{\backslash n}^{(i)}) p(\mathcal{X}_{\backslash n} \mid s_n, \mathcal{S}_{\backslash n}^{(i)}) p(s_n \mid \mathcal{S}_{\backslash n}^{(i)}) p(\mathcal{S}_{\backslash n}^{(i)})$$

$$= p(x_n \mid \mathcal{X}_{\backslash n}, s_n, \mathcal{S}_{\backslash n}^{(i)}) \underbrace{p(\mathcal{X}_{\backslash n} \mid \mathcal{S}_{\backslash n}^{(i)})}_{s_n\text{に依らない}} p(s_n \mid \mathcal{S}_{\backslash n}^{(i)}) \underbrace{p(\mathcal{S}_{\backslash n}^{(i)})}_{s_n\text{に依らない}}$$

$$\propto p(x_n \mid \mathcal{X}_{\backslash n}, s_n, \mathcal{S}_{\backslash n}^{(i)}) p(s_n \mid \mathcal{S}_{\backslash n}^{(i)}).$$

- $p(s_n \mid \mathcal{S}_{\backslash n}^{(i)})$ は，$\mathcal{S}_{\backslash n}^{(i)}$: given のときの s_n の予測分布と解釈可能：

$$p(s_n \mid \mathcal{S}_{\backslash n}^{(i)}) = \int p(s_n \mid \pi) p(\pi \mid \mathcal{S}_{\backslash n}^{(i)}) d\pi.$$

$p(\pi \mid \mathcal{S}_{\backslash n}^{(i)})$ は $\mathcal{S}_{\backslash n}^{(i)}$: given のときの「π の事後分布」.

この計算は"3.2.2 カテゴリ分布の学習と予測"で既にやった！

$$\therefore\ p(\pi \mid \mathcal{S}_{\backslash n}^{(i)}) = \mathrm{Dir}(\pi \mid \hat{\alpha}_{\backslash n}^{(i)}),$$

$$\hat{\alpha}_{\backslash n,k}^{(i)} := \sum_{n'=1}^{n-1} s_{n',k}^{(i)} + \sum_{n'=n+1}^{N} s_{n',k}^{(i-1)} + \alpha_k \quad (k = 1, \cdots, K).$$

よって，"3.2.2 カテゴリ分布の学習と予測"の予測分布の結果より，

$$p(s_n \mid \mathcal{S}_{\backslash n}^{(i)}) = \int \mathrm{Cat}(s_n \mid \pi) \mathrm{Dir}(\pi \mid \hat{\alpha}_{\backslash n}^{(i)}) d\pi$$

$$= \mathrm{Cat}(s_n \mid \eta_{\backslash n}),$$

$$\eta_{\backslash n,k} \propto \hat{\alpha}_{\backslash n,k}^{(i)}, \quad \sum_{k=1}^{K} \eta_{\backslash n,k} = 1.$$

・$p(x_n | \mathcal{X}_{\backslash n}, s_n, \mathcal{S}_{\backslash n}^{(i)})$ は, $\mathcal{X}_{\backslash n}, \mathcal{S}_{\backslash n}^{(i)}$: given のときの x_n の予測分布のようなもの：

$$p(x_n | \mathcal{X}_{\backslash n}, s_n, \mathcal{S}_{\backslash n}^{(i)}) = \int p(x_n | s_n, \lambda) p(\lambda | \mathcal{X}_{\backslash n}, \mathcal{S}_{\backslash n}^{(i)}) d\lambda.$$

$p(\lambda | \mathcal{X}_{\backslash n}, \mathcal{S}_{\backslash n}^{(i)})$ は $\mathcal{X}_{\backslash n}, \mathcal{S}_{\backslash n}^{(i)}$: given のときの「λ の事後分布」.

$$\begin{aligned}
p(\lambda | \mathcal{X}_{\backslash n}, \mathcal{S}_{\backslash n}^{(i)}) &\propto p(\lambda, \mathcal{X}_{\backslash n}, \mathcal{S}_{\backslash n}^{(i)}) \\
&= p(\mathcal{X}_{\backslash n} | \mathcal{S}_{\backslash n}^{(i)}, \lambda) p(\mathcal{S}_{\backslash n}^{(i)} | \lambda) p(\lambda) \quad \text{独立性} \\
&= p(\mathcal{X}_{\backslash n} | \mathcal{S}_{\backslash n}^{(i)}, \lambda) \underbrace{p(\mathcal{S}_{\backslash n}^{(i)})}_{\lambda\text{によらない}} p(\lambda) \\
&\propto p(\mathcal{X}_{\backslash n} | \mathcal{S}_{\backslash n}^{(i)}, \lambda) p(\lambda).
\end{aligned}$$

$$\begin{aligned}
&\log p(\mathcal{X}_{\backslash n} | \mathcal{S}_{\backslash n}^{(i)}, \lambda) p(\lambda) \\
&= \log p(\mathcal{X}_{\backslash n} | \mathcal{S}_{\backslash n}^{(i)}, \lambda) + \log p(\lambda) \\
&= \sum_{n'=1}^{n-1} \log p(x_{n'} | s_{n'}^{(i)}, \lambda) + \sum_{n'=n+1}^{N} \log p(x_{n'} | s_{n'}^{(i-1)}, \lambda) + \log p(\lambda) \\
&= \sum_{n'=1}^{n-1} \sum_{k=1}^{K} s_{n',k}^{(i)} \log \mathrm{Poi}(x_{n'} | \lambda_k) + \sum_{n'=n+1}^{N} \sum_{k=1}^{K} s_{n',k}^{(i-1)} \log \mathrm{Poi}(x_{n'} | \lambda_k) \\
&\quad + \sum_{k=1}^{K} \log \mathrm{Gam}(\lambda_k | a, b) \\
&= \sum_{k=1}^{K} \Bigg(\sum_{n'=1}^{n-1} s_{n',k}^{(i)} (x_{n'} \log \lambda_k - \lambda_k) + \sum_{n'=n+1}^{N} s_{n',k}^{(i-1)} (x_{n'} \log \lambda_k - \lambda_k) \\
&\qquad + (a-1) \log \lambda_k - b\lambda_k \Bigg) + \text{const.}
\end{aligned}$$

$$= \sum_{k=1}^{K}\Bigg(\underline{\Bigg(\sum_{n'=1}^{n-1} s_{n',k}^{(i)} x_{n'} + \sum_{n'=n+1}^{N} s_{n',k}^{(i-1)} x_{n'} + a - 1\Bigg)\log\lambda_k}$$

$$\underline{- \Bigg(\sum_{n'=1}^{n-1} s_{n',k}^{(i)} + \sum_{n'=n+1}^{N} s_{n',k}^{(i-1)} + b\Bigg)\lambda_k}\Bigg) + \text{const.}$$

└ ガンマ分布のpdfの対数をとった形

$$\therefore\ p(\lambda | \mathcal{X}_{\backslash n}, \mathcal{S}_{\backslash n}^{(i)}) = \prod_{k=1}^{K} \mathrm{Gam}(\lambda_k | \hat{a}_{\backslash n,k}^{(i)}, \hat{b}_{\backslash n,k}^{(i)}),$$

$$\hat{a}_{\backslash n,k}^{(i)} := \sum_{n'=1}^{n-1} s_{n',k}^{(i)} x_{n'} + \sum_{n'=n+1}^{N} s_{n',k}^{(i-1)} x_{n'} + a,$$

$$\hat{b}_{\backslash n,k}^{(i)} := \sum_{n'=1}^{n-1} s_{n',k}^{(i)} + \sum_{n'=n+1}^{N} s_{n',k}^{(i-1)} + b$$

と書ける．これを用いて，λを積分除去し，$p(x_n | \mathcal{X}_{\backslash n}, s_n, \mathcal{S}_{\backslash n}^{(i)})$

を計算する．とくに，$s_n = (0, \cdots, \overset{k}{\overset{\downarrow}{1}}, \cdots, 0)^T$ $(s_{n,k} = 1)$ とすると，

$$p(x_n | \mathcal{X}_{\backslash n}, s_{n,k} = 1, \mathcal{S}_{\backslash n}^{(i)})$$

$$= \int p(x_n | \lambda_k) p(\lambda_k | \mathcal{X}_{\backslash n}, \mathcal{S}_{\backslash n}^{(i)}) d\lambda_k$$

$$= \int \mathrm{Poi}(x_n | \lambda_k) \mathrm{Gam}(\lambda_k | \hat{a}_{\backslash n,k}^{(i)}, \hat{b}_{\backslash n,k}^{(i)}) d\lambda_k$$

3.2.3でやった予測分布の計算

$$= \mathrm{NB}(x_n | \hat{a}_{\backslash n,k}^{(i)}, (\hat{b}_{\backslash n,k}^{(i)} + 1)^{-1}).$$

よって，

$$p(x_n | \mathcal{X}_{\backslash n}, s_n, \mathcal{S}_{\backslash n}^{(i)})$$

$$= \int p(x_n | s_n, \lambda) p(\lambda | \mathcal{X}_{\backslash n}, \mathcal{S}_{\backslash n}^{(i)}) d\lambda$$

$$= \int \prod_{k=1}^{K} \mathrm{Poi}(x_n | \lambda_k)^{s_{n,k}} \mathrm{Gam}(\lambda_k | \hat{a}_{\backslash n,k}^{(i)}, \hat{b}_{\backslash n,k}^{(i)})\, d\lambda$$

$$= \prod_{k=1}^{K} \int \mathrm{Poi}(x_n|\lambda_k)^{s_{n,k}} \mathrm{Gam}(\lambda_k|\hat{a}^{(i)}_{\backslash n,k}, \hat{b}^{(i)}_{\backslash n,k}) d\lambda_k$$

$s_{n,k}=1$ となる k は一つだけ.

$$= \prod_{k=1}^{K} \left(\int \mathrm{Poi}(x_n|\lambda_k) \mathrm{Gam}(\lambda_k|\hat{a}^{(i)}_{\backslash n,k}, \hat{b}^{(i)}_{\backslash n,k}) d\lambda_k \right)^{s_{n,k}}$$

$$= \prod_{k=1}^{K} \mathrm{NB}(x_n|\hat{a}^{(i)}_{\backslash n,k}, (\hat{b}^{(i)}_{\backslash n,k}+1)^{-1})^{s_{n,k}}.$$

- $\log p(s_n|\mathcal{X}, \mathcal{S}^{(i)}_{\backslash n})$

$$= \log p(x_n|\mathcal{X}_{\backslash n}, s_n, \mathcal{S}^{(i)}_{\backslash n}) + \log p(s_n|\mathcal{S}^{(i)}_{\backslash n}) + \text{const.}$$

$$= \sum_{k=1}^{K} s_{n,k} \log \mathrm{NB}(x_n|\hat{a}^{(i)}_{\backslash n,k}, (\hat{b}^{(i)}_{\backslash n,k}+1)^{-1}) + \sum_{k=1}^{K} s_{n,k} \log \eta_{\backslash n,k} + \text{const.}$$

$$= \sum_{k=1}^{K} s_{n,k} \log \eta_{\backslash n,k} \mathrm{NB}(x_n|\hat{a}^{(i)}_{\backslash n,k}, (\hat{b}^{(i)}_{\backslash n,k}+1)^{-1}) + \text{const.}$$

$\eta_{\backslash n,k} \propto \hat{\alpha}^{(i)}_{\backslash n,k}$

$$= \underbrace{\sum_{k=1}^{K} s_{n,k} \log \hat{\alpha}^{(i)}_{\backslash n,k} \mathrm{NB}(x_n|\hat{a}^{(i)}_{\backslash n,k}, (\hat{b}^{(i)}_{\backslash n,k}+1)^{-1})}_{\text{カテゴリ分布の pmf. の対数をとったもの.}} + \text{const.}$$

以上より, $s^{(i)}_n$ のサンプリングは, 次のように行えばよい:

$$s^{(i)}_n \sim p(s_n|\mathcal{X}, \mathcal{S}^{(i)}_{\backslash n}) = \mathrm{Cat}(s_n|\zeta_{\backslash n}),$$

$$\zeta_{\backslash n,k} \propto \hat{\alpha}^{(i)}_{\backslash n,k} \mathrm{NB}(x_n|\hat{a}^{(i)}_{\backslash n,k}, (\hat{b}^{(i)}_{\backslash n,k}+1)^{-1}), \quad \sum_{k=1}^{K} \zeta_{\backslash n,k} = 1.$$

- 次が成立していることに注意する.

- $\hat{\alpha}^{(i)}_{\backslash (n+1),k}$

$$= \sum_{n'=1}^{n} s^{(i)}_{n',k} + \sum_{n'=n+2}^{N} s^{(i-1)}_{n',k} + \alpha_k = \sum_{n'=1}^{n-1} s^{(i)}_{n',k} + \sum_{n'=n+1}^{N} s^{(i-1)}_{n',k} + \alpha_k + s^{(i)}_{n,k} - s^{(i-1)}_{n+1,k}$$

$$= \hat{\alpha}^{(i)}_{\backslash n,k} + s^{(i)}_{n,k} - s^{(i-1)}_{n+1,k}. \qquad (n=1,\cdots,N-1)$$

$$\hat{\alpha}^{(i)}_{\setminus 1,k}$$

$$= \sum_{n'=2}^{N} s^{(i-1)}_{n',k} + \alpha_k = \sum_{n'=1}^{N-1} s^{(i-1)}_{n',k} + \alpha_k + s^{(i-1)}_{N,k} - s^{(i-1)}_{1,k}$$

$$= \hat{\alpha}^{(i-1)}_{\setminus N,k} + s^{(i-1)}_{N,k} - s^{(i-1)}_{1,k}$$

- $\hat{a}^{(i)}_{\setminus(n+1),k}$

$$= \sum_{n'=1}^{n} s^{(i)}_{n',k} x_{n'} + \sum_{n'=n+2}^{N} s^{(i-1)}_{n',k} x_{n'} + a$$

$$= \sum_{n'=1}^{n-1} s^{(i)}_{n',k} x_{n'} + \sum_{n'=n+1}^{N} s^{(i-1)}_{n',k} x_{n'} + a + s^{(i)}_{n,k} x_n - s^{(i-1)}_{n+1,k} x_{n+1}$$

$$= \hat{a}^{(i)}_{\setminus n,k} + s^{(i)}_{n,k} x_n - s^{(i-1)}_{n+1,k} x_{n+1}. \qquad (n = 1, \cdots, N-1)$$

$$\hat{a}^{(i)}_{\setminus 1,k}$$

$$= \sum_{n'=2}^{N} s^{(i-1)}_{n',k} x_{n'} + a$$

$$= \sum_{n'=1}^{N-1} s^{(i-1)}_{n',k} x_{n'} + a + s^{(i-1)}_{N,k} x_N - s^{(i-1)}_{1,k} x_1$$

$$= \hat{a}^{(i-1)}_{\setminus N,k} + s^{(i-1)}_{N,k} x_N - s^{(i-1)}_{1,k} x_1$$

- $\hat{b}^{(i)}_{\setminus(n+1),k}$

$$= \sum_{n'=1}^{n} s^{(i)}_{n',k} + \sum_{n'=n+2}^{N} s^{(i-1)}_{n',k} + b$$

$$= \sum_{n'=1}^{n-1} s^{(i)}_{n',k} + \sum_{n'=n+1}^{N} s^{(i-1)}_{n',k} + b + s^{(i)}_{n,k} - s^{(i-1)}_{n+1,k}$$

$$= \hat{b}^{(i)}_{\setminus n,k} + s^{(i)}_{n,k} - s^{(i-1)}_{n+1,k}. \qquad (n = 1, \cdots, N-1)$$

$$\hat{b}^{(i)}_{\backslash 1,k}$$

$$= \sum_{n'=2}^{N} s^{(i-1)}_{n',k} + b$$

$$= \sum_{n'=1}^{N-1} s^{(i-1)}_{n',k} + b + s^{(i-1)}_{N,k} - s^{(i-1)}_{1,k}$$

$$= \hat{b}^{(i-1)}_{\backslash N,k} + s^{(i-1)}_{N,k} - s^{(i-1)}_{1,k} .$$

・実装上は、上記に注意することで、計算速度とメモリの面で効率的。

崩壊型 Gibbsサンプリングのアルゴリズムを以下にまとめる：

$\mathcal{S} = \{ s_1^{(0)}, \cdots, s_N^{(0)} \}$ を initialize;

for $k = 1, \cdots, K$ do

$\hat{\alpha}_k := \sum_{n=1}^{N} s^{(0)}_{n,k} + \alpha_k$;

$\hat{a}_k := \sum_{n=1}^{N} s^{(0)}_{n,k} x_n + a$;

$\hat{b}_k := \sum_{n=1}^{N} s^{(0)}_{n,k} + b$

end for;

for $i = 1, \cdots, N_{max}$ do ← 適当に上限を設けている

for $n = 1, \cdots, N$ do

for $k = 1, \cdots, K$ do

$\hat{\alpha}_k := \hat{\alpha}_k - s^{(i-1)}_{n,k}$; $\hat{a}_k := \hat{a}_k - s^{(i-1)}_{n,k} x_n$; $\hat{b}_k := \hat{b}_k - s^{(i-1)}_{n,k}$

end for;

for $k = 1, \cdots, K$ do

$\zeta_k := \hat{\alpha}_k \mathrm{NB}(x_n | \hat{a}_k, (\hat{b}_k + 1)^{-1})$

end for;

$\sum_{k=1}^{K} \zeta_k = 1$ となるよう ζ を正規化;

$s_n^{(i)} \sim \mathrm{Cat}(s_n | \zeta)$;

for $k = 1, \cdots, K$ do

$\hat{\alpha}_k := \hat{\alpha}_k + s_{n,k}^{(i)}$; $\hat{a}_k := \hat{a}_k + s_{n,k}^{(i)} x_n$; $\hat{b}_k := \hat{b}_k + s_{n,k}^{(i)}$

end for

end for

end for.

§4.4 Gauss混合モデルにおける推論

4.4.1 Gauss混合モデル

- Kクラスタある D次元データの生成過程を表現.

- $\boldsymbol{\mu} := \{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_K\}$, $\boldsymbol{\Lambda} := \{\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_K\}$, $\boldsymbol{\theta}_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k\}$: パラメータ.

  クラスタ k に対する観測モデル: $p(\boldsymbol{x}_n | \boldsymbol{\theta}_k) := \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})$

  → 混合分布における条件付き分布は,

$$p(\boldsymbol{x}_n | \boldsymbol{s}_n, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{k=1}^{K} \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})^{s_{n,k}}$$

  $\boldsymbol{s}_n$ によって, Kコの Gauss分布の一つが指定される.

- $\boldsymbol{\mu}, \boldsymbol{\Lambda}$ に対して, 事前分布を設定する.

  Gauss分布の共役事前分布 → Gauss-Wishart分布.

$$p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) := \mathrm{NW}(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k | \boldsymbol{m}, \beta, \nu, \boldsymbol{W})$$
$$= \mathcal{N}(\boldsymbol{\mu}_k | \boldsymbol{m}, (\beta \boldsymbol{\Lambda}_k)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_k | \nu, \boldsymbol{W})$$

クラスタごとに違うパラメータにしてもよい

  $\boldsymbol{m} \in \mathbb{R}^D, \beta \in \mathbb{R}_{>0}, \nu > D-1, \boldsymbol{W} \in M_D(\mathbb{R}), \boldsymbol{W} > O$: ハイパーパラメータ.

- S に対する事前分布は, $p(\boldsymbol{s}_n | \boldsymbol{\pi}) := \mathrm{Cat}(\boldsymbol{s}_n | \boldsymbol{\pi})$,

  $\boldsymbol{\pi}$ に対する事前分布は, $p(\boldsymbol{\pi}) := \mathrm{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha})$

  ($\boldsymbol{\alpha}$: ハイパーパラメータ) と設定する.

4.4.2 Gibbsサンプリング

- 事後分布 $p(\mathcal{S}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\pi} | \mathcal{X})$ からサンプリングをする。

次のように、パラメータと潜在変数を分けてサンプリング：

$$\mathcal{S}^{(i)} \sim p(\mathcal{S} | \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}, \boldsymbol{\pi}^{(i-1)}, \mathcal{X}) \quad -(1)$$

$$\boldsymbol{\mu}^{(i)}, \boldsymbol{\Lambda}^{(i)}, \boldsymbol{\pi}^{(i)} \sim p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\pi} | \mathcal{S}^{(i)}, \mathcal{X}) \quad -(2)$$

- (1)について、$p(\mathcal{S} | \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}, \boldsymbol{\pi}^{(i-1)}, \mathcal{X})$ を求める。

$$p(\mathcal{S} | \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}, \boldsymbol{\pi}^{(i-1)}, \mathcal{X})$$

$$\propto p(\mathcal{S}, \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}, \boldsymbol{\pi}^{(i-1)}, \mathcal{X})$$

$$= p(\mathcal{X} | \mathcal{S}, \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}, \boldsymbol{\pi}^{(i-1)}) p(\mathcal{S} | \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}, \boldsymbol{\pi}^{(i-1)})$$

$$\underbrace{\times p(\boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}, \boldsymbol{\pi}^{(i-1)})}_{\mathcal{S}\text{に依らない}}$$

↓ 独立性

$$\propto p(\mathcal{X} | \mathcal{S}, \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}) p(\mathcal{S} | \boldsymbol{\pi}^{(i-1)})$$

$$= \prod_{n=1}^{N} p(\boldsymbol{x}_n | \boldsymbol{s}_n, \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}) p(\boldsymbol{s}_n | \boldsymbol{\pi}^{(i-1)}).$$

→ $\boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}, \boldsymbol{\pi}^{(i-1)}, \mathcal{X}$: givenで、各 $\boldsymbol{s}_n$ は条件付き独立.

↳ $\boldsymbol{s}_n$ のサンプリングを考えればよい。

- $\log p(\boldsymbol{x}_n | \boldsymbol{s}_n, \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)})$

$$= \sum_{k=1}^{K} s_{n,k} \log \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{\mu}_k^{(i-1)}, (\boldsymbol{\Lambda}_k^{(i-1)})^{-1})$$

$$= \sum_{k=1}^{K} s_{n,k} \left(-\frac{1}{2} (\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(i-1)})^T \boldsymbol{\Lambda}_k^{(i-1)} (\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(i-1)}) + \frac{1}{2} \log (\det \boldsymbol{\Lambda}_k^{(i-1)}) \right) + \text{const}$$

・ $\log p(\boldsymbol{s}_n|\boldsymbol{\pi}^{(i-1)}) = \log \mathrm{Cat}(\boldsymbol{s}_n|\boldsymbol{\pi}^{(i-1)}) = \sum_{k=1}^{K} s_{n,k} \log \pi_k .$

であった．

$$\log p(\boldsymbol{x}_n|\boldsymbol{s}_n, \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}) p(\boldsymbol{s}_n|\boldsymbol{\pi}^{(i-1)})$$

$$= \log p(\boldsymbol{x}_n|\boldsymbol{s}_n, \boldsymbol{\mu}^{(i-1)}, \boldsymbol{\Lambda}^{(i-1)}) + \log p(\boldsymbol{s}_n|\boldsymbol{\pi}^{(i-1)})$$

$$= \sum_{k=1}^{K} s_{n,k} \Big(-\frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(i-1)})^{\mathsf{T}} \boldsymbol{\Lambda}_k^{(i-1)} (\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(i-1)}) + \frac{1}{2}\log(\det \boldsymbol{\Lambda}_k^{(i-1)}) + \log \pi_k \Big) + \mathrm{const}.$$

制約 $\sum_{k=1}^{K} s_{n,k} = 1$ を考慮すると，これは $\boldsymbol{s}_n$ に関する

カテゴリ分布（の定数倍）の対数をとったもの．

$$\therefore \boldsymbol{s}_n^{(i)} \sim \mathrm{Cat}(\boldsymbol{s}_n|\boldsymbol{\eta}_n^{(i)}),$$

$$\eta_{n,k}^{(i)} \propto \exp\Big(-\frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(i-1)})^{\mathsf{T}} \boldsymbol{\Lambda}_k^{(i-1)} (\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(i-1)}) + \frac{1}{2}\log(\det \boldsymbol{\Lambda}_k^{(i-1)}) + \log \pi_k\Big),$$

$$\sum_{k=1}^{K} \eta_{n,k}^{(i)} = 1.$$

◦ (2) について，$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\pi}|\mathcal{S}^{(i)}, \mathcal{X})$ を求める．

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\pi}|\mathcal{S}^{(i)}, \mathcal{X})$$

$$\propto p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \mathcal{S}^{(i)}, \mathcal{X})$$

独立性

$$= p(\mathcal{X}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \mathcal{S}^{(i)}) p(\mathcal{S}^{(i)}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\pi}) p(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\boldsymbol{\pi}) p(\boldsymbol{\pi})$$

$$= p(\mathcal{X}|S^{(i)},\mu,\Lambda)p(S^{(i)}|\pi)p(\mu,\Lambda)p(\pi)$$

$$= \underbrace{p(\mathcal{X}|S^{(i)},\mu,\Lambda)p(\mu,\Lambda)}_{\mu,\Lambda\text{についての項}}\underbrace{p(S^{(i)}|\pi)p(\pi)}_{\pi\text{についての項}}$$

→ $S^{(i)},\mathcal{X}$: given のもとで, (μ,Λ) と π は条件付き独立:

$$p(\mu,\Lambda|S^{(i)},\mathcal{X}) \propto p(\mathcal{X}|S^{(i)},\mu,\Lambda)p(\mu,\Lambda),$$

$$p(\pi|S^{(i)},\mathcal{X}) \propto p(S^{(i)}|\pi)p(\pi),$$

$$p(\mu,\Lambda,\pi|S^{(i)},\mathcal{X}) = p(\mu,\Lambda|S^{(i)},\mathcal{X})p(\pi|S^{(i)},\mathcal{X}).$$

・ $\log p(\mu,\Lambda|S^{(i)},\mathcal{X})$

$$= \log p(\mathcal{X}|S^{(i)},\mu,\Lambda) + \log p(\mu,\Lambda) + \text{const.}$$

$$= \sum_{n=1}^{N}\log p(x_n|s_n^{(i)},\mu,\Lambda) + \sum_{k=1}^{K}\log p(\mu_k,\Lambda_k) + \text{const.}$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} s_{n,k}^{(i)}\log\mathcal{N}(x_n|\mu_k,\Lambda_k^{-1}) + \sum_{k=1}^{K}\log NW(\mu_k,\Lambda_k|m,\beta,\nu,W) + \text{const.}$$

$$= \sum_{k=1}^{K}\left(\sum_{n=1}^{N} s_{n,k}^{(i)}\log\mathcal{N}(x_n|\mu_k,\Lambda_k^{-1}) + \log NW(\mu_k,\Lambda_k|m,\beta,\nu,W)\right) + \text{const.}$$

→ $p(\mu,\Lambda|S^{(i)},\mathcal{X}) = \prod_{k=1}^{K} p(\mu_k,\Lambda_k|S^{(i)},\mathcal{X})$,

← 条件付き独立

$$\log p(\mu_k,\Lambda_k|S^{(i)},\mathcal{X})$$

$$= \sum_{n=1}^{N} s_{n,k}^{(i)}\log\mathcal{N}(x_n|\mu_k,\Lambda_k^{-1}) + \log NW(\mu_k,\Lambda_k|m,\beta,\nu,W) + \text{const.}$$

$$p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k \mid \mathcal{S}^{(i)}, \mathcal{X}) = p(\boldsymbol{\mu}_k \mid \boldsymbol{\Lambda}_k, \mathcal{S}^{(i)}, \mathcal{X})\, p(\boldsymbol{\Lambda}_k \mid \mathcal{S}^{(i)}, \mathcal{X}) \quad -(*)$$

であるから、Gauss-Wishart分布の推論でやったように、

$\boldsymbol{\mu}_k$ の分布を求めたあとに $\boldsymbol{\Lambda}_k$ の分布を求める。

・ $\log p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k \mid \mathcal{S}^{(i)}, \mathcal{X}) \quad \left(= \log p(\boldsymbol{\mu}_k \mid \boldsymbol{\Lambda}_k, \mathcal{S}^{(i)}, \mathcal{X}) + \text{const.} \right)$

$$= -\frac{1}{2} \sum_{n=1}^{N} s_{n,k}^{(i)} \Big((\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Lambda}_k (\boldsymbol{x}_n - \boldsymbol{\mu}_k) \underbrace{- \log(\det \boldsymbol{\Lambda}_k)}_{\text{const.}} \Big)$$

$$+ \log \mathcal{N}(\boldsymbol{\mu}_k \mid \boldsymbol{m}, (\beta \boldsymbol{\Lambda}_k)^{-1}) \underbrace{+ \log \mathcal{W}(\boldsymbol{\Lambda}_k \mid \nu, \boldsymbol{W})}_{\text{const.}} + \text{const.}$$

$$= -\frac{1}{2} \sum_{n=1}^{N} s_{n,k}^{(i)} (\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Lambda}_k (\boldsymbol{x}_n - \boldsymbol{\mu}_k)$$

$$-\frac{1}{2} \Big(\beta (\boldsymbol{\mu}_k - \boldsymbol{m})^T \boldsymbol{\Lambda}_k (\boldsymbol{\mu}_k - \boldsymbol{m}) \underbrace{- \log(\det(\beta \boldsymbol{\Lambda}_k))}_{\text{const.}} \Big) + \text{const.}$$

$$= -\frac{1}{2} \Big(\sum_{n=1}^{N} s_{n,k}^{(i)} \big(\underbrace{\boldsymbol{x}_n^T \boldsymbol{\Lambda}_k \boldsymbol{x}_n}_{\text{const.}} - 2\boldsymbol{\mu}_k^T \boldsymbol{\Lambda}_k \boldsymbol{x}_n + \boldsymbol{\mu}_k^T \boldsymbol{\Lambda}_k \boldsymbol{\mu}_k \big)$$

$$\beta \big(\boldsymbol{\mu}_k^T \boldsymbol{\Lambda}_k \boldsymbol{\mu}_k - 2\boldsymbol{\mu}_k^T \boldsymbol{\Lambda}_k \boldsymbol{m} \underbrace{+ \boldsymbol{m}^T \boldsymbol{\Lambda}_k \boldsymbol{m}}_{\text{const.}} \big) \Big) + \text{const.}$$

$$= -\frac{1}{2} \Big(\boldsymbol{\mu}_k^T \Big(\sum_{n=1}^{N} s_{n,k}^{(i)} + \beta \Big) \boldsymbol{\Lambda}_k \boldsymbol{\mu}_k - 2\boldsymbol{\mu}_k^T \boldsymbol{\Lambda}_k \Big(\sum_{n=1}^{N} s_{n,k}^{(i)} \boldsymbol{x}_n + \beta \boldsymbol{m} \Big) \Big) + \text{const.}$$

$$\therefore \boldsymbol{\mu}_k^{(i)} \sim p(\boldsymbol{\mu}_k \mid \boldsymbol{\Lambda}_k, \mathcal{S}^{(i)}, \mathcal{X}) = \mathcal{N}(\boldsymbol{\mu}_k \mid \hat{\boldsymbol{m}}_k, (\hat{\beta}_k \boldsymbol{\Lambda}_k)^{-1}),$$

$$\hat{\beta}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + \beta, \quad \hat{\boldsymbol{m}}_k := \hat{\beta}_k^{-1} \Big(\sum_{n=1}^{N} s_{n,k}^{(i)} \boldsymbol{x}_n + \beta \boldsymbol{m} \Big).$$

・ $p(\boldsymbol{\Lambda}_k \mid \mathcal{S}^{(i)}, \mathcal{X})$ を求める。(*) より

$\log p(\boldsymbol{\Lambda}_k \mid \mathcal{S}^{(i)}, \mathcal{X})$

$$= \log p(\mu_k, \Lambda_k \mid S^{(i)}, \mathcal{X}) - \log p(\mu_k \mid \Lambda_k, S^{(i)}, \mathcal{X})$$

$$= -\frac{1}{2}\sum_{n=1}^{N} s_{n,k}^{(i)}\left((x_n-\mu_k)^T\Lambda_k(x_n-\mu_k) - \log(\det\Lambda_k)\right)$$

$$-\frac{1}{2}\left(\beta(\mu_k-m)^T\Lambda_k(\mu_k-m) - \log(\det(\beta\Lambda_k))\right)$$

前ページの式変形を用いて計算

$$+\log\mathcal{W}(\Lambda_k \mid \nu, W) - \log\mathcal{N}(\mu_k \mid \hat{m}_k, (\hat{\beta}_k\Lambda_k)^{-1}) + \text{const.}$$

$$= -\frac{1}{2}\left(\mu_k^T\left(\sum_{n=1}^{N} s_{n,k}^{(i)} + \beta\right)\Lambda_k\mu_k - 2\mu_k^T\Lambda_k\left(\sum_{n=1}^{N} s_{n,k}^{(i)}x_n + \beta m\right)\right)$$

$$-\frac{1}{2}\sum_{n=1}^{N} s_{n,k}^{(i)} x_n^T\Lambda_k x_n - \frac{1}{2}\beta m^T\Lambda_k m$$

$$+\frac{1}{2}\sum_{n=1}^{N} s_{n,k}^{(i)}\log(\det\Lambda_k) + \frac{1}{2}\log(\det\Lambda_k) + \underbrace{\frac{1}{2}D\log\beta}_{\text{const.}}$$

$$+\frac{\nu-D-1}{2}\log(\det\Lambda_k) - \frac{1}{2}\mathrm{tr}(W^{-1}\Lambda_k)$$ ← $\log\mathcal{W}(\Lambda_k \mid \nu, W)$ から

$$+\frac{1}{2}\left((\mu_k-\hat{m}_k)^T(\hat{\beta}_k\Lambda_k)(\mu_k-\hat{m}_k) - \log(\det(\hat{\beta}_k\Lambda_k))\right) + \text{const.}$$

$$= -\frac{1}{2}\left(\cancel{\mu_k^T(\hat{\beta}_k\Lambda_k)\mu_k - 2\mu_k^T(\hat{\beta}_k\Lambda_k)\hat{m}_k}\right)$$

$$-\frac{1}{2}\left(\mathrm{tr}\left(\sum_{n=1}^{N} s_{n,k}^{(i)} x_n x_n^T\Lambda_k\right) + \mathrm{tr}(\beta m m^T\Lambda_k)\right)$$

$$+\frac{1}{2}\sum_{n=1}^{N} s_{n,k}^{(i)}\log(\det\Lambda_k) + \cancel{\frac{1}{2}\log(\det\Lambda_k)}$$

$$+\frac{\nu-D-1}{2}\log(\det\Lambda_k) - \frac{1}{2}\mathrm{tr}(W^{-1}\Lambda_k)$$

$$+\frac{1}{2}\left(\cancel{\mu_k^T(\hat{\beta}_k\Lambda_k)\mu_k - 2\mu_k^T(\hat{\beta}_k\Lambda_k)\hat{m}_k} + \hat{m}_k^T(\hat{\beta}_k\Lambda_k)\hat{m}_k\right)$$

$$\cancel{-\frac{1}{2}\log(\det\Lambda_k)} - \underbrace{\frac{1}{2}D\log\hat{\beta}_k}_{\text{const.}} + \text{const.}$$

$$= \frac{1}{2}\left(\sum_{n=1}^{N} s_{n,k}^{(i)} + \nu - D - 1\right)\log(\det \Lambda_k)$$

$$-\frac{1}{2}\mathrm{tr}\left(\left(\sum_{n=1}^{N} s_{n,k}^{(i)} \mathbf{x}_n \mathbf{x}_n^T + \beta \mathbf{m}\mathbf{m}^T - \hat{\beta}_k \hat{\mathbf{m}}_k \hat{\mathbf{m}}_k^T + W^{-1}\right)\Lambda_k\right) + \text{const.}$$

$$\therefore \Lambda_k^{(i)} \sim p(\Lambda_k \mid S^{(i)}, \mathcal{X}) = \mathcal{W}(\Lambda_k \mid \hat{\nu}_k, \hat{W}_k),$$

$$\hat{W}_k^{-1} := \sum_{n=1}^{N} s_{n,k}^{(i)} \mathbf{x}_n \mathbf{x}_n^T + \beta \mathbf{m}\mathbf{m}^T - \hat{\beta}_k \hat{\mathbf{m}}_k \hat{\mathbf{m}}_k^T + W^{-1},$$

$$\hat{\nu}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + \nu.$$

・以上より，各 k に対して，

$$\Lambda_k^{(i)} \sim \mathcal{W}(\Lambda_k \mid \hat{\nu}_k, \hat{W}_k) \quad \text{としてから，}$$

$$\boldsymbol{\mu}_k^{(i)} \sim \mathcal{N}(\boldsymbol{\mu}_k \mid \hat{\mathbf{m}}_k, (\hat{\beta}_k \Lambda_k^{(i)})^{-1})$$

とすることでサンプリングできる．

。 $$\boldsymbol{\pi}^{(i)} \sim p(\boldsymbol{\pi} \mid S^{(i)}, \mathcal{X}) \propto p(S^{(i)} \mid \boldsymbol{\pi}) p(\boldsymbol{\pi})$$

であるが，これは Poisson 混合モデルのときに計算済み：

$$\boldsymbol{\pi}^{(i)} \sim \mathrm{Dir}(\boldsymbol{\pi} \mid \hat{\boldsymbol{\alpha}}), \ \hat{\alpha}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + \alpha_k \ (k = 1, \cdots, K).$$

。以上より，Gibbs サンプリングのアルゴリズムは次のとおり：

$\boldsymbol{\mu}^{(0)}, \Lambda^{(0)}, \boldsymbol{\pi}^{(0)}$ をランダムに生成；

for $i = 1, \cdots, N_{max}$ do

  for $n = 1, \cdots, N$ do

$$\eta_{n,k} \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(i-1)})^\top \Lambda_k^{(i-1)} (\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(i-1)}) + \frac{1}{2}\log(\det \Lambda_k^{(i-1)}) + \log \pi_k\right) \ (k = 1, \cdots, K)$$

$\sum_{k=1}^{K} \eta_{n,k} = 1$ となる η_n をつくる;

$$\boldsymbol{s}_n^{(i)} \sim \mathrm{Cat}(\boldsymbol{s}_n \mid \eta_n)$$

end for;

for $k = 1, \cdots, K$ do

$$\hat{\beta}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + \beta; \quad \hat{\boldsymbol{m}}_k := \hat{\beta}_k^{-1}\left(\sum_{n=1}^{N} s_{n,k}^{(i)} \boldsymbol{x}_n + \beta \boldsymbol{m}\right);$$

$$\hat{W}_k^{-1} := \sum_{n=1}^{N} s_{n,k}^{(i)} \boldsymbol{x}_n \boldsymbol{x}_n^\top + \beta \boldsymbol{m}\boldsymbol{m}^\top - \hat{\beta}_k \hat{\boldsymbol{m}}_k \hat{\boldsymbol{m}}_k^\top + W^{-1};$$

$$\hat{\nu}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + \nu;$$

$$\Lambda_k^{(i)} \sim \mathcal{W}(\Lambda_k \mid \hat{\nu}_k, \hat{W}_k);$$

$$\boldsymbol{\mu}_k^{(i)} \sim \mathcal{N}(\boldsymbol{\mu}_k \mid \hat{\boldsymbol{m}}_k, (\hat{\beta}_k \Lambda_k^{(i)})^{-1})$$

end for;

$\hat{\alpha}_k := \sum_{n=1}^{N} s_{n,k}^{(i)} + \alpha_k \ (k = 1, \cdots, K)$ となる $\hat{\boldsymbol{\alpha}}$ をつくる;

$$\boldsymbol{\pi}^{(i)} \sim \mathrm{Dir}(\boldsymbol{\pi} \mid \hat{\boldsymbol{\alpha}})$$

end for.