Variational Bayes Derivation of Poisson Mixture Model

@Shusei-E

September 13, 2016

1 Model

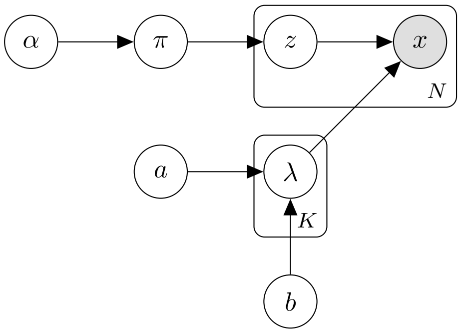

Figure 1: Poisson Mixture Model

Variables:

- N: the number of data
- D: dimension
- K: the number of cluster
- x: data, $D \times N$ matrix
- z: latent category, $K \times N$ matrix
- λ: average parameter of Poisson distribution, $K \times D$ (each cluster has D parameters)
- π: mixture coefficient, K dimension

Data Generating Process:

$$p(\pi) = \mathrm{Dir}(\pi|\alpha) \tag{1}$$

$$p(\lambda) = \prod_{k=1}^{K}\prod_{d=1}^{D} \mathrm{Gam}(\lambda_d^{(k)}|a,b) \tag{2}$$

$$p(z|\pi) = \prod_{n=1}^{N} \mathrm{Cat}(z_n|\pi) \tag{3}$$

$$p(x|z,\lambda) = \prod_{n=1}^{N}\prod_{k=1}^{K}\prod_{d=1}^{D} \mathrm{Poi}(x_{n,d}|\lambda_d^{(k)})^{z_n^{(k)}} \tag{4}$$

K-of-1 encoding is used for z_n. In Equation (4), z_n takes 1 for a certain k, so z_n is a "switch" for Poisson distribution.

2 Derivation

2.1 Induced Factorization

Assuming $q(A, B, C) = q(A, B)q(C)$, $q(A, B) = q(A)q(B)$ is true if $A \perp\!\!\!\perp B | X, C$. In this case, we can say $\lambda \perp\!\!\!\perp \pi | x, z$ from the graphical model, so assuming $p(z, \lambda, \pi | x) \approx q(z)q(\lambda, \pi)$, induced factorization leads to $q(z)q(\lambda)q(\pi)$.

2.2 Joint Distribution

$$
\begin{aligned}
p(x, \pi, z, \lambda | \alpha, a, b) &= p(x | \pi, z, \lambda, \alpha, a, b) p(\pi, z, \lambda | \alpha, a, b) && (5)\\
&= p(x | \pi, z, \lambda, \alpha, a, b) p(z | \pi, \lambda, \alpha, a, b) p(\lambda, \pi | \alpha, a, b) && (6)\\
&= p(x | z, \lambda) p(z | \pi) p(\lambda | a, b) p(\pi | \alpha) \qquad (\because \text{graphical model}) && (7)\\
&= \left[\prod_{n=1}^{N} p(x_i | \lambda_{z_i = k}) p(z_i = k | \pi) \right] \left[\prod_{k=1}^{K} p(\lambda_k | a, b) \right] p(\pi | \alpha) && (8)
\end{aligned}
$$

2.3 Evidence lower bound

Keep in mind that from factorization assumption (and induced factorization):

$$
p(z, \lambda, \pi | x) \approx q(z)q(\lambda)q(\pi) \tag{9}
$$

Evidence lower bound (ELBO) is:

$$
\begin{aligned}
F[q(\pi, z, \lambda)] &= \iint \sum_{z} q(z)q(\lambda)q(\pi) \log \frac{p(x|z, \lambda)p(z|\pi)p(\lambda|a, b)p(\pi|\alpha)}{q(z)q(\lambda)q(\pi)} d\pi d\lambda && (10)\\
&= \iint \sum_{z} q(z)q(\lambda)q(\pi) \log \{p(x|z, \lambda)p(z|\pi)\} d\pi d\lambda \\
&\quad - \sum_{z} q(z) \log q(z) + \int q(\pi) \log \frac{p(\pi|\alpha)}{q(\pi)} d\pi + \int q(\lambda) \log \frac{p(\lambda|a, b)}{q(\lambda)} d\lambda && (11)\\
&= \iint \sum_{i=1}^{N} \sum_{k=1}^{K} q(z_i = k) q(\lambda_k) q(\pi) \log \left\{ p(x_i^{(k)} | \lambda_{z_i = k}) p(z_i = k | \pi) \right\} d\pi d\lambda \\
&\quad - \sum_{i=1}^{K} \sum_{k=1}^{K} q(z_i = k) \log q(z_i = k) + \underbrace{\int q(\pi) \log \frac{p(\pi|\alpha)}{q(\pi)} d\pi}_{\mathrm{KL}[q(\pi)||p(\pi|\alpha)]} + \sum_{k=1}^{K} \underbrace{\int q(\lambda_k) \log \frac{p(\lambda_k|a, b)}{q(\lambda_k)} d\lambda_k}_{\mathrm{KL}[q(\lambda_k)||p(\lambda_k|a,b)]} && (12)
\end{aligned}
$$

3 Update Equation of $q(\pi)$

$$
\widetilde{F}[q(\pi)] = \int \sum_{i=1}^{N} q(\pi) \log \{p(z_i = k|\pi)\} d\pi + \int q(\pi) \log \frac{p(\pi|\alpha)}{q(\pi)} d\pi \tag{13}
$$

Using variational inference,

$$\frac{\delta \widetilde{F}[q(\pi)]}{\delta q(\pi)} = \frac{\partial \widetilde{F}[q(\pi)]}{\partial q(\pi)} = \sum_{i=1}^{N} \log p(z_i = k|\pi) + \log \frac{p(\pi|\alpha)}{q(\pi)} - 1 = 0 \tag{14}$$

$$\iff \qquad \log q(\pi) \propto \sum_{i=1}^{N} \log p(z_i = k|\pi) + \log p(\pi|\alpha) \tag{15}$$

$$= \log \left[\left\{ \prod_{i=1}^{N} p(z_i = k|\pi) \right\} p(\pi|\alpha) \right] \tag{16}$$

$$\iff \qquad q(\pi) \propto \exp \left[\log \left\{ \prod_{k=1}^{K} \pi_k^{\alpha_k - 1} \prod_{i=1}^{N} p(z_i = k|\pi) \right\} \right] \tag{17}$$

$$= \exp \left[\log \left\{ \prod_{k=1}^{K} \pi_k^{\alpha_k - 1} \prod_{i=1}^{N} \pi_k^{z_i = k} \right\} \right] \tag{18}$$

$$= \exp \left[\sum_{i=1}^{N} \sum_{k=1}^{K} \log(\pi_k)^{(\alpha_k - 1)(z_i = k)} \right] \tag{19}$$

$$= \prod_{k=1}^{K} \pi_k^{\alpha_k - 1 + \sum_{i=1}^{N} (z_i = k)} \tag{20}$$

$$= \mathrm{Dir}(\pi|\hat{\alpha}), \qquad \hat{\alpha} = \alpha_k + \sum_{i=1}^{N} (z_i = k) \tag{21}$$

Note:

$$\sum_{i=1}^{3} \log a^{l_i} + \log_a^n = \log a^{l_1} + \log a^{l_2} + \log a^{l_3} + \log a^n \tag{22}$$

$$= \log a^{l_1 + l_2 + l_3 + n} \tag{23}$$

$$= \log a^{\sum_{i=1}^{3} l_i + n} \tag{24}$$

4 Update Equation of $q(\lambda)$

$$\widetilde{F}[q(\lambda_k)] = \int \sum_{i=1}^{N} q(\lambda_k) \log \left\{ p(x_i^{(k)}|\lambda_{z_i = k}) \right\} d\lambda_k + \int q(\lambda_k) \log \frac{p(\lambda_k|a, b)}{q(\lambda_k)} d\lambda_k \tag{25}$$

Recall distributions:

$$\mathrm{Gam}(\lambda_d^{(k)}|a, b) = \frac{b^a}{\Gamma(a)} (\lambda_d^{(k)})^{a-1} \exp(-b\lambda_d^{(k)}) \tag{26}$$

$$\mathrm{Poi}(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \tag{27}$$

Using variational inference,

$$\frac{\delta\widetilde{F}[q(\lambda_k)]}{\delta q(\lambda_k)} = \frac{\partial\widetilde{F}[q(\lambda_k)]}{\partial q(\lambda_k)} = \sum_{i=1}^{N} \log p(x_i^{(k)}|\lambda_{z_i=k}) + \log\frac{p(\lambda_k|a,b)}{q(\lambda_k)} - 1 = 0 \tag{28}$$

$$\Longleftrightarrow \qquad \log q(\lambda_k) \propto \sum_{i=1}^{N} \log p(x_i^{(k)}|\lambda_{z_i=k}) + \log p(\lambda_k|a,b) \tag{29}$$

$$\Longleftrightarrow \qquad q(\lambda_k) \propto (\lambda_k)^{a-1}\exp(-b\lambda_k)\exp\left[\sum_{i=1}^{N}\log p(x_i^{(k)}|\lambda_{z_i=k})\right] \tag{30}$$

$$= \exp\left[\log\lambda_k^{a-1}\right]\exp(-b\lambda_k)\exp\left[\sum_{i=1}^{N}\log e^{-\lambda_k}\lambda_k^{x_i^{(k)}}\right] \tag{31}$$

$$= \exp\left[\log\lambda_k^{a-1} - b\lambda_k + \sum_{i=1}^{N}\log\lambda_k^{x_i^{(k)}} + \sum_{i=1}^{N}\log e^{-\lambda_{z_i=k}}\right] \tag{32}$$

$$= \exp\left[\log\lambda_k^{a+\sum_{i=1}^{n}x_i^{(k)}-1} + \left\{-\left(b+\sum_{i=1}^{N}(z_i=k)\right)\lambda_k\right\}\right] \tag{33}$$

Recall $x_i^{(k)}$ infers we need to consider a "switch" for category, that is z.

5 Update Equation of $q(z)$

$$\widetilde{F}[q(z)] = \sum_{i=1}^{N} q(z_i=k)\log\left\{p(x_i^{(k)})p(z_i=k|\pi)\right\} - \sum_{i=1}^{N} q(z_i=k)\log q(z_i=k) \tag{34}$$

Using variational inference,

$$\frac{\delta\widetilde{F}[q(z)]}{\delta q(z)} = \frac{\partial\widetilde{F}[q(z)]}{\partial q(z)} = \log\left\{p(x_i^{(k)})p(z_i=k|\pi)\right\} - \log q(z_i=k) - 1 = 0 \tag{35}$$

$$\Longleftrightarrow \qquad \log q(z_i=k) \propto p(x_i^{(k)}|\lambda_{z_i=k})p(z_i=k|\pi) \tag{36}$$

$$= \left(e^{-\lambda_k}\lambda_{z_i=k}^{x_i^{(k)}}\right)\left(\prod_{k=1}^{K}\pi_k^{z_i=k}\right) \tag{37}$$

$$= \prod_{d=1}^{D}\left(e^{-\lambda_{k,d}}\lambda_{z_i=k,d}^{x_{i,d}^{(k)}}\right)\left(\prod_{k=1}^{K}\pi_k^{z_i=k}\right) \tag{38}$$

$$= \pi_k\exp\left[\sum_{d=1}^{D}\left\{-\lambda_{k,d} + x_{i,d}^{(k)}\log\lambda_{k,d}\right\}\right] \tag{39}$$

Note that $\prod_{k=1}^{K}\pi_k^{z_i=k} = \pi_k$.

Reference

1. sammy-suyama「ベイズ混合モデルにおける近似推論 1 ～変分近似～」 `http://machine-learning.hatenablog.com/entry/2016/07/06/200605`