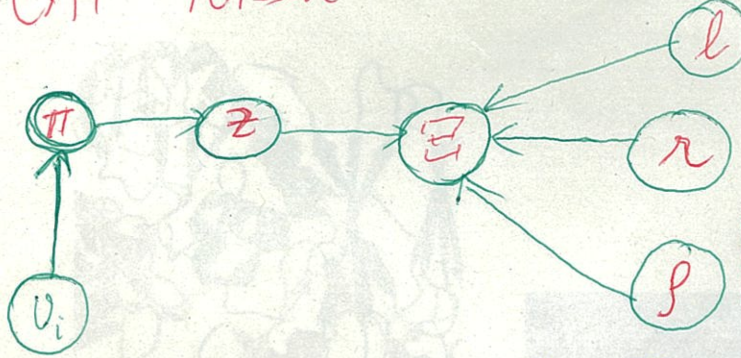
CAT – Poisson



$Q_{ab} = \pi(b)\rho_{ab}$ (1)

$$P(z, \pi, l, \lambda, \rho \mid \Xi) = \prod_i p_i(\lambda) \prod_j p(l) \prod_{ij} p(\Xi)\, p(z_l \mid z_{l-1}, \alpha)$$

$$p(\Xi, z, \boxed{\pi}, l, \lambda, \boxed{\rho}) = P(\Xi \mid z, \boxed{\pi}, l, \lambda, \rho)\, p(z \mid \pi)\, p(\pi)\, p(l)\, p(\lambda) \boxed{p(\rho)}$$

~~$v_k \sim Beta(1, k)$~~

~~$\pi_k = v_k \prod_{j<k}(1 - v_j)$~~

$$p(l) = \prod_j \gamma_{1,\beta}(l_j) = \prod \frac{\beta^1}{\Gamma(1)} l_j^{1-1} e^{-\beta \cdot l_j} = \prod_j \beta e^{-\beta \cdot l_j}$$

$$p(\lambda) = \prod_i \gamma_{\alpha,\alpha}(\lambda) = \prod_i \frac{\alpha^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} \cdot e^{-\alpha\lambda_i}$$

$$\otimes\ p(\rho) = \prod \gamma_{1,1}(\rho) = \prod \frac{1^1}{\Gamma(1)} \rho^{1-1} \cdot e^{-1\cdot\rho} = \prod e^{-\rho}$$

$$p(\Xi \mid \lambda_i, l_j) = Possion(\Xi_i \mid \lambda_i, l_j) = \frac{\lambda_i l_j^{n_{ij}} e^{-\lambda_i l_j}}{n_{ij}!}$$

$z_l \mid \pi_{1:K} \sim Discrete(\pi_1, \pi_2, \ldots, \pi_K)$

$\pi_1, \pi_2, \ldots, \pi_K \sim Dirichelet(\alpha/K, \ldots, \alpha/K)$

$$q(z, \pi, l, \lambda, \boxed{\rho}) = \left[\prod_{ij} \prod_{l=1}^{K} q(z_{l})\right] \left[\prod_{l=1}^{K} q(\pi)\right] \left[\prod_i q(\lambda)\right] \left[\prod_j q(l)\right] \boxed{\left[\prod_{ij} q(\rho)\right]}$$

ELBO

$$\mathcal{L} = \sum_z \int_\pi \int_l \int_\lambda \boxed{\int_\rho} q(z, \pi, l, \lambda, \boxed{\rho}) \ln \left\{ \frac{p(\Xi, z, \pi, l, \lambda, \boxed{\rho})}{q(z, \pi, l, \lambda, \boxed{\rho})} \right\} d\pi\, dl\, d\lambda \boxed{d\rho}$$

$$= E_q[\log p(\Xi, z, \pi, l, \lambda, \boxed{\rho})] - E[\log q(z, \pi, l, \lambda, \boxed{\rho})]$$

$$= E_q[\log p(\Xi \mid z, \pi, l, \lambda) + \log p(z \mid \pi) + \log p(\pi) + \log p(l) + \log p(\lambda)] - E_q[\log q(z, \pi, l, \lambda)]$$

$$p(\varphi \mid \Xi) = \frac{p(\Xi, \varphi)}{\int p(\Xi, \varphi)\, d\varphi}$$

Jensen's inequality

$$\log \int p(\Xi, \varphi)\, d\varphi = \log \int q(\varphi \mid \nu) \frac{p(\Xi, \varphi)}{q(\varphi \mid \nu)}\, d\varphi = \log E_{q(\varphi \mid \nu)} \frac{p(\Xi, \varphi)}{q(\varphi, \nu)} \geq E_{q(\varphi \mid \nu)} \log \frac{p(\Xi, \varphi)}{q(\varphi, \nu)}$$

$$\geq \underbrace{\int q(\varphi \mid \nu) \log \frac{p(\Xi, \varphi)}{q(\varphi, \nu)}\, d\varphi}_{\text{Evidence Lower Bound (ELBO)}}$$

Target: Maximize ELBO respect to ν

Solution: Use stochastic gradient opimization to maximize ELBO.

parameter $(z, \pi, l, \lambda) = \varphi$

varitional parameter $(\tilde{\alpha}_\pi ; \phi_{le} ; \tilde{\alpha}_\lambda ; \tilde{\beta}_\lambda ; \tilde{\alpha}_l ; \tilde{\beta}_l) = r$

ELBO

$$\mathcal{L} = \int q(\varphi|r) \log \frac{p(\Xi ; \varphi)}{q(\varphi|r)} d(\varphi)$$

$$\nabla_r \mathcal{L} = \nabla_r \int q(\varphi|r) \log \frac{p(\Xi ; \varphi)}{q(\varphi|r)} d(\varphi)$$

$$\nabla_r \mathcal{L} = \nabla_r \int [\log p(\Xi, \varphi) - \log q(\varphi|r)] q(\varphi|r) d(\varphi)$$

$$\nabla_r \mathcal{L} = \int \nabla_r [\log p(\Xi, \varphi) - \log q(\varphi|r)] q(\varphi|r) d(\varphi)$$

$$\nabla_r \mathcal{L} = \int \nabla_r [\log p(\Xi, \varphi) - \log q(\varphi|r)] q(\varphi|r) d\varphi$$

$$\nabla_r \mathcal{L} = \int \nabla_r [\log p(\Xi, \varphi) - \log q(\varphi|r)] d\varphi$$

$$+ \int [\nabla_r q(\varphi|r)] [\log p(\Xi, \varphi)]$$

$$= -E_{q(\varphi|r)} [\nabla_r \log q(\varphi|r)]$$

$$+ \int [\nabla_r q(\varphi|r)] [\log p(\Xi, \varphi) - \log q(\varphi|r)] d\varphi$$

$$= \int [\nabla_r \log q(\varphi|r) . q(\varphi|r)] [\log p(\Xi, \varphi) - \log q(\varphi|r)] d\varphi$$

$$= \int [\nabla_r \log q(\varphi|r)] [\log p(\Xi, \varphi) - \log q(\varphi|r)] q(\varphi|r) d\varphi$$

$$= E_q [[\nabla_r \log q(\varphi|r)] [\log p(\Xi, \varphi) - \log q(\varphi|r)]]$$

$$-E_{q(\varphi|r)} [\nabla_r \log q(\varphi|r)] = \quad (3)$$

$$-E_{q(\varphi|r)} \left[\frac{\nabla_r q(\varphi|r)}{q(\varphi|r)} \right] =$$

$$- \int \nabla_r q(\varphi|r) d\varphi = -\nabla_r \int q(\varphi|r) d\varphi$$

$$= -\nabla_r 1 = 0$$

$$\boxed{\nabla_r \log q(\varphi|r) = \frac{\nabla_r q(\varphi|r)}{q(\varphi|r)}}$$

$$\boxed{E_{q(\varphi|r)} [\nabla_r \log q(\varphi|r)] = 0}$$

$\boxed{\text{The score junction trick}}$

$$\log q(z|\phi) = \sum_{ij}\sum_{l=1}^{k} z_{ijk} \ln \phi_{ijk} \qquad \frac{d\log q(z|\phi)}{d\phi} = \sum_{ij}\sum_{l=1}^{k} z_{ijk} \cdot \frac{1}{\phi_{ijk}}$$

$$\log q(\lambda|\tilde{\alpha}_\lambda, \tilde{\beta}_\lambda) = \log\left(\prod_i \frac{\tilde{\beta}_\lambda^{\tilde{\alpha}_\lambda}}{\Gamma(\tilde{\alpha}_\lambda)} \lambda_i^{\tilde{\alpha}_\lambda - 1} e^{-\tilde{\beta}_\lambda \lambda}\right) \qquad \frac{d\log q(\lambda|\tilde{\alpha}_\lambda, \tilde{\beta}_\lambda)}{d\tilde{\alpha}_\lambda} = \sum_i \cancel{\ } + \log(\lambda_i) - \psi(\tilde{\alpha}_\lambda) + \log(\tilde{\beta}_\lambda)$$

$$= \sum_i (-\tilde{\beta}_\lambda \lambda) + (\tilde{\alpha}_\lambda - 1)\log(\lambda_i) + \qquad \frac{d\log q(\lambda|\tilde{\alpha}_\lambda, \tilde{\beta}_\lambda)}{d\tilde{\beta}_\lambda} = \sum_i -\lambda_i + \frac{\tilde{\alpha}_\lambda}{\tilde{\beta}_\lambda}$$

$$- \log\Gamma(\tilde{\alpha}_\lambda) + \tilde{\alpha}_\lambda \log(\tilde{\beta}_\lambda)$$

$$\log q(l|\tilde{\alpha}_l, \tilde{\beta}_l) = \log\left(\prod_j \frac{\tilde{\beta}_l^{\tilde{\alpha}_l}}{\Gamma(\tilde{\alpha}_l)} l^{\tilde{\alpha}_l - 1} e^{-\tilde{\beta}_l l}\right) \qquad \frac{d\log q(l|\tilde{\alpha}_l, \tilde{\beta}_l)}{d\tilde{\alpha}_l} = \sum_j \log(l_j) + \log\tilde{\beta}_l - \psi(\tilde{\alpha}_l)$$

$$= \sum_j -\tilde{\beta}_l l + (\tilde{\alpha}_l - 1)\log(l) + \qquad \frac{d\log q(l|\tilde{\alpha}_l, \tilde{\beta}_l)}{d\tilde{\beta}_l} = \sum_j -l_j + \frac{\tilde{\alpha}_l}{\tilde{\beta}_l}$$

$$\tilde{\alpha}_l \log\tilde{\beta}_l - \log\Gamma(\tilde{\alpha}_l)$$

$$\log q(\pi|\tilde{\alpha}_\pi) = \sum_{l=1}^{K}\left(\frac{\tilde{\alpha}_{\pi l}}{K} - 1\right)\log\pi_l + \log\Gamma\left(\sum_{l=1}^{K}\frac{\tilde{\alpha}_{\pi l}}{K}\right) - \sum_{l=1}^{K}\log\Gamma\left(\frac{\tilde{\alpha}_{\pi l}}{K}\right)$$

$$\frac{d\log q(\pi|\tilde{\alpha}_\pi)}{d\tilde{\alpha}_\pi} = \sum_{l=1}^{K}\frac{1}{K}\cdot\log\pi_l + \psi\left(\sum_{l=1}^{K}\frac{\tilde{\alpha}_{\pi l}}{K}\right) - \psi\left(\frac{\tilde{\alpha}_\pi}{K}\right)$$

$$q(z,\pi,l,\lambda) = \left[\prod_{ij}\prod_{l=1}^{K} q(z)\right]\left[\prod_{l=1}^{k} q(\pi)\right]\left[\prod_{i} q(\lambda)\right]\left[\prod_{j} q(l)\right]$$

$q(\pi) \sim \text{Dirichlet}\left(\tilde{\alpha}_\pi/K, \tilde{\alpha}_\pi/K, \dots, \tilde{\alpha}_\pi/K\right)$

$q(z_k) \sim \text{Discrete}(\phi_k)$

Variational parameter $(\tilde{\alpha}_\pi; \tilde{\phi}_k; \tilde{\alpha}_\lambda, \tilde{\beta}_\lambda; \tilde{\alpha}_l, \tilde{\beta}_l) = v$

$q(\lambda) = \prod_i \gamma_{\tilde{\alpha}_\lambda, \tilde{\alpha}_\lambda}(\lambda)$

$\mathcal{L}(v) = E_q[\log p(z,\pi,l,\lambda,\Xi)] - E_q[\log q(z,\pi,l,\lambda)]$

$q(l) = \prod_j \gamma_{\tilde{\alpha}_l, \tilde{\beta}_l}(l)$

use stochastic optimization to maximize $\mathcal{L}(v)$ respect to v

Monte Carlo to approximate $\mathcal{L}(v)$

for $s = 1$ to S do

$$\mathcal{L}(v) \approx \frac{1}{S}\sum_{s=1}^{S}\left[\log p(z_s,\pi_s,l_s,\lambda_s,\Xi_s) - \log q(z_s,\pi_s,l_s,\lambda_s)\right]$$

$$\nabla_v \mathcal{L}(v) \simeq \frac{1}{S}\sum_{s=1}^{S}\left[\nabla_v \log q(z_s,\pi_s,l_s,\lambda_s|v)\right]\left[\log p(z_s,\pi_s,l_s,\lambda_s,\Xi_s) - \log q(z_s,\pi_s,l_s,\lambda_s|v)\right]$$

$$\nabla_{\tilde{\alpha}_\pi}\mathcal{L} \simeq \frac{1}{S}\sum_{s=1}^{S}\left[\nabla_{\tilde{\alpha}_\pi}\log q(\pi|\tilde{\alpha}_\pi)\right]\left[\log p\left(\pi_s \middle| \frac{\alpha_\pi}{K}\right) - \log q\left(\pi_s \middle| \frac{\tilde{\alpha}_\pi}{K}\right) + \sum_{l=1}^{K}\left(\log p(z_{ls}|\pi_s) + \log p(\Xi_l | z_{ls}, \pi_{ls}, l_{ls}, \lambda_{ls})\right)\right]$$

$$\nabla_{\phi_\ell} \mathcal{L} \simeq \frac{1}{S}\sum_{s=1}^{S}\left[\nabla_{\phi_\ell} q(z_{\ell s}|\phi_i)\right]\left[\log p(z_{\ell s}|\phi_{\ell s}) + \log p(z_{\ell s}|\pi_s) + p(\Xi_{\ell s}|z_{\ell s}, \pi_{\ell s}, l_{\ell s}, \lambda_{\ell s})\right]$$

$$\nabla_{\tilde{\alpha}_\lambda} \mathcal{L} \simeq \frac{1}{S}\sum_{s=1}^{S}\left[\nabla_{\tilde{\alpha}_\lambda} q(\lambda_s|\tilde{\alpha}_{s\lambda}, \tilde{\beta}_{s\lambda})\right]\left[\log p(\lambda_s|\alpha_{s\lambda}, \beta_{s\lambda}) - \log q(\lambda_s|\tilde{\alpha}_{s\lambda}, \tilde{\beta}_{s\lambda}) + p(\Xi_s|z_s, \pi_s, l_s, \lambda_s)\right]$$

$$\nabla_{\tilde{\beta}_\lambda} \mathcal{L} \simeq \frac{1}{S}\sum_{s=1}^{S}\left[\nabla_{\tilde{\beta}_\lambda} q(\lambda_s|\tilde{\alpha}_{s\lambda}, \tilde{\beta}_{s\lambda})\right]\left[\log p(\lambda_s|\alpha_{s\lambda}, \beta_{s\lambda}) - \log q(\lambda_s|\tilde{\alpha}_{s\lambda}, \tilde{\beta}_{s\lambda}) + p(\Xi_s|z_s, \pi_s, l_s, \lambda_s)\right]$$

$$\nabla_{\tilde{\alpha}_\ell} \mathcal{L} \simeq \frac{1}{S}\sum_{s=1}^{S}\left[\nabla_{\tilde{\alpha}_\ell} q(l_s|\tilde{\alpha}_{\ell s}; \tilde{\beta}_{\ell s})\right]\left[\log p(l_s|\alpha_{\ell s}, \beta_{\ell s}) - \log q(l_s|\tilde{\alpha}_{\ell s}, \tilde{\beta}_{\ell s}) + p(\Xi_s|\pi_s, z_s, l_s, \lambda_s)\right]$$

$$\nabla_{\tilde{\beta}_\ell} \mathcal{L} = \frac{1}{S}\sum_{s=1}^{S}\left[\nabla_{\tilde{\beta}_\ell} q(l_s|\tilde{\alpha}_{\ell s}; \tilde{\beta}_{\ell s})\right]\left[\log p(l_s|\alpha_{\ell s}, \beta_{\ell s}) - \log q(l_s|\tilde{\alpha}_{\ell s}, \tilde{\beta}_{\ell s}) + p[\Xi_s|\pi_s, z_s, l_s, \lambda_s]\right]$$

$$\begin{cases} f_{\tilde{\alpha}_{\pi_d}}(\pi) = \nabla_{\tilde{\alpha}_{\pi_d}} \mathcal{L}(\cancel{\pi}) \\ h_{\tilde{\alpha}_{\pi_d}}(\pi) = \nabla_{\tilde{\alpha}_{\pi_d}} \log q(\pi|\tilde{\alpha}_\pi) \end{cases}$$

$$a^*_{\tilde{\alpha}_{\pi_d}} = \frac{\mathrm{cov}(f, h)}{\mathrm{Var}(h)}$$

$$\nabla_{\tilde{\alpha}_{\pi_d}} \mathcal{L}_{new} = \sum_{s=1}^{S} \frac{1}{S} f_{\tilde{\alpha}_{\pi_d}}(\pi) - a^*_{\tilde{\alpha}_{\pi_d}} \cdot h_{\tilde{\alpha}_{\pi_d}}(\pi$$

$$\begin{cases} f_{\phi_d}(z) = \nabla_{\phi_d} \mathcal{L} \\ h_{\phi_d}(z) = \nabla_{\phi_d} q(z|\phi_d) \end{cases}$$

$$a^*_{\phi_d} = \frac{\mathrm{cov}(f_{\phi_d}, h_{\phi_d})}{\mathrm{Var}(h_{\phi_d})}$$

$$\nabla_{\phi_d} \mathcal{L}_{new} = \sum_{s=1}^{S} \frac{1}{S} f_{\phi_d}(z) - a^*_{\phi_d} \cdot h_{\phi_d}(z)$$

$$\begin{cases} j_{\tilde{\alpha}_{1d}}(l) = \nabla_{\tilde{\alpha}_{1d}} \mathcal{L} \\ h_{\tilde{\alpha}_{1d}}(l) = \nabla_{\tilde{\alpha}_{1d}} q(l|\tilde{\alpha}_{1d}, \tilde{\beta}_{1d}) \end{cases} \qquad a^*_{\tilde{\alpha}_{1d}} = \frac{cov(j_{\tilde{\alpha}_{1d}}, h_{\tilde{\alpha}_{1d}})}{Var(h_{\tilde{\alpha}_{1d}})} \qquad \nabla_{\tilde{\alpha}_{1d}} \mathcal{L}\,(new) = \sum_{s=1}^{S} \frac{1}{S} j_{\tilde{\alpha}_{1d}}(l) - a^*_{\tilde{\alpha}_{1d}} \cdot h_{\tilde{\alpha}_{1d}}(l)$$

$$\begin{cases} j_{\tilde{\beta}_{1d}}(l) = \nabla_{\tilde{\beta}_{1d}} \mathcal{L} \\ h_{\tilde{\beta}_{1d}}(l) = \nabla_{\tilde{\beta}_{1d}} q(l|\tilde{\alpha}_{1d}, \tilde{\beta}_{1d}) \end{cases} \qquad a^*_{\tilde{\beta}_{1d}} = \frac{cov(j_{\tilde{\beta}_{1d}}, h_{\tilde{\beta}_{1d}})}{Var(h_{\tilde{\beta}_{1d}})} \qquad \nabla_{\tilde{\beta}_{1d}} \mathcal{L}_{(new)} = \sum_{s=1}^{S} \frac{1}{S} j_{\tilde{\beta}_{1d}}(l) - a^*_{\tilde{\beta}_{1d}} \cdot h_{\tilde{\beta}_{1d}}(l)$$

$$\begin{cases} j_{\tilde{\alpha}_{ld}}(l) = \nabla_{\tilde{\alpha}_{ld}} \mathcal{L} \\ h_{\tilde{\alpha}_{ld}}(l) = \nabla_{\tilde{\alpha}_{ld}} q(l|\tilde{\alpha}_{ld}, \tilde{\beta}_{ld}) \end{cases} \qquad a^*_{\tilde{\alpha}_{ld}} = \frac{cov(j_{\tilde{\alpha}_{ld}}, h_{\tilde{\alpha}_{ld}})}{Var(h_{\tilde{\alpha}_{ld}})} \qquad \nabla_{\tilde{\alpha}_{ld}} \mathcal{L}_{new} = \sum_{s=1}^{S} \frac{1}{S} j_{\tilde{\alpha}_{ld}}(l) - a^*_{\tilde{\alpha}_{ld}} \cdot h_{\tilde{\alpha}_{ld}}(l)$$

$$\begin{cases} j_{\tilde{\beta}_{ld}}(l) = \nabla_{\tilde{\beta}_{ld}} \mathcal{L} \\ h_{\tilde{\beta}_{ld}}(l) = \nabla_{\tilde{\beta}_{ld}} q(l|\tilde{\alpha}_{ld}, \tilde{\beta}_{ld}) \end{cases} \qquad a^*_{\tilde{\beta}_{ld}} = \frac{cov(j_{\tilde{\beta}_{ld}}, h_{\tilde{\beta}_{ld}})}{Var(h_{\tilde{\beta}_{ld}})} \qquad \nabla_{\tilde{\beta}_{ld}} \mathcal{L}_{new} = \sum_{s=1}^{S} \frac{1}{S} j_{\tilde{\beta}_{ld}}(l) - a^*_{\tilde{\beta}_{ld}} h_{\tilde{\beta}_{ld}}(l)$$

$$\rho_t^{\tilde{\alpha}_\pi} = \frac{\gamma_{\tilde{\alpha}_\pi}}{\sqrt{\sum_{t'=1}^{t} \left(\nabla_{\tilde{\alpha}_\pi} \mathcal{L}_{new}\right)^2 + \varepsilon_{\tilde{\alpha}_\pi}}}$$

$$\rho_t^{\tilde{\phi}} = \frac{\gamma_{\tilde{\phi}}}{\sqrt{\sum_{t'=1}^{t} \left(\nabla_{\tilde{\phi}_t} \mathcal{L}_{new}\right)^2 + \varepsilon_{\tilde{\phi}}}}$$

$$\rho_t^{\tilde{\alpha}_\lambda} = \frac{\gamma_{\tilde{\alpha}_\lambda}}{\sqrt{\sum_{t'=1}^{t} \left(\nabla_{\tilde{\alpha}_\lambda} \mathcal{L}_{new}\right)^2 + \varepsilon_{\tilde{\alpha}_\lambda}}}$$

$$\rho_t^{\tilde{\beta}_\lambda} = \frac{\gamma_{\tilde{\beta}_\lambda}}{\sqrt{\sum_{t'=1}^{t} \left(\nabla_{\tilde{\beta}_\lambda} \mathcal{L}_{new}\right)^2 + \varepsilon_{\tilde{\beta}_\lambda}}}$$

$$\rho_t^{\tilde{\alpha}_\ell} = \frac{\gamma_{\tilde{\alpha}_\ell}}{\sqrt{\sum_{t'=1}^{t} \left(\nabla_{\tilde{\alpha}_\ell} \mathcal{L}_{new}\right)^2 + \varepsilon_{\tilde{\alpha}_\ell}}}$$

$$\rho_t^{\tilde{\beta}_\ell} = \frac{\gamma_{\tilde{\beta}_\ell}}{\sqrt{\sum_{t'=1}^{t} \left(\nabla_{\beta_\ell} \mathcal{L}_{new}\right)^2 + \varepsilon_{\tilde{\beta}_\ell}}}$$

$$\tilde{\alpha}_{\pi(t+1)} = \tilde{\alpha}_{\pi(t)} + \rho_t^{\tilde{\alpha}_\pi} \cdot \nabla_{\tilde{\alpha}_\pi} \mathcal{L}^{(t)}_{(New)}$$

$$\tilde{\phi}_{(t+1)} = \tilde{\phi}_{(t)} + \rho_t^{\tilde{\phi}} \cdot \nabla_{\tilde{\phi}} \mathcal{L}^{(t)}_{(New)}$$

$$\tilde{\alpha}_{\lambda(t+1)} = \tilde{\alpha}_{\lambda(t)} + \rho_t^{\tilde{\alpha}_\lambda} \cdot \nabla_{\tilde{\alpha}_\lambda} \mathcal{L}^{(t)}_{(New)}$$

$$\tilde{\beta}_{\lambda(t+1)} = \tilde{\beta}_{\lambda(t)} + \rho_t^{\tilde{\beta}_\lambda} \cdot \nabla_{\tilde{\beta}_\lambda} \mathcal{L}^{(t)}_{(New)}$$

$$\tilde{\alpha}_{\ell(t+1)} = \tilde{\alpha}_{\ell(t)} + \rho_t^{\tilde{\alpha}_\ell} \cdot \nabla_{\tilde{\alpha}_\ell} \mathcal{L}^{(t)}_{(New)}$$

$$\tilde{\beta}_{\ell(t+1)} = \tilde{\alpha}_{\ell(t)} + \rho_t^{\tilde{\beta}_\ell} \cdot \nabla_{\tilde{\beta}_\ell} \mathcal{L}^{(t)}_{(New)}$$

Algorithm

Input: mapping Ξ, $\pi \sim$ Dirichelet, $l \sim \gamma_{1,\beta}$, $\lambda \sim \gamma_{\alpha_\lambda,\alpha_r}$
number of components K_{max}

Output: Variational distribution $q(z_i, k_i)$, $q(\pi)$, $q(\lambda)q(l)$

Initialize: Variational parameter $\{\tilde{\alpha}_\pi, \tilde{\phi}, \tilde{\alpha}_\lambda, \tilde{\beta}_\lambda, \tilde{\alpha}_l, \tilde{\beta}_l\} = V$

While the ELBO has not convergence do

// Draw S samples from the variational approximation //

for s = 1 to S do

Pick $\pi_s \sim$ Dirichlet$(\tilde{\alpha}_\pi)$; $\pi_s \to \nabla_{\tilde{\alpha}_{\pi_s}} \log q(\pi_s | \tilde{\alpha}_{\pi_s})$

Pick $\lambda_s \sim$ Gamma$(\tilde{\alpha}_\lambda, \tilde{\beta}_\lambda)$; $\lambda_s \to \nabla_{\tilde{\alpha}_{\lambda s}, \tilde{\beta}_{\lambda s}} \log q(\lambda_s | \tilde{\alpha}_{\lambda s}, \tilde{\beta}_{\lambda s})$

Pick $l_s \sim$ Gamma$(\tilde{\alpha}_l, \tilde{\beta}_l)$; $l_s \to \nabla_{\tilde{\alpha}_{l_s}, \tilde{\beta}_{l_s}} \log q(l_s | \tilde{\alpha}_{l_s}, \tilde{\beta}_{l_s})$

for l = 1 to K do

Pick $z_{ls} \sim q(z_{ls} | \phi_{ls})$; $z_{ls} \to \nabla_{\phi_{ls}} \log q(z_{ls} | \phi_{ls})$

end

end

for d = 1 to D do (dth entry of gradient)

for s = 1 to S do

$f^s_{\tilde{\alpha}_{\pi d}}(\pi)$; $h^s_{\tilde{\alpha}_{\pi d}}(\pi)$

$f^s_{\phi_d}(z)$; $h^s_{\phi_d}(z)$

$f^s_{\tilde{\alpha}_{\lambda d}}(\lambda)$; $h^s_{\tilde{\alpha}_{\lambda d}}(\lambda)$

$f^s_{\tilde{\beta}_{\lambda d}}(\lambda)$; $h^s_{\tilde{\beta}_{\lambda d}}(\lambda)$

$f^s_{\tilde{\alpha}_{l d}}(l)$; $h^s_{\tilde{\alpha}_{l d}}(l)$

$f^s_{\tilde{\beta}_{l d}}(l)$; $h^s_{\tilde{\beta}_{l d}}(l)$

end

$a^*_{\tilde{\alpha}_{\pi d}}$; $a^*_{\phi d}$; $a^*_{\tilde{\alpha}_{\lambda d}}$; $a^*_{\tilde{\beta}_{\lambda d}}$; $a^*_{\tilde{\alpha}_{\ell d}}$; $a^*_{\tilde{\beta}_{\ell d}}$

end

Update variational bayesian

$$\tilde{\alpha}_{\pi(t+1)} = \tilde{\alpha}_{\pi(t)} + \rho_t^{\tilde{\alpha}_\pi} \cdot \nabla_{\tilde{\alpha}_\pi} \mathcal{L}^{(t)}_{(New)}$$

⋮

until convergened.