LLDA Variational Approximation Derivation

Model Parameters

$$
\begin{aligned}
\beta &= P(S|Z) \\
\pi &= P(L|Z) \\
\alpha &= P(\theta) \\
\theta &= P(Z|\theta)
\end{aligned} \tag{1}
$$

Variational Parameters

Where D is the experiment.

$$
\begin{aligned}
\gamma &= P(\theta|D) \\
\phi &= P(Z|S,D) \\
\psi &= P(L|S,D)
\end{aligned} \tag{2}
$$

Node Distributions

$$p(\theta;\alpha) = \frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}\prod_{i=1}^{K}\theta_i^{\alpha_i-1} \tag{3}$$

$$p(Z=z|\theta) = \theta_z \tag{4}$$

$$p(S=s|Z=z;\beta) = \beta_{sz} \tag{5}$$

$$p(L=l|Z=z;\pi) = \pi_{lz} \tag{6}$$

Likelihood

$$p(\theta,Z,S,L|\alpha,\beta,\pi) = p(\theta;\alpha)\prod_{n=1}^{N}p(Z_n|\theta)p(S_n|Z_n;\beta)p(L_n|Z_n;\pi) \tag{7}$$

$$p(S,L|\alpha,\beta,\pi) = \int_\theta p(\theta;\alpha)\left(\prod_{n=1}^{N}\sum_{Z_n=1}^{K}p(Z_n|\theta)p(S_n|Z_n;\beta)p(L_n|Z_n;\pi)\right)d\theta \tag{8}$$

We will use the notation $\mathcal{D}$ to indicate the corpus of experiments. $\mathcal{D}$ is a $V\times M$ matrix where there are V genes and M experiments in the corpus.

The log–likelihood function becomes

$$ll(\alpha,\beta,\pi) = \sum_{d=1}^{M}\log p(s_d,l_d|\alpha,\beta,\pi). \tag{9}$$

Multiplying and dividing by a variational distribution q gives

$$ll(\alpha,\beta,\pi) = \log\int_\theta\sum_{\underline{z}}\sum_{\underline{l}}p(\theta,Z,S,L|\alpha,\beta,\pi)\frac{q(\theta,Z,L)}{q(\theta,Z,L)}d\theta. \tag{10}$$

We have chosen to marginalize the label in this equation because it is only a partially observed random variable. In the cases the label is observed for a gene, the average becomes degenerate.

Applying Jensen's Inequality gives

$$ll(\alpha,\beta,\pi) \geq E_q[\log p(\theta,Z,L,S|\alpha,\beta,\pi)] - E_q[\log q(\theta,Z,L)]. \tag{11}$$

Choosing the fully factorized distribution for q as a variational distribution gives

$$q(\theta,Z,L|\gamma,\phi,\psi) = q(\theta;\gamma)q(Z;\phi)q(L;\psi). \tag{12}$$

Then the bound to the log–likelihood in 11 can be factorized to give

$$\begin{aligned}\mathcal{L}(\gamma,\phi,\psi;\alpha,\beta,\pi) = \\ &E_q[p(\theta;\alpha)] + E_q[\log p(Z|\theta)] + E_q[p(S|Z;\beta)] + E_q[p(L|Z;\pi)] \\ &-E_q[\log q(\theta;\gamma)] - E_q[\log q(Z;\phi)] - E_q[\log q(L;\psi)]\end{aligned} \tag{13}$$

Variational Bound Expectations

We now compute each of expectations in equation 13.

$$E_q[p(\theta|\alpha)] = \log\Gamma\left(\sum_i \alpha_i\right) - \sum_i \log\Gamma(\alpha_i) + \sum_{i=1}^{K}(\alpha_i - 1)\left[\Psi(\gamma_i) - \Psi(\sum_j \gamma_j)\right] \tag{14}$$

$$E_q[p(Z = i|\theta)] = \sum_{n=1}^{N}\sum_{i=1}^{K}\phi_{ni}\left[\Psi(\gamma_i) - \Psi(\sum_j \gamma_j)\right] \tag{15}$$

$$E_q[p(S = v|Z = i;\beta)] = \sum_{n=1}^{N}\sum_{k=1}^{V}\phi_{ni}s_n^k \log\beta_{ki} \tag{16}$$

$$E_q[p(L = j|Z = i;\pi)] = \sum_{n=1}^{N}\sum_{i=1}^{K}\sum_{j=1}^{L}\phi_{ni}\psi_{nj}\log\pi_{ji} \tag{17}$$

$$E_q[\log q(\theta)] = \log\Gamma\left(\sum_i \gamma_i\right) - \sum_i \log\Gamma(\gamma_i) + \sum_{i=1}^{K}(\gamma_i - 1)\left[\Psi(\gamma_i) - \Psi(\sum_j \gamma_j)\right] \tag{18}$$

$$E_q[\log q(Z;\phi)] = \sum_{n=1}^{N}\sum_{i=1}^{K}\phi_{ni}\log\phi_{ni} \tag{19}$$

$$E_q[\log q(L;\psi)] = \sum_{n=1}^{N}\sum_{j=1}^{L}\psi_{nj}\log\psi_{nj} \tag{20}$$

Variational Parameter Estimation

Estimate of ϕ Isolating the terms in equation 13 that reference ϕ and including the constraint that $\sum_i \phi_{ni} = 1, \forall n$ gives

$$\mathcal{L}_{[\phi]} = \phi_{ni}\left[\Psi(\gamma_i) - \Psi(\sum_j \gamma_j)\right] + \sum_{k=1}^{V}\phi_{ni}s_n^k\log\beta_{ki} + \sum_{j=1}^{L}\phi_{ni}\psi_{nj}\log\pi_{ji} - \phi_{ni}\log\phi_{ni} + \lambda_n(\sum_{j=1}^{K}\phi_{nj} - 1). \tag{21}$$

Differentiating with respect to the parameter and setting equal to 0 gives

$$\phi_{ni} \propto \exp\left(\Psi(\gamma_i) + \sum_{j=1}^{L}\psi_{nj}\log\pi_{ji} + \log\beta_{vi}\right) \tag{22}$$

Estimate of ψ Isolating the terms in equation 13 that reference ψ and including the constraint that $\sum_i \psi_{ni} = 1, \forall n$ gives

$$\mathcal{L}_{[\psi]} = \sum_{i=1}^{K}\phi_{ni}\psi_{nj}\log\pi_{ji} - \psi_{nj}\log\psi_{nj} + \lambda_n(\sum_{j=1}^{L}\psi_{nj} - 1). \tag{23}$$

Differentiating with respect to the parameter and setting equal to 0 gives

$$\psi_{nj} \propto \exp\left(\sum_{i=1}^{K}\phi_{ni}\log\pi_{ji}\right) \tag{24}$$

Estimate of γ Isolating the terms in equation 13 that reference γ gives

$$\mathcal{L}_{[\gamma]} = (\alpha_i - 1)\left[\Psi(\gamma_i) - \Psi(\sum_j \gamma_j)\right] + \sum_{n=1}^{N} \phi_{ni}\left[\Psi(\gamma_i) - \Psi(\sum_j \gamma_j)\right]. \tag{25}$$

Differentiating with respect to the parameter and setting equal to 0 gives

$$\gamma_i = \alpha_i + \sum_{n=1}^{N} \phi_{ni} \tag{26}$$

Model Parameter Estimation

Estimate of β

$$\mathcal{L}_{[\beta]} = \sum_{d=1}^{M}\sum_{n=1}^{N_d} \phi_{ni}^{d} s_{dn}^{v} \log \beta_{vi} + \lambda_i(\sum_{k=1}^{V} \beta_{vi} - 1) \tag{27}$$

The update equation becomes

$$\beta_{vi} \propto \sum_{d=1}^{M}\sum_{n=1}^{N_d} \phi_{ni}^{d} s_{dn}^{v} \tag{28}$$

Estimate of π

$$\mathcal{L}_{[\pi]} = \sum_{d=1}^{M}\sum_{n=1}^{N_d} \phi_{ni}\psi_{nj} \log \pi_{ji} + \lambda_i(\sum_{j=1}^{L} \pi_{ji} - 1) \tag{29}$$

The update equation becomes

$$\pi_{ji} \propto \sum_{d=1}^{M}\sum_{n=1}^{N_d} \phi_{ni}\psi_{nj} \tag{30}$$

Estimate of α Since the parameters in the α vector are coupled a gradient descent procedure must be used. See [Blei et. al. 2002] for details.