# Machine Learning in Complex Domains: Assignment 3

Ryan Cotterell, Daniel Deutsch, Dan Crankshaw

## 4 Blocked Gibbs Sampler

We have the following equation

$$p(z_{d,i} = k, x_{d,i} = c | \mathbf{z} - z_{d,i}, \mathbf{x} - x_{d,i}, \mathbf{c}, \mathbf{w}; \alpha, \beta, \lambda) = \frac{p(\mathbf{z}, \mathbf{x}, \mathbf{c}, \mathbf{w} | \alpha, \beta, \lambda)}{p(\mathbf{z} - z_{d,i}, \mathbf{x} - x_{d,i}, \mathbf{c}, \mathbf{w}; \alpha, \beta, \lambda)}$$

The full join (numerator) is given to us in the assignment. So every term that does not involve $z_{d,i}$ or $x_{d,i}$ will cancel. This is basically a union of (15) and (18). This yields the following:

$$p(z_{d,i} = k, x_{d,i} = 0 | \mathbf{z} - z_{d,i}, \mathbf{x} - x_{d,i}, \mathbf{c}, \mathbf{w}; \alpha, \beta, \lambda) = \frac{p(x = 0|\lambda) \frac{\Gamma(n^k_{w_{d,i}} + 1 + \beta)}{\Gamma(1 + \sum_w n^k_w + \beta)} \frac{\Gamma(n^d_k + 1 + \alpha)}{\Gamma(1 + \sum_{k'} n^d_{k'} + \alpha}}{\frac{\Gamma(n^k_{w_{d,i}} + \beta)}{\Gamma(\sum_w n^k_w + \beta)} \frac{\Gamma(n^d_k + \alpha)}{\Gamma(\sum_{k'} n^d_{k'} + \alpha)}}$$

$$p(z_{d,i} = k, x_{d,i} = 0 | \mathbf{z} - z_{d,i}, \mathbf{x} - x_{d,i}, \mathbf{c}, \mathbf{w}; \alpha, \beta, \lambda) \propto \frac{(1 - \lambda)(n^k_{w_{d,i}} + \beta)(n^d_k + \alpha)}{(n^k_* + K\beta)(n^d_* + V\alpha)}$$

### 4.1 Analysis Questions

## 5 Text Analysis with MCLDA

### 5.1 Empirical Questions

1.

2.

3.

4.

5.

6. (a)

   (b)

   (c)

## 6 Variational Inference

## Ryan's derivation of the update for $\delta$

$\delta$ is the variational parameter for $x$ and we need an update rule for it.

It is a generally known fact that

$$\mathcal{L}_{[\delta]} = \mathbb{E}_q\left[\log(p(\mathbf{w}|\mathbf{z}, x, c, \phi))\right] + \mathbb{E}_q\left[\log(p(x|\lambda))\right] - \mathbb{E}_q\left[\log(q(x|\lambda))\right]$$

We can expand this to (note that $T = \{\text{g,l}\}$

$$\sum_{n=1}^{N_d}\sum_{k=1}^{K}\sum_{t=1}^{T} w_n^w c_d^c\, \delta_{d,t,n}\epsilon_{d,n,k}\log(\phi_{d,n,v,k,c}) + \sum_{n=1}^{N_d}\sum_{t=1}^{T}\delta_{d,t,n}\log(\lambda_t)$$
$$- \sum_{n=1}^{N_d}\sum_{t=1}^{T}\delta_{d,t,n}\log(\delta_{d,t,n}) + \zeta_n(\sum_{t=1}^{T}\delta_{d,t,n} - 1)$$

Now we take the partial for the global and local $\delta$ and $\zeta_i$ is a Lagrance multiplier.

$$\frac{\partial}{\partial\delta_{d,g,n}}\left[\mathcal{L}_{[\delta]}\right] = \sum_{k=1}^{K}\epsilon_{d,n,k}\log(\phi_{d,v,g,k,c}) + \log(\lambda_g) - \log(\delta_{d,g,n}) - \zeta_g$$

Setting this equal to 0 and relocating the $\log(\delta)d, g, n$ to the left hand size and then exponentiating yields.

$$\delta_{d,g,n} \propto w_n^v\lambda_g \exp\left\{\sum_{k=1}^{K}\epsilon_{d,n,k}\log(\phi_{d,v,g,k,c})\right\}$$

By analogy, we achieve the update rule for each local corpus

$$\delta_{d,l,n} \propto w_n^v c_n^c\lambda_l \exp\left\{\sum_{k=1}^{K}\epsilon_{d,n,k}\log(\phi_{d,v,l,k,c})\right\}$$

## Ryan's derivation of the update for $\epsilon$

$\epsilon$ is the variational parameter for $\theta$. We get

$$\mathcal{L}_{[\epsilon]} = \mathbb{E}_q\left[\log(p(\mathbf{w}|\mathbf{z}, x, c, \phi))\right] + \mathbb{E}_q\left[\log(p(\mathbf{z}|\theta))\right] - \mathbb{E}_q\left[\log(q(\mathbf{z}|\epsilon))\right]$$

Now since we are clever, we note that the only difference between this and the $\mathcal{L}_{[\phi]}$ in Blei et al. (2003) is $\mathbb{E}_q\left[\log(p(\mathbf{w}|\mathbf{z}, x, c, \phi))\right]$. Thus we can simply write down adapation of equation (16) to include the $\delta$ term.

$$\epsilon_{d,n,k} \propto \exp\left\{\sum_{t}^{T} w_n^v c_n^c\, \delta_{d,t,n}\log(\phi_{d,v,l,k,c})\left(\Psi(\gamma_k) - \Psi(\sum_{k'=1}^{K}\gamma_{k'})\right)\right\}$$

Finally to get the updates for $\phi$ we look at A.4.1 in Blei et al. (2003). We notice that only addition will be $\delta_{d,t,n}$ which will be constant with respect to the derivative. So we can write down by inspection the global update

$$\phi_{d,v,g,k} \propto \beta + \sum_{d=1}^{D}\sum_{n=1}^{N_d} w_{d,n}^v\, \delta_{d,g,n}\epsilon_{d,n,k}$$

Now to get the local update, we have

$$\phi_{d,v,l,k,c} \propto \beta + \sum_{d=1}^{D}\sum_{n=1}^{N_d} w_{d,n}^v c_d^c\, \delta_{d,l,n}\epsilon_{d,n,k}$$

The updates for $\gamma$ are identical.

### 6.1 Analysis Questions

### 6.3 Empirical Questions

1. 

2. 

3. 

4. 