Deriving Gibbs Sampling for LDA.

In this question we derive two Gibbs sampling algorithms for latent Dirichlet allocation (LDA). Recall that LDA is a topic model that defines multiple mixtures of discrete distributions with shared components. The archetypical application is to words in documents. Suppose there are W possible words, D documents and K topics. The LDA model specifies the distribution of the ith word in the dth document, $x_{id} \in \{1 \dots W\}$, in terms of the hyperparameters α and β, by way of latent Dirichlet parameters:

$\phi_{kw} = \frac{B_{kw}}{M_k}$

$\theta_{dk} = \frac{A_{dk}}{N_d}$

topic distribution for dth document	$\boldsymbol{\theta}_d \mid \alpha \sim \text{Dirichlet}(\alpha, \dots, \alpha)$	(1)
word distribution for kth topic	$\boldsymbol{\phi}_k \mid \beta \sim \text{Dirichlet}(\beta, \dots, \beta)$	(2)
topic for ith word in dth document	$z_{id} \mid \boldsymbol{\theta}_d \sim \text{Discrete}(\boldsymbol{\theta}_d)$	(3)
identity of ith word in dth document	$x_{id} \mid z_{id}, \boldsymbol{\phi}_{z_{id}} \sim \text{Discrete}(\boldsymbol{\phi}_{z_{id}})$	(4)

Let $A_{dk} = \sum_i \delta(z_{id} = k)$ be the number of z_{id} variables taking on value k in document d, and $B_{kw} = \sum_d \sum_i \delta(x_{id} = w)\delta(z_{id} = k)$ be the number of times word w is assigned to topic k across all the documents. Let N_d be the total number of words in document d and let $M_k = \sum_w B_{kw}$ be the total number of words assigned to topic k.

Joint.

(a) ✓ Write down the joint probability over the observed data and latent variables, expressing the joint probability in terms of the counts N_d, M_k, A_{dk}, and B_{kw}. [5 points]

(b) Derive the Gibbs sampling updates for all the latent variables z_{id} and parameters $\boldsymbol{\theta}_d$ and $\boldsymbol{\phi}_k$. [10 points]

(c) ✓ Integrate out all the parameters $\boldsymbol{\theta}_d$ and $\boldsymbol{\phi}_k$ from the joint probability in (a), resulting in a joint probability over only the z_{id} topic assignment variables and x_{id} observed variables. Again this expression should relate to z_{id}'s and x_{id}'s only through the counts N_d, M_k, A_{dk}, and B_{kw}. [5 points]

(d) Derive the Gibbs sampling updates for the z_{id} with all parameters integrated out. This is called **collapsed Gibbs sampling**. You will need the the following identity of the Gamma function: $\Gamma(1 + x) = x\Gamma(x)$ for $x > 0$. [10 points]

(e) What hyperpriors would you give to α and β? How would you generate samples of α and β from the appropriate conditionals? [You should suggest an algorithm and justify its feasibility, but do not need to derive the update equations; 5 points]

(b) $P(z_i \mid \underset{\sim}{z}_{-i}, \underset{\sim}{\theta}, \underset{\sim}{\phi})$, $P(\theta_d \mid \underset{\sim}{\theta}_{-d}, \underset{\sim}{z}, \underset{\sim}{\phi})$, $P(\phi_k \mid \underset{\sim}{\phi}_{-k}, \underset{\sim}{z}, \underset{\sim}{\theta})$.

(c) $P(\underset{\sim}{w}, \underset{\sim}{z})$

(d). $P(z_i \mid \underset{\sim}{z}_{-i}, \underset{\sim}{w})$.

$$
\underset{=}{\phi} = K \left(\begin{array}{cccc} w_1 & w_2 & \cdots & w_W \\ & & & \\ & & & \\ & & & \end{array} \right) \begin{array}{l} \text{Topic 1} \\ \text{Topic 2} \\ \vdots \\ \text{Topic } K \end{array} \in \mathbb{R}^{K \times W}
$$

$$
\underset{=}{\theta} = D \left(\begin{array}{cccc} \text{Topic 1} & \text{Topic 2} & \cdots & \text{Topic } K \\ & & & \\ & & & \\ & & & \end{array} \right) \begin{array}{l} \text{Document 1} \\ \text{Document 2} \\ \vdots \\ \text{Document } D. \end{array} \in \mathbb{R}^{D \times K}
$$

1. Joint :

$$
P(\underset{\sim}{x}, \underset{\sim}{z}, \underset{=}{\theta}, \underset{=}{\phi} \mid \alpha, \beta) = \prod_{i,d} P(x_{id} \mid z_{id}, \underset{\sim}{\phi}_{z_{id}}) \cdot \prod_{i,d} P(z_{id} \mid \underset{\sim}{\theta}_d) \cdot \prod_{d} P(\underset{\sim}{\theta}_d \mid \alpha) \cdot \prod_{k} P(\underset{\sim}{\phi}_k \mid \beta)
$$

$$
= \prod_{i,d} \underset{\sim}{\phi}_{z_{id}, x_{id}} \cdot \prod_{i,d} \underset{\sim}{\theta}_{d, z_{id}} \cdot \prod_{d} \left(\frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \cdot \prod_{k} \theta_{dk}^{\alpha - 1} \right) \cdot \prod_{k} \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \cdot \prod_{w} \phi_{kw}^{\beta - 1} \right)
$$

x_{id}-th component of vector $\underset{\sim}{\phi}_{z_{id}}$

K : # of possible topics $\quad W$: # of possible words.

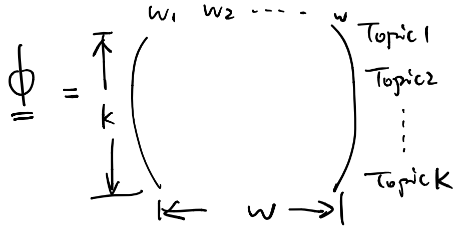
$$
= \prod_{i,d} \left(\prod_{kw} \phi_{kw}^{\sum_{id} \delta(z_{id} = k) \cdot \delta(x_{id} = w)} \right) \cdot \prod_{i,d} \left(\prod_{k} \theta_{dk}^{\delta(z_{id} = k)} \right) \cdot \left(\prod_{dk} \theta_{dk}^{\alpha - 1} \right) \cdot \left(\prod_{kw} \phi_{kw}^{\beta - 1} \right) \cdot \left(\frac{\Gamma(K\alpha)}{\Gamma(\alpha)} \right)^D \cdot \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)} \right)^K
$$

$$
= \prod_{kw} \phi_{kw}^{\sum_{id} \delta(z_{id} = k) \cdot \delta(x_{id} = w) + \beta - 1} \cdot \prod_{kd} \theta_{dk}^{\delta(z_{id} = k) + \alpha - 1} \cdot \left(\frac{\Gamma(K\alpha)}{\Gamma(\alpha)} \right)^D \cdot \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)} \right)^K
$$

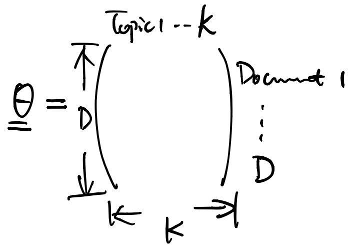
$$
= \prod_{kw} \phi_{kw}^{B_{kw} + \beta - 1} \cdot \prod_{dk} \theta_{dk}^{A_{dk} + \alpha - 1} \cdot \left(\frac{\Gamma(K\alpha)}{\Gamma(\alpha)} \right)^D \cdot \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)} \right)^K
$$

$\underline{\theta} \in \mathbb{R}^{D\times K}$

$\underline{\phi} \in \mathbb{R}^{K\times W}$

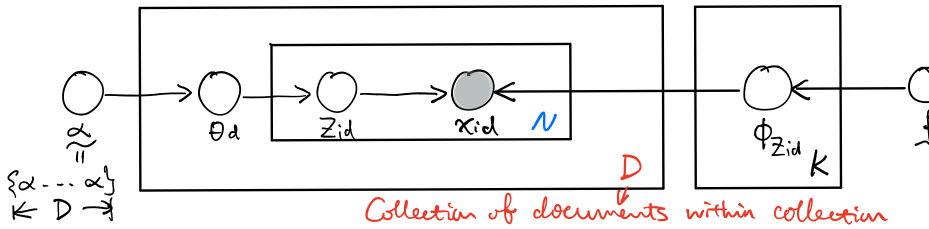


Probability of word belongs to topic



$B(\cdot)$ is multinomial Beta function.

$\Gamma(\cdot)$ is Gamma function



1(a)

$$P(\underset{\sim}{x}, \underset{\sim}{z}, \underline{\theta}, \underline{\phi} \mid \alpha, \beta) = P(\underset{\sim}{x} \mid \underset{\sim}{z}, \underline{\phi})\, P(\underset{\sim}{z} \mid \underline{\theta}) \cdot P(\underline{\theta} \mid \alpha) \cdot P(\underline{\phi} \mid \beta).$$

$$* \; P(\underset{\sim}{x} \mid \underset{\sim}{z}, \underline{\phi}) = \prod_{k=1}^{K} \prod_{w=1}^{W} \phi_{kw}^{B_{kw}}$$

$\longleftarrow$ Topics $z_1, \; z_2, \; ----, \; z_k$

Words: $w_1 \quad w_2 \cdots$

Follows multinomial distribution

$$* \; P(\underline{\phi} \mid \underline{\beta}) = \frac{1}{B(\underline{\beta})} \prod_{w=1}^{W} \phi_{kw}^{\beta - 1},$$

$$\Rightarrow P(\underset{\sim}{x} \mid \underset{\sim}{z}, \underline{\phi})\, P(\underline{\phi} \mid \underline{\beta})$$

$$= \prod_{k=1}^{K} \frac{1}{B(\underline{\beta})} \prod_{w=1}^{W} \phi_{kw}^{B_{kw} + \beta - 1} = \left(\frac{1}{B(\underline{\beta})}\right)^{K} \prod_{k=1}^{K} \prod_{w=1}^{W} \phi_{kw}^{B_{kw} + \beta - 1}$$

Dirichlet distribution, where $\Delta(\underline{\beta})$ is the normaliser.

$$* \; P(\underset{\sim}{z} \mid \underline{\theta}) = \prod_{d=1}^{D} \prod_{k=1}^{K} \theta_{dk}^{A_{dk}}$$

Follows multinomial

$$\Rightarrow P(\underset{\sim}{z} \mid \underline{\theta}) \cdot P(\underline{\theta} \mid \underset{\sim}{\alpha})$$

$$= \left(\frac{1}{B(\alpha)}\right)^{D} \prod_{d=1}^{D} \prod_{k=1}^{K} \theta_{dk}^{A_{dk} + \alpha - 1}$$

$$* \; P(\underline{\theta} \mid \underset{\sim}{\alpha}) = \frac{1}{B(\alpha)} \prod_{k=1}^{K} \theta_{dk}^{\alpha - 1}$$

Dirichlet.

$$P(\underset{\sim}{x}, \underset{\sim}{z}, \underline{\theta}, \underline{\phi} \mid \alpha, \beta) = \Big(P(\underset{\sim}{x} \mid \underset{\sim}{z}, \underline{\phi}) \cdot P(\underline{\phi} \mid \beta)\Big) \cdot \Big(P(\underset{\sim}{z} \mid \underline{\theta}) \cdot P(\underline{\theta} \mid \underset{\sim}{\alpha})\Big)$$

$$= \left(\frac{1}{B(\underline{\beta})}\right)^{K} \left(\prod_{k=1}^{K} \prod_{w=1}^{W} \phi_{kw}^{B_{kw} + \beta - 1}\right) \cdot \left(\frac{1}{B(\alpha)}\right)^{D} \left(\prod_{d=1}^{D} \prod_{k=1}^{K} \theta_{dk}^{A_{dk} + \alpha - 1}\right)$$

$$= \left(\frac{\Gamma(K\beta)}{\Gamma(\beta)^{K}}\right)^{K} \left(\prod_{k=1}^{K} \prod_{w=1}^{W} \phi_{kw}^{B_{kw} + \beta - 1}\right) \cdot \left(\frac{\Gamma(W\alpha)}{\Gamma(\alpha)^{W}}\right)^{D} \left(\prod_{d=1}^{D} \prod_{k=1}^{K} \theta_{dk}^{A_{dk} + \alpha - 1}\right)$$

where $\phi_{kw} = \dfrac{B_{kw}}{M_k}$, $\quad \theta_{dk} = \dfrac{A_{dk}}{N_d}$.

(b) We want to find the following for Gibbs Sampling

$P(z_i \mid \underline{z}_{-i}, \underline{\theta}, \underline{\phi}, \alpha, \beta)$, $P(\underset{\sim}{\theta_d} \mid \underline{\theta}_{-d}, \underset{\sim}{w}, \underline{z}, \underline{\phi}, \alpha, \beta)$, $P(\underset{\sim}{\phi_k} \mid \underline{\phi}_{-k}, \underset{\sim}{w}, \underline{z}, \underline{\theta}, \alpha, \beta)$

We leave α, β for simplicity

1. $$P(z_{id}=k \mid \underset{\sim}{z}_{-id}, \underset{\sim}{x}, \underline{\theta}, \underline{\phi}) = \frac{P(z_{id}=k, \underset{\sim}{z}_{-id}, \underset{\sim}{x}, \underline{\theta}, \underline{\phi})}{P(\underset{\sim}{z}_{-id}, \underset{\sim}{x}, \underline{\theta}, \underline{\phi})} = \frac{\phi_{k,x_{id}}\theta_{dk}}{\sum_{k'} \phi_{k',x_{id}}\theta_{dk'}}$$

$$\left(\propto \frac{\phi_{kw}^{\beta+B_{kw}-1} \cdot \theta_{dk}^{\alpha+A_{dk}-1}}{\phi_{kw}^{\beta-1+B_{kw}-1} \cdot \theta_{dk}^{\alpha-1+A_{dk}-1}} = \theta_{dk}\phi_{kw} = \theta_{dk}\phi_{kx_{id}} \right)$$

2. $$P(\underset{\sim}{\theta_d} = \theta_* \mid \underline{\theta}_{-d}, \underset{\sim}{x}, \underset{\sim}{z}, \underline{\phi}) = \frac{P(\underset{\sim}{\theta_d} = \underset{\sim}{\theta_*}, \underline{\theta}_{-d}, \underset{\sim}{x}, \underset{\sim}{z}, \underline{\phi})}{P(\underline{\theta}_{-d}, \underset{\sim}{x}, \underset{\sim}{z}, \underline{\phi})} = \frac{P(\underset{\sim}{\theta_d} = \underset{\sim}{\theta_*}, \underline{\theta}_{-d}, \underset{\sim}{x}, \underset{\sim}{z}, \underline{\phi})}{\int P(\underline{\theta}, \underline{\phi}, \underset{\sim}{x}, \underset{\sim}{z})\, d\underset{\sim}{\theta_d}}$$

$$= \frac{\frac{1}{\cancel{B(\alpha)}} \cdot \prod_{k=1}^{K} \underset{\sim}{\theta_{*k}}^{A_{dk}+\alpha-1}}{\int \frac{1}{\cancel{B(\alpha)}} \prod_{k=1}^{K} \theta_{dk}^{A_{dk}+\alpha-1} d\underset{\sim}{\theta_d}} = \frac{\prod_{k=1}^{K} \underset{\sim}{\theta_{*k}}^{A_{dk}+\alpha-1}}{B(A_{dk}+\alpha)}$$

$$= \frac{\left(\prod_{k=1}^{K} \underset{\sim}{\theta_{*k}}^{A_{dk}+\alpha-1}\right) \Gamma\left(\sum_{k=1}^{K} A_{dk}+\alpha_k\right)}{\prod_{k=1}^{K} \Gamma(A_{dk}+\alpha)}$$

$$= \Gamma(N_d + K\alpha) \prod_{k=1}^{K} \frac{\underset{\sim}{\theta_{*k}}^{A_{dk}+\alpha-1}}{\Gamma(A_{dk}+\alpha)} = \mathrm{Dir}(A_{dk}+\alpha)$$

3. $$P(\underset{\sim}{\phi_k} = \underset{\sim}{\phi_*} \mid \underline{\phi}_{-k}, \underset{\sim}{z}, \underline{\theta}) = \frac{P(\underset{\sim}{\phi_k} = \underset{\sim}{\phi_*}, \underline{\phi}_{-k}, \underset{\sim}{x}, \underset{\sim}{z}, \underline{\theta})}{P(\underline{\phi}_{-k}, \underset{\sim}{x}, \underset{\sim}{z}, \underline{\theta})} = \frac{P(\underset{\sim}{\phi_k} = \underset{\sim}{\phi_*}, \underline{\phi}_{-k}, \underset{\sim}{x}, \underset{\sim}{z}, \underline{\theta})}{\int P(\underline{\theta}, \underline{\phi}, \underset{\sim}{x}, \underset{\sim}{z})\, d\underset{\sim}{\phi_k}}$$

$$= \frac{\frac{1}{\cancel{B(\beta)}} \prod_{w=1}^{W} \phi_{kw}^{B_{kw}+\beta-1}}{\int \frac{1}{\cancel{B(\beta)}} \prod_{w=1}^{W} \phi_{kw}^{B_{kw}+\beta-1} d\underset{\sim}{\phi_k}} = \frac{\prod_{w=1}^{W} \phi_{kw}^{B_{kw}+\beta-1}}{B(B_{kw}+\beta)}$$

$$= \Gamma\left(\sum_{w} B_{kw}+\beta_w\right) \prod_{w=1}^{W} \frac{\phi_{kw}^{B_{kw}+\beta-1}}{\Gamma(B_{kw}+\beta)}.$$

$$= \Gamma(M_k + W\beta) \prod_{w=1}^{W} \frac{\phi_{kw}^{B_{kw}+\beta-1}}{\Gamma(B_{kw}+\beta)} = \mathrm{Dir}(B_{kw}+\beta_k).$$

(c)

$$p(\underset{\sim}{z},\underset{\sim}{x}\mid\alpha,\beta)=\iint p(\underset{\sim}{z},\underset{\sim}{x},\underset{\sim}{\phi},\underset{\sim}{\theta}\mid\alpha,\beta)\,d\underset{\sim}{\phi}\,d\underset{\sim}{\theta} \quad \leftarrow \text{Integrate out } \underset{\sim}{\theta}, \underset{\sim}{\phi}.$$

$$=\iint p(\underset{\sim}{x}\mid\underset{\sim}{z},\underset{\sim}{\phi})\,p(\underset{\sim}{z}\mid\underset{\sim}{\theta})\cdot p(\underset{\sim}{\theta}\mid\alpha)\cdot p(\underset{\sim}{\phi}\mid\beta)\,d\underset{\sim}{\phi}\,d\underset{\sim}{\theta}$$

$$=\int p(\underset{\sim}{x}\mid\underset{\sim}{z},\underset{\sim}{\phi})\cdot p(\underset{\sim}{\phi}\mid\beta)\,d\underset{\sim}{\phi}\int p(\underset{\sim}{z}\mid\underset{\sim}{\theta})\cdot p(\underset{\sim}{\theta}\mid\alpha)\,d\underset{\sim}{\theta}$$

$$* \int p(\underset{\sim}{x}\mid\underset{\sim}{z},\underset{\sim}{\phi})\,p(\underset{\sim}{\phi}\mid\beta)\,d\underset{\sim}{\phi}=\int\prod_{k=1}^{K}\frac{1}{B(\beta)}\prod_{w=1}^{W}\phi_{kw}^{B_{kw}+\beta-1}\,d\underset{\sim}{\phi}$$

$$=\prod_{k=1}^{K}\frac{1}{B(\beta)}\int\prod_{w=1}^{W}\phi_{kw}^{B_{kw}+\beta-1}\,d\underset{\sim}{\phi}$$

$$=\prod_{k=1}^{K}\frac{1}{B(\beta)}\cdot B(B_{kw}+\beta)$$

$$* \int p(\underset{\sim}{z}\mid\underset{\sim}{\theta})\cdot p(\underset{\sim}{\theta}\mid\alpha)\,d\underset{\sim}{\theta}=\int\prod_{d=1}^{D}\frac{1}{B(\alpha)}\cdot\prod_{k=1}^{K}\theta_{dk}^{A_{dk}+\alpha-1}\,d\underset{\sim}{\theta}$$

$$=\prod_{d=1}^{D}\frac{1}{B(\alpha)}\cdot\int\prod_{k=1}^{K}\theta_{dk}^{A_{dk}+\alpha-1}\,d\underset{\sim}{\theta}$$

$$=\prod_{d=1}^{D}\frac{1}{B(\alpha)}\cdot B(A_{dk}+\alpha).$$

$$B(\alpha)=\frac{\Gamma(\sum_k\alpha_k)}{\prod_k\Gamma(\alpha_k)}=\frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K}$$

$$B(A_{dk}+\alpha)=\frac{\prod_k\Gamma(A_{kd}+\alpha_k)}{\Gamma(\sum_k A_{dk}+\alpha_k)}$$

$$\Rightarrow p(\underset{\sim}{z},\underset{\sim}{x}\mid\alpha,\beta)=\left(\prod_{d=1}^{D}\frac{B(A_{dk}+\alpha)}{B(\alpha)}\right)\cdot\left(\prod_{k=1}^{K}\frac{B(B_{kw}+\beta)}{B(\beta)}\right)$$

$$=\left(\prod_{d=1}^{D}\frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K}\cdot\frac{\prod_{k=1}^{K}\Gamma(A_{dk}+\alpha)}{\Gamma(\sum_{k=1}^{K}A_{dk}+\alpha_k)}\right)\cdot\left(\prod_{k=1}^{K}\frac{\Gamma(W\beta)}{\Gamma(\beta)^W}\cdot\frac{\prod_{w=1}^{W}\Gamma(B_{kw}+\beta)}{\Gamma(\sum_{w=1}^{W}B_{kw}+\beta)}\right)$$

$$=\left(\prod_{d=1}^{D}\frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K}\cdot\frac{\prod_{k=1}^{K}\Gamma(A_{dk}+\alpha)}{\Gamma(N_d+K\alpha)}\right)\cdot\left(\prod_{k=1}^{K}\frac{\Gamma(W\beta)}{\Gamma(\beta)^W}\cdot\frac{\prod_{w=1}^{W}\Gamma(B_{kw}+\beta)}{\Gamma(M_k+W\beta)}\right)$$

$$=\left(\frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K}\right)^D\cdot\left(\prod_{d=1}^{D}\frac{1}{\Gamma(N_d+K\alpha)}\prod_{k=1}^{K}\Gamma(A_{dk}+\alpha)\right)$$

$$\times\left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W}\right)^K\cdot\left(\prod_{k=1}^{K}\frac{1}{\Gamma(M_k+W\beta)}\cdot\prod_{w=1}^{W}\Gamma(B_{kw}+\beta)\right).$$

$$B_{k\cdot}=\sum_w B_{kw}=M_k$$

$$A_{d\cdot}=\sum_k A_{dk}=N_d$$

(d). We want: $P(z_i \mid z_{-i}, x, \alpha, \beta)$ leave α, β for simplicity.

Define $A_{dk}^{(-id)} = \sum_{j \neq i} \delta(z_{jd} = k) \Rightarrow A_{dk} = A_{dk}^{(-id)} + \delta(z_{id} = k)$

$$B_{kw}^{(-id)} = \sum_{(p,q) \neq (i,d)} \delta(x_{pq} = w) \cdot \delta(z_{pq} = k) \Rightarrow B_{kw} = B_{kw}^{(-id)} + \delta(x_{id} = w) \cdot \delta(z_{id} = k).$$

$$B_{k,x_{id}} = B_{k,x_{id}}^{(-id)} + \delta(z_{id} = k).$$

$$M_k^{(-id)} = \sum_{(p,q) \neq (i,d)} \delta(z_{pq} = k) \Rightarrow M_k = M_k^{(-id)} + \delta(z_{id} = k).$$

We can simplify further: $\Gamma(B_{k,x_{id}} + \beta) = \left(B_{k,x_{id}}^{(-id)} + \beta\right)^{\delta(z_{id} = k)} \cdot \Gamma(B_{k,x_{id}}^{(-id)} + \beta)$.

Similar for other terms.

The joint probability is:

$$P(z \mid x, \alpha, \beta) \propto \prod_k \left(\frac{\left(B_{k,x_{id}}^{(-id)} + \beta\right) \cdot \left(A_{dk}^{(-id)} + \alpha\right)}{M_k^{(-id)} + W\beta} \right)^{\delta(z_{id} = k)}.$$

So the collapsed Gibbs Sampling updates for z_{id} still follows a categorical distribution:

$$P(z_{id} = k) = \frac{r_k}{\sum_{k'} r_{k'}}, \quad \text{where } r_k = \frac{\left(B_{k,x_{id}}^{(-id)} + \beta\right) \cdot \left(A_{dk}^{(-id)} + \alpha\right)}{M_k^{(-id)} + W\beta}.$$