LDA using Gibbs Sampling

Hrishikesh Pable

January 2023

Contents

Chapter 1

Notations

- M : Total number of Documents
- N_m : Total number of Words in the mth Document
- K : Total Number of Topics
- V : Total Number of words in the Vocabulary
- Θ : $\{\theta_1, \theta_2, \ldots, \theta_M\}$
- Φ : $\{\phi_1, \phi_2, \ldots, \phi_K\}$
- θ_m : $\{\theta_{m,1}, \theta_{m,2}, \ldots, \theta_{m,K}\}$ The probability that kth topic is assigned to mth document for $k \in \{1, \ldots, K\}$
- ϕ_k : $\{\phi_{k,1}, \phi_{k,2}, \ldots, \phi_{k,V}\}$ The probability that vth word is assigned to kth topic for $v \in \{1, \ldots, V\}$.
- α : Hyperparameter of the Dirichlet Distribution over Θ
- β : Hyperparameter of the Dirichlet Distribution over Φ
- W : Matrix consisting of words in each document.
- $w_{m,n}$: The index of the word present in nth location of the mth document. $w_{m,n} \in \{1, \ldots, V\}$
- Z : Matrix consisting of topic index of words in each document.
- $z_{m,n}$: The index of the topic assigned to the word present at the nth location in the mth document. $z_{m,n} \in \{1, \ldots, K\}$
- $z_{m,n,k}$: Indicator variable $\mathrm{I}(z_{m,n} = k)$
- $c_{m,k,v}$: Indicator variable which is 1, if the mth document contains the vth word assigned the kth topic.

- $c_{m,k,\cdot}$: The number of times the words in mth document are assigned the kth topic. $c_{m,k,\cdot} = \sum_{v=1}^{V} c_{m,k,v}$
- $c_{\cdot,k,v}$: The number of times the vth word is assigned the kth topic across all documents. $c_{\cdot,k,v} = \sum_{m=1}^{M} c_{m,k,v}$
- $c_{m,:,\cdot}$: $\{c_{m,1,\cdot}, c_{m,2,\cdot}, \ldots, c_{m,K,\cdot}\}$
- $c_{\cdot,k,:}$: $\{c_{\cdot,k,1}, c_{\cdot,k,2}, \ldots, c_{\cdot,k,V}\}$

Chapter 2

Graphical Model of LDA

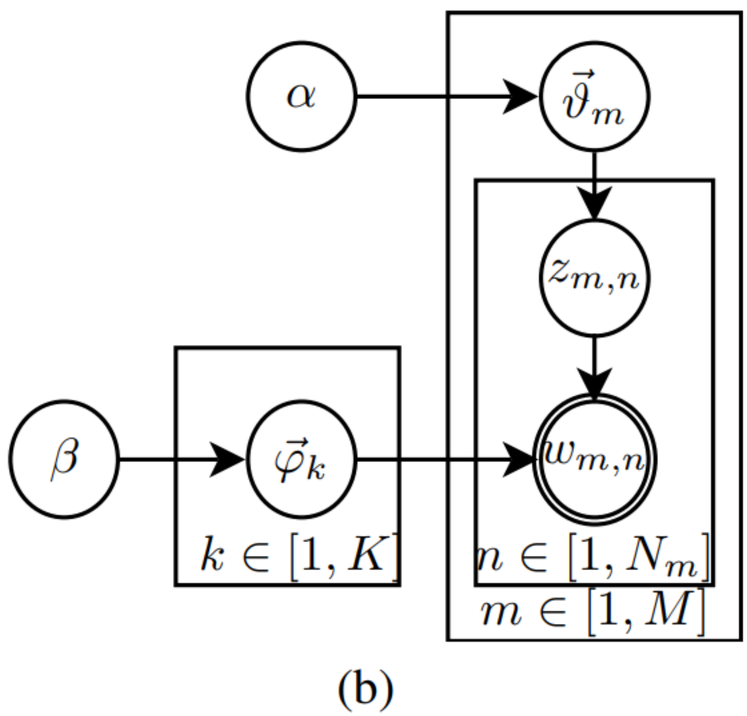

Figure 2.1: LDA Graphical Model

Chapter 3

Gibbs Sampling Derivation

We need to find the following 3 conditional distributions:

1. $P(Z|\Theta, \Phi, W)$
2. $P(\Theta|Z, \Phi, W)$
3. $P(\Phi|Z, \Theta, W)$

First, we will find $P(Z|\Theta, \Phi, W)$.

$$P(Z|\Theta, \Phi, W) = \frac{P(Z, \Theta, \Phi, W)}{P(\Theta, \Phi, W)}$$

where $P(\Theta, \Phi, W) = \sum_{Z} P(Z, \Theta, \Phi, W)$

Now, we can factorize the joint distribution by using the above bayesian network as follows:

$$\begin{aligned}
P(Z, \Theta, \Phi, W) &= p(\Theta|\alpha)\ p(Z|\Theta)\ p(\Phi|\beta)\ p(W|Z, \Phi) \\
&= \prod_{m=1}^{M} p(\theta_m|\alpha_m)\ \prod_{m=1}^{M}\prod_{n=1}^{N_m} p(z_{m,n}|\theta_m)\ \prod_{k=1}^{K} p(\phi_k|\beta_k)\ \prod_{m=1}^{M}\prod_{n=1}^{N} p(w_{m,n}|z_{m,n}, \phi_k) \\
&= \prod_{m=1}^{M} Dir(\theta_m; \alpha_m)\ \prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K} \theta_{m,k}^{z_{m,n,k}}\ \prod_{k=1}^{K} Dir(\phi_k; \beta_k)\ \prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K} \phi_{k,w_{m,n}}^{z_{m,n,k}}
\end{aligned}$$

$P(Z|\Theta, \Phi, W)$ can be found by dividing this joint distribution by the joint distribution marginalised over z. This causes the $Dir(\theta_m; \alpha_m)$ and $Dir(\phi_k; \beta_k)$ terms to be cancelled from the numerator and the denominator. So, we get,

$$
\begin{aligned}
P(Z|\Theta,\Phi,W) &= \frac{\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K}(\theta_{m,k}\phi_{k,w_{m,n}})^{z_{m,n,k}}}{\sum_{Z}\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K}(\theta_{m,k}\phi_{k,w_{m,n}})^{z_{m,n,k}}} \\
&= \frac{\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K}(\theta_{m,k}\phi_{k,w_{m,n}})^{z_{m,n,k}}}{\prod_{m=1}^{M}\prod_{n=1}^{N_m}\sum_{z_{m,n}\in\{1,\ldots,K\}}\prod_{k=1}^{K}(\theta_{m,k}\phi_{k,w_{m,n}})^{z_{m,n,k}}} \quad \text{(Since each } z_{m,n} \text{ is independent)} \\
&= \frac{\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K}(\theta_{m,k}\phi_{k,w_{m,n}})^{z_{m,n,k}}}{\prod_{m=1}^{M}\prod_{n=1}^{N_m}\sum_{k'=1}^{K}(\theta_{m,k'}\phi_{k',w_{m,n}})} \quad \left(\forall k', \prod_{k=1}^{K}(\theta_{m,k}\phi_{k,w_{m,n}})^{z_{m,n,k}} = \theta_{m,k'}\phi_{k',w_{m,n}}\right) \\
&= \prod_{m=1}^{M}\prod_{n=1}^{N_m}\left(\frac{\prod_{k=1}^{K}(\theta_{m,k}\phi_{k,w_{m,n}})^{z_{m,n,k}}}{\sum_{k'=1}^{K}(\theta_{m,k'}\phi_{k',w_{m,n}})}\right)
\end{aligned}
$$

Hence, we have the conditional distribution for Z given Θ, Φ, W.
Now we will find $P(\Theta|Z,\Phi,W)$ and $P(\Phi|Z,\Theta,W)$

$$
\begin{aligned}
P(Z,\Theta,\Phi,W) &= p(\Theta|\alpha)\ p(Z|\Theta)\ p(\Phi|\beta)\ p(W|Z,\Phi) \\
&= \prod_{m=1}^{M} p(\theta_m|\alpha_m)\ \prod_{m=1}^{M}\prod_{n=1}^{N_m} p(z_{m,n}|\theta_m)\ \prod_{k=1}^{K} p(\phi_k|\beta_k)\ \prod_{m=1}^{M}\prod_{n=1}^{N} p(w_{m,n}|z_{m,n},\phi_k) \\
&= \prod_{m=1}^{M} Dir(\theta_m;\alpha_m)\ \prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K}\theta_{m,k}^{z_{m,n,k}}\ \prod_{k=1}^{K} Dir(\phi_k;\beta_k)\ \prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K}\phi_{k,w_{m,n}}^{z_{m,n,k}}
\end{aligned}
$$

$$
\begin{aligned}
\text{let, } A &= \left(\prod_{m=1}^{M} Dir(\theta_m;\alpha_m)\right)\ \left(\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K}\theta_{m,k}^{z_{m,n,k}}\right) \\
B &= \left(\prod_{k=1}^{K} Dir(\phi_k;\beta_k)\right)\left(\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K}\phi_{k,w_{m,n}}^{z_{m,n,k}}\right) \\
\frac{1}{B(\alpha)} &= \frac{\prod_{k=1}^{K}\Gamma(\alpha_{m,k})}{\Gamma\left(\sum_{k=1}^{K}\alpha_{m,k}\right)} \text{ and } \frac{1}{B(\beta)} = \frac{\prod_{k=1}^{K}\Gamma(\beta_{m,k})}{\Gamma\left(\sum_{k=1}^{K}\beta_{m,k}\right)}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
A &= \left(\prod_{m=1}^{M} Dir(\theta_m;\alpha_m)\right) \left(\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K} \theta_{m,k}^{z_{m,n,k}}\right) \\
&= \left(\prod_{m=1}^{M} \left(\frac{1}{B(\alpha)}\right) \prod_{k=1}^{K} \theta_{m,k}^{\alpha-1}\right) \left(\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K} \theta_{m,k}^{z_{m,n,k}}\right) \\
&= \left(\prod_{m=1}^{M} \left(\frac{1}{B(\alpha)}\right) \prod_{k=1}^{K} \theta_{m,k}^{\alpha-1}\right) \left(\prod_{m=1}^{M}\prod_{k=1}^{K} \theta_{m,k}^{\sum_{n=1}^{N_m} z_{m,n,k}}\right)
\end{aligned}
\tag{3.1}
$$

Note that, for a given document m and for a given topic k,

$$
\sum_{n=1}^{N_m} z_{m,n,k} = \sum_{v=1}^{V} c_{m,k,v} = c_{m,k,\cdot}
$$

Therefore,

$$
\begin{aligned}
A &= \left(\prod_{m=1}^{M} \left(\frac{1}{B(\alpha)}\right) \prod_{k=1}^{K} \theta_{m,k}^{\alpha-1}\right) \left(\prod_{m=1}^{M}\prod_{k=1}^{K} \theta_{m,k}^{c_{m,k,\cdot}}\right) \\
&= \left(\prod_{m=1}^{M} \frac{1}{B(\alpha)} \prod_{k=1}^{K} \theta_{m,k}^{\alpha+c_{m,k,\cdot}-1}\right)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
B &= \left(\prod_{k=1}^{K} Dir(\phi_k;\beta_k)\right) \left(\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K} \phi_{k,w_{m,n}}^{z_{m,n,k}}\right) \\
&= \left(\prod_{k=1}^{K} \frac{1}{B(\beta)} \prod_{v=1}^{V} \phi_{k,v}^{\beta_{k,v}-1}\right) \left(\prod_{m=1}^{M}\prod_{n=1}^{N_m}\prod_{k=1}^{K} \phi_{k,w_{m,n}}^{z_{m,n,k}}\right)
\end{aligned}
\tag{3.2}
$$

Consider the second term in equation 3.2,

$$\begin{aligned}
\prod_{k=1}^{K}\prod_{m=1}^{M}\prod_{n=1}^{N_m}\phi_{k,w_{m,n}}^{z_{m,n,k}} &= \prod_{k=1}^{K}\prod_{m=1}^{M}\prod_{v\in\{w_{m,n}|1\le n\le N_m\}}\phi_{k,v}^{c_{m,k,v}} \\
&= \prod_{k=1}^{K}\prod_{m=1}^{M}\prod_{v=1}^{V}\phi_{k,v}^{c_{m,k,v}} && (\text{Since } \forall v \notin \{w_{m,n}|1\le n\le N_m\}, c_{m,k,v}=0) \\
&= \prod_{k=1}^{K}\prod_{v=1}^{V}\phi_{k,v}^{\sum_{m=1}^{M} c_{m,k,v}} \\
&= \prod_{k=1}^{K}\prod_{v=1}^{V}\phi_{k,v}^{c_{\cdot,k,v}}
\end{aligned}$$

Hence, B can be written as:

$$\begin{aligned}
B &= \left(\prod_{k=1}^{K}\frac{1}{B(\beta)}\prod_{v=1}^{V}\phi_{k,v}^{\beta_{k,v}-1}\right)\left(\prod_{k=1}^{K}\prod_{v=1}^{V}\phi_{k,v}^{c_{\cdot,k,v}}\right) \\
&= \left(\prod_{k=1}^{K}\frac{1}{B(\beta)}\prod_{v-1}^{V}\phi_{k,v}^{\beta+c_{\cdot,k,v}-1}\right)
\end{aligned}$$

The joint distribution can now be written as,

$$P(Z,\Theta,\Phi,W) = \left(\prod_{m=1}^{M}\frac{1}{B(\alpha)}\prod_{k=1}^{K}\theta_{m,k}^{\alpha+c_{m,k,\cdot}-1}\right)\left(\prod_{k=1}^{K}\frac{1}{B(\beta)}\prod_{v=1}^{V}\phi_{k,v}^{\beta+c_{\cdot,k,v}-1}\right)$$

So, the joint conditional of Θ and Φ is given by,

$$\begin{aligned}
P(\Theta,\Phi|Z,W) &= \frac{P(\Theta,\Phi,Z,W)}{P(Z,W)} \\
&= \frac{\left(\prod_{m=1}^{M}\frac{1}{B(\alpha)}\prod_{k=1}^{K}\theta_{m,k}^{\alpha+c_{m,k,\cdot}-1}\right)\left(\prod_{k=1}^{K}\frac{1}{B(\beta)}\prod_{v=1}^{V}\phi_{k,v}^{\beta+c_{\cdot,k,v}-1}\right)}{\left(\prod_{m=1}^{M}\frac{B(\alpha+c_{m,k,\cdot})}{B(\alpha)}\right)\left(\prod_{k=1}^{K}\frac{B(\beta+c_{\cdot,k,v}}{B(\beta)}\right)} \\
&= \left(\prod_{m=1}^{M}\frac{1}{B(\alpha+c_{m,k,\cdot})}\prod_{k=1}^{K}\theta_{m,k}^{\alpha+c_{m,k,\cdot}-1}\right)\left(\prod_{k=1}^{K}\frac{1}{B(\beta+c_{\cdot,k,v})}\prod_{v=1}^{V}\phi_{k,v}^{\beta+c_{\cdot,k,v}-1}\right) \\
&= \left(\prod_{m=1}^{M}\mathrm{Dir}(\theta_m;\alpha+c_{m,k,\cdot})\right)\left(\prod_{k=1}^{K}\mathrm{Dir}(\phi_k;\beta+c_{\cdot,k,v})\right)
\end{aligned}$$

Hence we get the individual conditional distributions as:

$$
\begin{aligned}
P(\Theta|\Phi, Z, W) &= \prod_{m=1}^{M} \mathrm{Dir}(\theta_m; \alpha + c_{m,k,\cdot}) \\
P(\Phi|\Theta, Z, W) &= \prod_{k=1}^{K} \mathrm{Dir}(\phi_k; \beta + c_{\cdot,k,v})
\end{aligned}
$$

Given these conditional distributions, we can now use Gibbs Sampling to get a sample from the joint distribution of Θ, Φ, Z

Chapter 4

Collapsed Gibbs Sampling Derivation

Here, we directly want to find the probablity of Z given the documents W. Since Θ and Φ are intermediate statistics, and they can be obtained from Z.

So, we want,

$$p(Z|W) = \frac{p(Z,W)}{p(W)} = \frac{\prod_{m=1}^{M} p(z_m, w_m)}{\prod_{m=1}^{M} \sum_{k=1}^{K} p(z_m = k, w_m)}$$

The numerator can be factored as:

$$p(W, Z|\alpha, \beta) = p(W|Z, \beta)p(Z, \alpha) \tag{4.1}$$

For finding the first term in equation 4.1, we first find,

$$\begin{aligned} p(W|Z, \Phi, \beta) &= \prod_{k=1}^{K} \prod_{\{i:z_i=k\}} p(w_i = v|z_i = k, \phi_k) \\ &= \prod_{k=1}^{K} \prod_{v=1}^{V} \phi_{k,v}^{n_{\cdot,k,v}} \end{aligned}$$

The target $p(W|Z, \beta)$ is obtained by integrating over Φ (component-wise within the product over Z)

$$
\begin{aligned}
p(W|Z,\beta) &= \int p(W|Z,\Phi)p(\Phi|\beta)\dot{d}\Phi \\
&= \int \prod_{k=1}^{K} \left(\frac{1}{B(\beta_k)} \right) \prod_{v=1}^{V} \phi_{k,v}^{(c_{\cdot,k,v}+\beta_{k,v}-1)} \, d\phi_k \\
&= \prod_{k=1}^{K} \frac{1}{B(\beta_k)} \prod_{v=1}^{V} \phi_{k,v}^{(c_{\cdot,k,v}+\beta_{k,v}-1)} \, d\phi_k \\
&= \prod_{k=1}^{K} \frac{B(c_{\cdot,k,:}+\beta_k)}{B(\beta_k)}
\end{aligned}
$$

Analogous to $p(W|Z,\beta)$ we can find $(Z|\alpha)$. We start with $p(Z|\theta,\alpha)$ and then integrate out Θ.

$$
p(Z|\Theta,\alpha) = \prod_{m=1}^{M} p(z_m|\theta_m)
$$

$$
p(z_m|\theta_m,\alpha_m) = \prod_{k=1}^{K} p(z_{m,n}=k|\theta_{m,k})
$$

$$
\begin{aligned}
p(Z|\Theta,\alpha) &= \prod_{m=1}^{M} \prod_{k=1}^{K} p(z_{m,n}=k|\theta_{m,k}) \\
&= \prod_{m=1}^{M} \prod_{k=1}^{K} \theta_{m,k}^{c_{m,k,\cdot}}
\end{aligned}
$$

$$
\begin{aligned}
p(Z|\alpha) &= \int p(Z|\Theta)p(\Theta|\alpha)\dot{d}\Theta \\
&= \int \prod_{m=1}^{M} \frac{1}{B(\alpha_m)} \prod_{k=1}^{K} \theta_{m,k}^{(c_{m,k,\cdot}+\alpha_{m,k}-1)} \, d\theta_m \\
&= \prod_{m=1}^{M} \frac{B(c_{m,:,.}+\alpha_m)}{B(\alpha_m)}
\end{aligned}
$$

$$
p(Z,W|\alpha,\beta) = \prod_{k=1}^{K} \frac{B(c_{\cdot,k,:}+\beta_k)}{B(\beta_k)} \prod_{m=1}^{M} \frac{B(c_{m,:,.}+\alpha_m)}{B(\alpha_m)}
$$

Now, we will derive the conditional distribution for each z_i ($i = (m, n)$) given the other variables, from this joint distribution of Z and W.

$$
\begin{aligned}
p(z_i = k|Z_{\neg i}, W) &= \frac{p(W, Z)}{p(W, Z_{\neg i})} \\
&= \frac{p(W|Z)}{p(W_{\neg i}|Z_{\neg i})p(w_i)} \frac{p(Z)}{p(Z_{\neg i})} \\
&\propto \frac{p(W|Z)}{p(W_{\neg i}|Z_{\neg i})} \frac{p(Z)}{p(Z_{\neg i})} \\
&\propto \frac{\left[\prod_{m=1}^{M} B(\alpha_m + c_{m,:,\cdot}) \prod_{k=1}^{K} B(\beta_k + c_{\cdot,k,:})\right]}{\left[\prod_{m=1}^{M} B(\alpha_m + c_{m,:,\cdot}) \prod_{k=1}^{K} B(\beta_k + c_{\cdot,k,:})\right]_{\neg i}} \\
&\propto \left[\frac{\prod_{m=1}^{M} B(\alpha_m + c_{m,:,\cdot})}{\prod_{m=1}^{M} B(\alpha_m + c_{m,:,\cdot})_{\neg i}}\right] \left[\frac{\prod_{k=1}^{K} B(\beta_k + c_{\cdot,k,:})}{\prod_{k=1}^{K} B(\beta_k + c_{\cdot,k,:})_{\neg i}}\right]
\end{aligned} \tag{4.2}
$$

Hence, the the distribution for $p(z_i = k)$ is proportional to the full joint distribution of the model divided by the joint considering the token/word w_i and its associated topical assignment did not exist in our data/model.

Observing that α remains fixed, we can try to find the expression for the first term of equation 4.2, i.e. for a fixed position n' in the document m' assigned to the topic k'.

$$
\begin{aligned}
p(z_i = k'|Z_{\neg i}, W) &\propto \left[\frac{\prod_{m=1}^{M} B(\alpha_m + c_{m,:,\cdot})}{\prod_{m=1}^{M} B(\alpha_m + c_{m,:,\cdot})_{\neg i}}\right] \\
&\propto \left[\frac{B(\alpha_{m'} + c_{m',:,\cdot})}{B(\alpha_{m'} + c_{m',:,\cdot})_{\neg i}}\right] \\
&\propto \left[\frac{B(\alpha_{m'} + c_{m',:,\cdot})}{B(\alpha_{m'} + [c_{m',:,\cdot}]_{\neg i})}\right]
\end{aligned}
$$

For all documents other than m', the numerator and denominator is the same. Hence, only the m' term remains in the numerator and denominator.

Now, let us see what changes happen in the count vector $c_{m',:,\cdot}$ with and without the term at $i = (m', n')$ whose latent topic assignment is $z_i = k'$.

$$
c_{m',k,\cdot} = \begin{cases} [c_{m',k,\cdot}]_{\neg i} & k \neq k' \\ [c_{m',k',\cdot}]_{\neg i} + 1 & k = k' \end{cases}
$$

Now, using the definition of Beta function,

$$\frac{B(\alpha'_m + c_{m',:,\cdot})}{B(\alpha'_m + [c_{m',:,\cdot}]_{\neg i})} = \frac{\prod_{k=1}^{K} \Gamma(\alpha_{m',k'} + c_{m',k'})}{\prod_{k=1}^{K} \Gamma(\alpha_{m',k'} + [c_{m',k'}]_{\neg i})} \times$$
$$\frac{\Gamma(\sum_{k \neq k'} (\alpha_{m',k} + [c_{m',k,\cdot}]_{\neg i}) + [\alpha_{m',k'} + [c_{m',k',\cdot}]_{\neg i}])}{\Gamma(\sum_{k \neq k'} (\alpha_{m',k} + c_{m',k,\cdot}) + [\alpha_{m',k'} + c_{m',k',\cdot}])}$$
$$= \frac{\prod_{k=1}^{K} \Gamma(\alpha_{m',k'} + [c_{m',k'}]_{\neg i} + 1)}{\prod_{k=1}^{K} \Gamma(\alpha_{m',k'} + [c_{m',k'}]_{\neg i})} \times$$
$$\frac{\Gamma(\sum_{k \neq k'} (\alpha_{m',k} + [c_{m',k,\cdot}]_{\neg i}) + [\alpha_{m',k'} + [c_{m',k',\cdot}]_{\neg i}])}{\Gamma(\sum_{k \neq k'} (\alpha_{m',k} + [c_{m',k,\cdot}]_{\neg i}) + [\alpha_{m',k'} + [c_{m',k',\cdot}]_{\neg i} + 1])}$$

Using the identity $\Gamma(x+1) = x\Gamma(x)$

$$\frac{B(\alpha'_m + c_{m',:,\cdot})}{B(\alpha_m + [c_{m',:,\cdot}]_{\neg i})} = \frac{\alpha_{m',k'} + [c_{m',k',\cdot}]_{\neg i}}{\sum_{k=1}^{K} (\alpha_{m',k} + [c_{m',k,\cdot}]_{\neg i})}$$

Similarly, we can simplify the second term of equation 4.2 to get,

$$\left[\frac{\prod_{k=1}^{K} B(\beta_k + c_{\cdot,k,:})}{\prod_{k=1}^{K} B(\beta_k + [c_{\cdot,k,:}]_{\neg i})} \right] = \frac{B(\beta_{k'} + c_{\cdot,k',:})}{B(\beta_{k'} + [c_{\cdot,k',:}]_{\neg i})} = \frac{\beta_{k',v'} + [c_{\cdot,k',v'}]_{\neg i}}{\sum_{v=1}^{V} (\beta_{k',v} + [c_{\cdot,k',v}]_{\neg i})}$$

where v' refers to the vocabulary token assigned to w_i. Hence, the full Gibbs conditional is as follows:

$$P(z_i = k' | Z_{\neg i}, W_{\neg i}, w_i) \propto \left[\frac{\alpha_{m',k'} + [c_{,m',k',\cdot}]_{\neg i}}{\sum_{k=1}^{K} (\alpha_{m',k} + [c_{m',k,\cdot}]_{\neg i})} \right] \times \left[\frac{\beta_{k',v'} + [c_{,k',v',\cdot}]_{\neg i}}{\sum_{v=1}^{V} (\beta_{k',v} + [c_{k',v,\cdot}]_{\neg i})} \right]$$

Chapter 5

Pseudocode for Collapsed Gibbs Sampling

⊡ initialisation
zero all count variables, $n_m^{(k)}, n_m, n_k^{(t)}, n_k$
for all documents $m \in [1, M]$ **do**
  for all words $n \in [1, N_m]$ in document m **do**
    sample topic index $z_{m,n}{=}k \sim \text{Mult}(1/K)$
    increment document–topic count: $n_m^{(k)} + 1$
    increment document–topic sum: $n_m + 1$
    increment topic–term count: $n_k^{(t)} + 1$
    increment topic–term sum: $n_k + 1$
  end for
end for
⊡ Gibbs sampling over burn-in period and sampling period
while not finished **do**
  for all documents $m \in [1, M]$ **do**
    for all words $n \in [1, N_m]$ in document m **do**
      ⊡ for the current assignment of k to a term t for word $w_{m,n}$:
      decrement counts and sums: $n_m^{(k)} - 1; n_m - 1; n_k^{(t)} - 1; n_k - 1$
      ⊡ multinomial sampling acc. to Eq. 79 (decrements from previous step):
      sample topic index $\tilde{k} \sim p(z_i|\vec{z}_{\neg i}, \vec{w})$
      ⊡ use the new assignment of $z_{m,n}$ to the term t for word $w_{m,n}$ to:
      increment counts and sums: $n_m^{(\tilde{k})} + 1; n_m + 1; n_{\tilde{k}}^{(t)} + 1; n_{\tilde{k}} + 1$
    end for
  end for
  ⊡ check convergence and read out parameters
  if converged and L sampling iterations since last read out **then**
    ⊡ the different parameters read outs are averaged.
    read out parameter set $\underline{\Phi}$ according to Eq. 82
    read out parameter set $\underline{\Theta}$ according to Eq. 83
  end if
end while

Figure 5.1: LDA using collapsed gibbs sampling