1 Background

$$\mathrm{Dir}(x \mid \alpha) = \frac{1}{B(\alpha)} \prod_i x_i^{\alpha_i - 1}$$

$$B(\alpha) = \frac{\prod_i \Gamma(\alpha_i)}{\Gamma(\sum_i \alpha_i)}$$

Define $x^\beta = \prod_i x_i^{\beta_i}$.

$$\begin{aligned}
&\int_x x^\beta \mathrm{Dir}(x \mid \alpha)\,\mathrm{d}x \\
=&\frac{1}{B(\alpha)} \int_x x^\beta \prod_i x_i^{\alpha_i - 1}\,\mathrm{d}x \\
=&\frac{1}{B(\alpha)} \int_x \prod_i x_i^{\alpha_i + \beta_i - 1}\,\mathrm{d}x \\
=&\frac{B(\alpha + \beta)}{B(\alpha)}
\end{aligned}$$

$$\frac{\Gamma(a+1)}{\Gamma(a)} = a$$

2 LDA

Assume K topics, M documents and N_m words for document m. The words are generated in the following way:

$$\begin{aligned}
\phi_k &\sim \mathrm{Dir}(\beta) & k &= 1..K \\
\theta_m &\sim \mathrm{Dir}(\alpha) & m &= 1..M \\
z_{mn} &\sim \mathrm{Multi}(\theta_m) & n &= 1..N_m,\ m = 1..M \\
w_{mn} &\sim \mathrm{Multi}(\phi_{z_{mn}}) & n &= 1..N_m,\ m = 1..M
\end{aligned}$$

$$\Pr[w, z, \theta, \phi \mid \alpha, \beta] = \left\{ \prod_k \Pr[\phi_k \mid \beta] \right\} \prod_m \left\{ \Pr[\theta_m \mid \alpha] \prod_n \left(\Pr[w_{mn} | \phi, z_{mn}] \Pr[z_{mn} | \theta_m] \right) \right\}$$

$$\Pr[z_{mn} = k \mid \theta_m] = \theta_{mk}$$

$$\Pr[w_{mn} = t \mid \phi_k, z_{mn} = k] = \phi_{kt}$$

$$\Pr[w_{mn} = t, z_{mn} = k \mid \theta_m, \phi] = \Pr[w_{mn} = t \mid \phi_k, z_{mn} = k] \Pr[z_{mn} = t \mid \theta_m] = \phi_{kt}\theta_{mk}$$

$$\Pr[w_m, z_m \mid \theta_m, \phi] = \prod_k \left\{ \theta_{mk}^{n_{mk}} \prod_t \phi_{kt}^{n_{mkt}} \right\}$$

where n_{mk} is the number of words in document m that has topic k, and n_{mkt} is the number of words t that has topic k in document m. We have

$$n_{mk} = \sum_t n_{mkt}.$$

$$\begin{aligned}
\Pr[w, z \mid \theta, \phi] = \prod_m \Pr[w_m, z_m | \theta_m, \phi] &= \prod_m \prod_k \left\{ \theta_{mk}^{n_{mk}} \prod_t \phi_{kt}^{n_{mkt}} \right\} \\
&= \left\{ \prod_{mk} \theta_{mk}^{n_{mk}} \right\} \left\{ \prod_{mkt} \phi_{kt}^{n_{mkt}} \right\} \\
&= \left\{ \prod_{mk} \theta_{mk}^{n_{mk}} \right\} \left\{ \prod_{kt} \phi_{kt}^{n_{kt}} \right\}
\end{aligned}$$

where

$$n_{kt} = \sum_m n_{mkt}$$

that is, number of term t assigned to topic k across all documents.

$$\begin{aligned}
\Pr[w, z \mid \alpha, \beta] &= \int_{\theta,\phi} \Pr[w, z, \theta, \phi \mid \alpha, \beta] \, \mathrm{d}\theta \, \mathrm{d}\phi \\
&= \int_{\theta,\phi} \Pr[w, z \mid \theta, \phi] \ \Pr[\theta \mid \alpha] \ \Pr[\phi \mid \beta] \, \mathrm{d}\theta \, \mathrm{d}\phi \\
&= \int_{\theta,\phi} \left\{ \prod_{mk} \theta_{mk}^{n_{mk}} \right\} \left\{ \prod_{kt} \phi_{kt}^{n_{kt}} \right\} \prod_m \mathrm{Dir}[\theta_m \mid \alpha] \prod_k \mathrm{Dir}[\phi_k \mid \beta] \, \mathrm{d}\theta \, \mathrm{d}\phi \\
&= \left\{ \int_\theta \prod_{mk} \theta_{mk}^{n_{mk}} \prod_m \mathrm{Dir}[\theta_m \mid \alpha] \, \mathrm{d}\theta \right\} \left\{ \int_\phi \prod_{kt} \phi_{kt}^{n_{kt}} \prod_k \mathrm{Dir}[\phi_k \mid \beta] \, \mathrm{d}\phi \right\} \\
&= \prod_m \left\{ \int_{\theta_m} \prod_k \theta_{mk}^{n_{mk}} \mathrm{Dir}[\theta_m \mid \alpha] \, \mathrm{d}\theta_m \right\} \prod_k \left\{ \int_{\phi_k} \prod_t \phi_{kt}^{n_{kt}} \mathrm{Dir}[\phi_k \mid \beta] \, \mathrm{d}\phi_k \right\} \\
&= \prod_m \frac{B(n_m + \alpha)}{B(\alpha)} \prod_k \frac{B(n_k + \beta)}{B(\beta)}
\end{aligned}$$

$$\Pr[w, z \mid \alpha, \beta] = \prod_m \frac{\prod_k \Gamma(n_{mk} + \alpha)}{\Gamma^K(\alpha)} \frac{\Gamma(K\alpha)}{\Gamma(\sum_k n_{mk} + K\alpha)} \prod_k \frac{\prod_t \Gamma(n_{kt} + \beta)}{\Gamma^T(\beta)} \frac{\Gamma(T\beta)}{\Gamma(\sum_t n_{kt} + T\beta)}$$

$$
\begin{aligned}
\Pr[w_{mn} = t \mid \alpha, \beta] &= \int_{\theta_m, \phi, z_{mn}=k} \Pr[t, k, \theta_m, \phi \mid \alpha, \beta] \, \mathrm{d}\theta_m \, \mathrm{d}\phi \\
&= \int_{\theta_m, \phi, z_{mn}=k} \Pr[t, k \mid \theta_m, \phi] \; \Pr[\theta_m \mid \alpha] \; \Pr[\phi \mid \beta] \, \mathrm{d}\theta \, \mathrm{d}\phi \\
&= \int_{\theta, \phi, z_{mn}=k} \phi_{kt} \theta_{mk} \mathrm{Dir}[\theta_m \mid \alpha] \mathrm{Dir}[\phi \mid \beta] \, \mathrm{d}\theta \, \mathrm{d}\phi \\
&= \sum_{z_{mn}=k} \left\{ \int_{\theta_m} \theta_{mk} \mathrm{Dir}[\theta_m \mid \alpha] \, \mathrm{d}\theta_m \right\} \left\{ \int_{\phi_k} \phi_{kt} \mathrm{Dir}[\phi_k \mid \beta] \, \mathrm{d}\phi_k \right\} \\
&= \sum_{z_{mn}=k} \frac{B(e_k + \alpha)}{B(\alpha)} \frac{B(e_t + \beta)}{B(\beta)}
\end{aligned}
$$

3 Gibbs Sampling

We already defined the follow two.

$$
\begin{aligned}
n_{mk} &= \text{number of words in document } m \text{ that has topic } k. \\
n_{tk} &= \text{number of term } t \text{ assigned to topic } k \text{ across all document.}
\end{aligned}
$$

We define the following two to be the same statistics without the term j of document i taken into consideration.

$$
n_{mk}^{\setminus ij} \qquad n_{kt}^{\setminus ij}
$$

Define

$$
\begin{aligned}
n_{mk}^{ij} &= \delta(m - i)\delta(k - z_{ij}) \\
n_{kt}^{ij} &= \delta(t - w_{ij})\delta(k - z_{ij})
\end{aligned}
$$

and we have

$$
n_{mk} = n_{mk}^{\setminus ij} + n_{mk}^{ij} \qquad n_{tk} = n_{kt}^{\setminus ij} + n_{kt}^{ij}.
$$

$$
\begin{aligned}
&\Pr[z_{ij} = c \,|\, z \setminus z_{ij}, w] \\
=&\frac{\Pr[z_{ij} = c, z \setminus z_{ij}, w]}{\Pr[z \setminus z_{ij}, w]} \\
=&\frac{\Pr[z_{ij} = c, z \setminus z_{ij}, w]}{\Pr[z \setminus z_{ij}, w \setminus w_{ij}] \Pr[w_{ij}]} \propto \frac{\Pr[z_{ij} = c, z \setminus z_{ij}, w]}{\Pr[z \setminus z_{ij}, w \setminus w_{ij}]} \\
=&\left\{ \prod_m \frac{B(n_m^{\setminus ij} + n_m^{ij} + \alpha)}{B(\alpha)} \prod_k \frac{B(n_k^{\setminus ij} + n_k^{ij} + \beta)}{B(\beta)} \right\} / \left\{ \prod_m \frac{B(n_m^{\setminus ij} + \alpha)}{B(\alpha)} \prod_k \frac{B(n_k^{\setminus ij} + \beta)}{B(\beta)} \right\} \\
=&\prod_m \frac{B(n_m^{\setminus ij} + n_m^{ij} + \alpha)}{B(n_m^{\setminus ij} + \alpha)} \prod_k \frac{B(n_k^{\setminus ij} + n_k^{ij} + \beta)}{B(n_k^{\setminus ij} + \beta)} \\
=&\frac{B(n_i^{\setminus ij} + n_i^{ij} + \alpha)}{B(n_i^{\setminus ij} + \alpha)} \frac{B(n_c^{\setminus ij} + n_c^{ij} + \beta)}{B(n_c^{\setminus ij} + \beta)} \\
\propto&\frac{B(n_i^{\setminus ij} + n_i^{ij} + \alpha) B(n_c^{\setminus ij} + n_c^{ij} + \beta)}{B(n_c^{\setminus ij} + \beta)} \\
=&\frac{\prod_k \Gamma[n_{ik}^{\setminus ij} + \delta(k-c) + \alpha_k]}{\Gamma\left\{\sum_k \left[n_{ik}^{\setminus ij} + \delta(k-c) + \alpha_k\right]\right\}} \frac{\prod_t \Gamma[n_{ct}^{\setminus ij} + \delta(t - w_{ij}) + \beta_t]}{\Gamma\left\{\sum_t \left[n_{ct}^{\setminus ij} + \delta(t - w_{ij}) + \beta_t\right]\right\}} \frac{\Gamma\left\{\sum_t [n_{ct}^{\setminus ij} + \beta_t]\right\}}{\prod_t \Gamma[n_{ct}^{\setminus ij} + \beta_t]} \\
=&\frac{\prod_k \Gamma[n_{ik}^{\setminus ij} + \delta(k-c) + \alpha_k]}{\Gamma\left\{\sum_k \left[n_{ik}^{\setminus ij} + \delta(k-c) + \alpha_k\right]\right\}} \frac{\prod_t \Gamma[n_{ct}^{\setminus ij} + \delta(t - w_{ij}) + \beta_t]}{\prod_t \Gamma[n_{ct}^{\setminus ij} + \beta_t]} \frac{\Gamma\left\{\sum_t [n_{ct}^{\setminus ij} + \beta_t]\right\}}{\Gamma\left\{\sum_t \left[n_{ct}^{\setminus ij} + \delta(t - w_{ij}) + \beta_t\right]\right\}} \\
\propto&\frac{\Gamma(n_{ic}^{\setminus ij} + 1 + \alpha_c)}{\Gamma(n_{ic}^{\setminus ij} + \alpha_c)} \frac{\Gamma(n_{cw_{ij}}^{\setminus ij} + 1 + \beta_{w_{ij}})}{\Gamma(n_{cw_{ij}}^{\setminus ij} + \beta_{w_{ij}})} \frac{\Gamma\left\{\sum_t [n_{ct}^{\setminus ij} + \beta_t]\right\}}{\Gamma\left\{1 + \sum_t \left[n_{ct}^{\setminus ij} + \beta_t\right]\right\}} \\
=&\frac{(n_{ic}^{\setminus ij} + \alpha_c)(n_{cw_{ij}}^{\setminus ij} + \beta_{w_{ij}})}{\sum_t \left[n_{ct}^{\setminus ij} + \beta_t\right]} \\
=&\frac{(n_{ic}^{\setminus ij} + \alpha_c)(n_{cw_{ij}}^{\setminus ij} + \beta_{w_{ij}})}{n_c^{\setminus ij} + \sum_t \beta_t}
\end{aligned}
$$

where $n_c^{\setminus ij}$ is the total number of terms under topic c across all documents, except for w_{ij}. Intuitively, topic c is more likely to be sampled if

- More words in the current documents are under topic c;
- More times the current word is assigned to topic c across all documents;
- Topic c is used less for all words.

4 Parameter Estimation

$$\begin{aligned}
&\Pr[\theta_m | w, z, \alpha] \\
&\propto \Pr[z_m \,|\, \theta_m] \Pr[\theta_m \,|\, \alpha] \\
&= \left\{ \prod_n \Pr[z_{mn} \,|\, \theta_m] \right\} \Pr[\theta_m \,|\, \alpha] \\
&= \left\{ \prod_k \theta_{mk}^{n_{mk}} \right\} \Pr[\theta_m \,|\, \alpha] \\
&\sim \mathrm{Dir}(n_m + \alpha)
\end{aligned}$$

$$\begin{aligned}
&\Pr[\phi_k \,|\, w, z, \beta] \\
&\propto \Pr[w \,|\, z, \phi_k] \Pr[\phi_k \,|\, \beta] \\
&\propto \left\{ \prod_{mn : z_{mn} = k} \Pr[w_{mn} \,|\, \phi_k] \right\} \Pr[\phi_k \,|\, \beta] \\
&\propto \left\{ \prod_t \phi_{kt}^{n_{kt}} \right\} \Pr[\phi_k \,|\, \beta] \\
&\sim \mathrm{Dir}(n_k + \beta)
\end{aligned}$$

5 Log-Likelihood