CMSC 25025
Assignment 5

Luis Buenaventura

Problem 1

a) We wish to derive: $\mathbb{P}(z_n \mid z_{-n}, \beta_{1:K}, W_{1:N}, \alpha)$.

We can express as:

$$\mathbb{P}(z_n \mid z_{-n}, \beta_{1:K}, W_{1:N}, \alpha) = \frac{\mathbb{P}(z_{1:N}, \beta_{1:K}, W_{1:N}, \alpha)}{\mathbb{P}(z_{-n}, \beta_{1:K}, W_{1:N}, \alpha)}$$

We recall the joint distribution:

$$\mathbb{P}(z_{1:D,1:N}, W_{1:D,1:N}, \beta_{1:K}, \theta_{1:D}) = \prod_{k=1}^{K} p(\beta_k \mid \eta) \prod_{d=1}^{D} p(\theta_d \mid \alpha) \prod_{d=1}^{D} \prod_{n=1}^{N} p(z_{d,n} \mid \theta_d) \prod_{d=1}^{D} \prod_{n=1}^{N} p(w_{d,n} \mid \beta_{1:K}, z_{d,n})$$

Which we use to derive

$$\mathbb{P}(z_{1:N}, w_{1:N}, \beta_{1:K}, \theta, \alpha) = \text{(for some document } d\text{)}$$

$$p_\alpha(\theta) \prod_{n=1}^{N} p(z_n \mid \theta)\, p(w_n \mid \beta_{1:K}, z_n)$$

So

$$\mathbb{P}(z_{1:N}, W_{1:N}, \beta_{1:K}, \alpha) = \int_\theta p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta)\, p(w_n \mid \beta_{1:K}, z_n)\, d\theta$$

$$= \int_\theta \mathrm{Dir}(\theta, \alpha) \prod_{w=1}^{V} \left(\prod_{k=1}^{K} \theta_k^{n_{kw}} \right) \prod_{n=1}^{V} \left(\beta_k^{n_{kw}} \right)$$

$$= \int_\theta \mathrm{Dir}(\theta, \alpha) \prod_{k=1}^{K} \left(\theta_k^{n_{k\cdot}} \right) \prod_{n=1}^{V} \left(\beta_k^{n_{kn}} \right) d\theta$$

Let $n_{k\cdot} = \sum_{w}^{N} n_{kw}$, $n_{\cdot\cdot} = \sum_{k}^{K} \sum_{w} n_{kw}$

$$= \int_\theta \mathrm{Dir}(\theta, \alpha) \prod_{k=1}^{K} \theta^{n_{k\cdot}} \prod_{w=1}^{V} \beta_k^{n_{kw}}\, d\theta$$

$$= \prod_{k=1}^{K} \prod_{w=1}^{N} \beta_k^{n_{kw}} \times \int_\theta \prod_{n=1}^{K} \frac{1}{B(\alpha)} \theta^{n_{k\cdot} + \alpha_k - 1} d\theta$$

$$= \prod_{k=1}^{K} \prod_{w=1}^{V} \beta_k^{n_{kw}} \times \left[\frac{B(\sum_k n_k + \alpha_k)}{B(\alpha)} \right] \int_\theta \frac{1}{B(\sum_k n_k + \alpha_k)} \theta_k^{n_{\cdot\cdot} + \alpha_k - 1} d\theta$$

$$= \frac{B(n_{\cdot\cdot} + \alpha_k)}{B(\alpha)} \left[\prod_{k=1}^{K} \beta_k^{n_{k\cdot}} \right] = \chi \quad (1)$$

Also,

$$p(z_{-n}, \beta_{1:K}, W_{1:N}, \alpha) = \mathbb{P}(z_{-n}, W_{-n}, \beta_{1:K}, \alpha) \left[\sum_{k=1}^{K} \mathbb{P}(z_{-i} = k, W_{-i}) \right]$$

Let $\xi = \sum_{k=1}^{K} \mathbb{P}(z_{-i} = k, W_{-i})$

We can generalize (1) for

$$\gamma = \mathbb{P}(z_{-n}, W_{-n}, \beta_{1:K}, \alpha) = \frac{B(n_{\cdot\cdot}^{\setminus d_i} + \alpha)}{B(\alpha)} \left[\prod_{k=1}^{K} \beta_k^{n_{k\cdot}^{\setminus d_i}} \mathrm{Mult}(\beta_k) \right]$$

So, $\mathbb{P}(z_n \mid z_{-n}, w_{1:N}, \beta_{1:K}, \alpha)$ can be calculated as $= \dfrac{\gamma}{\xi \cdot \gamma}$ from our earlier calculations.

b) We can think of $\mathbb{E}(\theta_d \mid \beta_{1:K}, w_{1:N}, \alpha)$ as the expected value of the distribution of K topics in document d given the observable $w_{1:N}$ in document d and the prior assumption about the distribution of the K topics as well as the distribution of words in each topic.

If we use (a) to find $\mathbb{P}(z_n \mid z_{-n}, w_{1:N}, \beta_{1:K}, \alpha)$, which allows us to estimate the topic proportions of document d by evaluating at all words, then we can enumerate topic counts z_n to estimate $\theta_{d,k} \mid w_{1:N}, \beta_{1:K}, \alpha$.

From (d) we see that $\theta_{d,k} \propto \dfrac{n_{dk\cdot} + \alpha}{\sum_{k=1}^{K} n_{dk\cdot} + \alpha}$.

c) $$\mathbb{P}(w_{1:N}, z_{1:N} \mid \beta_{1:K}, \alpha) = \int_{\theta} \mathbb{P}(w_{1:N}, z_{1:N} \mid \theta_d, \beta_{1:K})\, p(\theta_d \mid \alpha)\, d\theta_d$$

$$= \int_{\theta} \mathbb{P}(w_{1:N} \mid z_{1:N}, \beta_{1:K})\, \mathbb{P}(z_{1:N} \mid \theta_d)\, p(\theta_d \mid \alpha)\, d\theta_d$$

$$= \int_{\theta} \prod_{k=1}^{K} \prod_{w=1}^{V} (\beta^{n_{kw}}) \left(\prod_{k=1}^{K} \left(\prod_{w=1}^{V} \theta_k^{n_{kv}} \right) \right) \mathrm{Dir}(\alpha)\, d\theta$$

$$= \frac{B(n_{\cdot\cdot} + \alpha)}{B(\alpha)} \left(\prod_{k=1}^{K} \beta_k^{n_{k\cdot}} \right)$$ from before

d) We now attempt to derive the conditional probability

$$p(z_n \mid z_{-n}, w, \alpha, \eta)$$ across all documents.

Recall:

$$p(z_n \mid z_{-n}, w, \alpha, \eta) = \frac{p(z, w, \alpha, \eta)}{p(z_{-n}, w, \alpha, \eta)} = \frac{p(z, w, \alpha, \eta)}{p(z_{-n}, w_{-n}, \alpha, \eta) \sum_{k=1}^{K} p(z_{-n}=k, w_{-n})}$$

where $n_{dk\cdot} = \sum_w n_{dkw}$, $n_{\cdot kw} = \sum_d n_{dkw}$, $Dim(\alpha) = K$, $Dim(\eta) = V$:

$$p(z, w, \alpha, \eta) = \int_\theta \int_\beta p(w, z, \theta_{1:D}, \beta_{1:K}, \alpha, \eta)\, d\theta\, d\beta$$

$$= \int_\theta \int_\beta p(\theta^d, \alpha) p(z \mid \theta) p(\beta, \eta) p(w \mid z, \beta)\, d\theta\, d\beta$$

$$= \int_\theta \int_\beta \prod_{d=1}^{D} Dir(\theta^d, \alpha) \prod_{k=1}^{K} Dir(\beta_k, \eta) \prod_{d=1}^{D} \prod_{k=1}^{K} \prod_{w=1}^{V} (\theta_k^d)^{n_{dkw}} \prod_{k=1}^{K} \prod_{d=1}^{D} \prod_{w=1}^{V} ((\beta_{k,w})^{n_{dkw}})\, d\theta\, d\beta$$

Due to independence:

$$= \int_{\beta_{1:K}} \prod_{k=1}^{K} Dir(\beta_k, \eta) \prod_{k=1}^{K} \prod_{d=1}^{D} \prod_{w=1}^{V} (\beta_{k,w})^{n_{dkw}}\, d\beta$$

$$\int_{\theta_{1:D}} \prod_{d=1}^{D} Dir(\theta^d, \alpha) \prod_{k=1}^{D} \prod_{k=1}^{K} (\theta_{d,k})^{n_{dkw}}\, d\theta$$

$$= \int_{\beta_{1:K}} \prod_{k=1}^{K} \frac{1}{B(\eta)} \prod_{k=1}^{K} \prod_{w=1}^{V} \beta_{k,w}^{\eta-1} \prod_{w=1}^{V} \beta_{k,w}^{n_{\cdot kw}}$$

$$\int_{\theta_{1:D}} \prod_{d=1}^{D} \frac{1}{B(\alpha)} \prod_{d=1}^{D} \prod_{k=1}^{K} \theta_{d,k}^{\alpha-1} \prod_{k=1}^{K} \theta_{d,k}^{n_{dk\cdot}}$$

$$= \int_{\beta_{1:K}} \prod_{k=1}^{K} \left[\frac{1}{B(\eta)} \prod_{w=1}^{V} \beta_{k,w}^{\eta-1+n_{\cdot kw}} \right] d\beta \int_{\theta_{1:K}} \prod_{d=1}^{D} \left(\frac{1}{B(\alpha)} \prod_{k=1}^{K} \theta_{d,k}^{\alpha-1+n_{dk\cdot}} \right) d\theta_{1:K}$$

$$= \prod_{k=1}^{K} \frac{B(\eta+n_{\cdot k\cdot})}{B(\eta)} \int_{\beta_{1:K}} \frac{1}{B(\eta+n_{\cdot k\cdot})} \prod_{w=1}^{V} \beta_{k,w}^{\eta-1+n_{\cdot kw}}\, d\beta_{1:K}$$

$$\prod_{d=1}^{D} \frac{B(\alpha+n_{d\cdot\cdot})}{B(\eta)} \int_{\theta_{1:K}} \frac{1}{B(\alpha+n_{d\cdot\cdot})} \prod_{d=1}^{D} \theta_{d,k}^{\alpha-1+n_{d\cdot\cdot}}\, d\theta_{1:K}$$

$$= \prod_{k=1}^{K} \frac{B(\eta+n_{\cdot k\cdot})}{B(\eta)} \prod_{d=1}^{D} \frac{B(\alpha+n_{d\cdot\cdot})}{B(\eta)}$$

We can evaluate $\mathbb{P}(z_{-n}, w_{-n} \mid \alpha, \eta)$ as =

$$\prod_{k=1}^{K} \frac{B(\eta + n^{\setminus di}_{\cdot k \cdot})}{B(\eta)} \prod_{d=1}^{D} \frac{B(\alpha + n^{\setminus di}_{d \cdot \cdot})}{B(\alpha)}$$

where n^{-n}_{dkv} omits word "n" from the count

So,

$$\mathbb{P}(z_n \mid z_{-n}, w, \alpha, \eta) \propto \prod_{k=1}^{K} \frac{B(\eta + n_{\cdot k \cdot})}{B(\eta + n^{\setminus di}_{\cdot k \cdot})} \prod_{d=1}^{D} \frac{B(\alpha + n_{d \cdot \cdot})}{B(\alpha + n^{\setminus di}_{d \cdot \cdot})}$$

$$\prod_{k=1}^{K} \frac{B(\eta + n^{\setminus di}_{\cdot k \cdot} + e_v)}{B(\eta + n^{\setminus di}_{\cdot k \cdot})} \prod_{d=1}^{D} \frac{B(\alpha + n^{\setminus di}_{d \cdot \cdot} + e_d)}{B(\alpha + n^{\setminus di}_{\cdot k \cdot})}$$ where e_i is a basis vector.

Let $x = \eta + n^{\setminus di}_{\cdot k \cdot}$ $\quad y = \alpha + n^{\setminus di}_{d \cdot \cdot}$

So,

$$\frac{B(x + e_v)}{B(x)} \times \frac{B(y + e_d)}{B(y)} = \prod_{v^*} \frac{\Gamma(x_{v^*}) \Gamma(x_v + 1)}{\Gamma(\sum_v x_v + 1)} \prod_{v^*} \frac{\Gamma(\sum x_v)}{\Gamma(x_{v^*}) \Gamma(x_v)}$$

$$\times \prod_{d^*} \frac{\Gamma(y_{d^*}) \Gamma(y_d + 1)}{\Gamma(\sum_d y_d + 1)} \prod_{d} \frac{\Gamma(\sum y_d)}{\Gamma(y_{d^*}) \Gamma(y_d)}$$

Using $z\Gamma(z) = \Gamma(z+1)$;

$$= \frac{\Gamma(\sum_v x_v)(x_v)\Gamma(x_v)}{\Gamma(x_v)(\sum_v x_v)(\Gamma(\sum_v x_v))} \cdot \frac{\Gamma(\sum_d y_d)(y_d)\Gamma(y_d)}{\Gamma(y_d)(\sum y_d)\Gamma(\sum_d y_d)}$$

$$= \frac{x_v}{\sum_v x_v} \times \frac{y_d}{\sum_k y_k} = \frac{\eta + n^{\setminus di}_{\cdot kv}}{\sum_v (n^{\setminus di}_{\cdot kv} + \eta)} \times \frac{\alpha + n^{\setminus di}_{d \cdot \cdot}}{\sum_k (n^{\setminus di}_{dk\cdot} + \alpha)}$$