1. **Algorithm 1:** Use the (collapsed) Gibbs sampling algorithm to approximate (using samples) the posterior distribution $p(\boldsymbol{Z}, \boldsymbol{\Theta} \mid \boldsymbol{X})$. Derive the posterior.

$\underline{p(z_n = k \mid \mathbf{x}_n, \mathbf{Z}_{-n}, \boldsymbol{\Theta}) :}$

$$\begin{aligned} p(z_n = k \mid \mathbf{x}_n, \mathbf{Z}_{-n}, \boldsymbol{\Theta}) &\propto p(z_n = k, \mathbf{x}_n \mid \mathbf{Z}_{-n}, \boldsymbol{\Theta}) \\ &= p(z_n = k \mid \mathbf{Z}_{-n}, \cancel{\boldsymbol{\Theta}}) p(\mathbf{x}_n \mid z_n = k, \mathbf{Z}_{-n}, \boldsymbol{\Theta}) \end{aligned}$$

$\underline{1 \leq k \leq K^+}$:

From Lecture 4:

$$p(z_n = k \mid \mathbf{Z}_{-n}) = \frac{m_k}{N - 1 + \alpha} \quad \text{, where } m_k = \sum_{i=1}^{n-1} [z_i = k]$$

$$p(\mathbf{x}_n \mid z_n = k, \cancel{\mathbf{Z}_{-n}}, \boldsymbol{\Theta}) = p(\mathbf{x}_n \mid \boldsymbol{\theta}_k) = \prod_{i=1}^{W_n} \text{Cat}(x_{ni} \mid \boldsymbol{\theta}_k)$$

$$\begin{aligned} \Rightarrow p(z_n = k \mid \mathbf{x}_n, \mathbf{Z}_{-n}, \boldsymbol{\Theta}) &\propto \frac{m_k}{N - 1 + \alpha} \cdot \prod_{i=1}^{W_n} \text{Cat}(x_{ni} \mid \boldsymbol{\theta}_k) \\ &= \frac{m_k}{N - 1 + \alpha} \cdot \prod_{m=1}^{|I|} \theta_{km}^{\sum_{i=1}^{W_n} [x_{ni} = m]} \end{aligned}$$

$$\begin{aligned} \Rightarrow \log p(z_n = k \mid \mathbf{x}_n, \mathbf{Z}_{-n}, \boldsymbol{\Theta}) &\propto \log p(z_n = k \mid \mathbf{Z}_{-n}, \cancel{\boldsymbol{\Theta}}) + \log p(\mathbf{x}_n \mid z_n = k, \mathbf{Z}_{-n}, \boldsymbol{\Theta}) \\ &= \log m_k - \log(N - 1 + \alpha) + \sum_{m=1}^{|I|} \sum_{i=1}^{W_n} [x_{ni} = m] \log \theta_{km} \end{aligned}$$

$\underline{k = K_{new}}$:

From Lecture 4:

$$p(z_n = K_{new} \mid \mathbf{Z}_{-n}) = \frac{\alpha}{N - 1 + \alpha}$$

$p(\mathbf{x}_n \mid z_n = K_{new}, \mathbf{Z}_{-n}, \boldsymbol{\Theta}) = p(\mathbf{x}_n \mid z_n = K_{new})$,

since $\mathbf{Z}_{-n}$ contains no assignments to K_{new} and $\boldsymbol{\Theta}$ does not contain $\boldsymbol{\theta}_{K_{new}}$

$$\begin{aligned} &= \int p(\mathbf{x}_n, \boldsymbol{\theta} \mid z_n = K_{new}) d\boldsymbol{\theta} \\ &= \int p(\mathbf{x}_n \mid z_n = K_{new}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{\gamma}) d\boldsymbol{\theta} \end{aligned}$$

$$
\begin{aligned}
p(\mathbf{x}_n \mid z_n = K_{new}) &= \int p(\mathbf{x}_n \mid z_n = K_{new}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{\gamma}) d\boldsymbol{\theta} \\
&= \int \mathrm{Dir}(\boldsymbol{\theta} \mid \boldsymbol{\gamma}) \prod_{i=1}^{W_n} \mathrm{Cat}(x_{ni} \mid \boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int \frac{1}{\mathrm{B}(\boldsymbol{\gamma})} \left(\prod_{m=1}^{|I|} \theta_m^{\gamma_m - 1} \right) \left(\prod_{m'=1}^{|I|} \theta_{m'}^{\sum_{i=1}^{W_n} [x_{ni} = m']} \right) d\boldsymbol{\theta} \\
&= \int \frac{\mathrm{B}(\boldsymbol{\gamma} + \boldsymbol{s})}{\mathrm{B}(\boldsymbol{\gamma})} \frac{1}{\mathrm{B}(\boldsymbol{\gamma} + \boldsymbol{s})} \prod_{m=1}^{|I|} \theta_m^{\gamma_m + s_m - 1} d\boldsymbol{\theta} \quad \text{, where } s_m = \sum_{i=1}^{W_n} [x_{ni} = m] \\
&= \frac{\mathrm{B}(\boldsymbol{\gamma} + \boldsymbol{s})}{\mathrm{B}(\boldsymbol{\gamma})} \int \mathrm{Dir}(\boldsymbol{\theta} \mid \boldsymbol{\gamma} + \boldsymbol{s}) d\boldsymbol{\theta} \\
&= \frac{\mathrm{B}(\boldsymbol{\gamma} + \boldsymbol{s})}{\mathrm{B}(\boldsymbol{\gamma})} \\
&= \frac{\prod_{m=1}^{|I|} \prod_{i=0}^{s_m - 1} (\gamma_m + i)}{\prod_{j=0}^{W_n - 1} \left(i + \sum_{m'=1}^{|I|} \gamma_{m'} \right)}, \qquad \text{(as shown in Tutorial 3)}
\end{aligned}
$$

$$
\Rightarrow p(z_n = K_{new} \mid \mathbf{x}_n, \mathbf{Z}_{-n}, \boldsymbol{\Theta}) \propto \frac{\alpha}{N - 1 + \alpha} \cdot \frac{\prod_{m=1}^{|I|} \prod_{i=0}^{s_m - 1} (\gamma_m + i)}{\prod_{j=0}^{W_n - 1} \left(i + \sum_{m'=1}^{|I|} \gamma_{m'} \right)}
$$

$\underline{p(\boldsymbol{\theta} \mid \mathbf{x}_n):}$

$$
\begin{aligned}
p(\boldsymbol{\theta} \mid \mathbf{x}_n) &\propto p(\mathbf{x}_n \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) = p(\mathbf{x}_n \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{\gamma}) \\
&= \prod_{i=1}^{W_n} \mathrm{Cat}(x_{ni} \mid \boldsymbol{\theta}) \, \mathrm{Dir}(\boldsymbol{\theta} \mid \boldsymbol{\gamma}) \\
&= \prod_{m=1}^{|I|} \theta_m^{\sum_{i=1}^{W_n} [x_{ni} = m]} \frac{1}{\mathrm{B}(\boldsymbol{\gamma})} \prod_{m'=1}^{|I|} \theta_{m'}^{\gamma_{m'} - 1} \\
&= \frac{1}{\mathrm{B}(\boldsymbol{\gamma})} \prod_{m=1}^{|I|} \theta_m^{\gamma_m + s_m - 1}, \quad \text{where } s_m = \sum_{i=1}^{W_n} [x_{ni} = m] \\
&\propto \mathrm{Dir}(\boldsymbol{\theta} \mid \boldsymbol{\gamma} + \boldsymbol{s})
\end{aligned}
$$

$$
\Rightarrow p(\boldsymbol{\theta} \mid \mathbf{x}_n) = \mathrm{Dir}(\boldsymbol{\theta} \mid \boldsymbol{\gamma} + \boldsymbol{s})
$$

$\underline{p(\boldsymbol{\theta}_k \mid \boldsymbol{X}, \boldsymbol{Z}):}$

$$
\begin{aligned}
p(\boldsymbol{\theta}_k \mid \boldsymbol{X}, \boldsymbol{Z}) &\propto p(\boldsymbol{\theta}_k, \boldsymbol{X} \mid \boldsymbol{Z}) = p(\boldsymbol{\theta}_k) p(\boldsymbol{X} \mid \boldsymbol{\theta}_k, \boldsymbol{Z}) \\
&= \mathrm{Dir}(\boldsymbol{\theta}_k \mid \boldsymbol{\gamma}) \prod_{n=1}^{N} p(\mathbf{x}_n \mid \boldsymbol{\theta}_k, z_n) \\
&= \mathrm{Dir}(\boldsymbol{\theta}_k \mid \boldsymbol{\gamma}) \prod_{n=1}^{N} p(\mathbf{x}_n \mid \boldsymbol{\theta}_k)^{[z_n = k]} \\
&= \mathrm{Dir}(\boldsymbol{\theta}_k \mid \boldsymbol{\gamma}) \prod_{n=1}^{N} \prod_{i=1}^{W_n} \mathrm{Cat}(x_{ni} \mid \boldsymbol{\theta}_k)^{[z_n = k]} \\
&= \mathrm{Dir}(\boldsymbol{\theta}_k \mid \boldsymbol{\gamma}) \prod_{n=1}^{N} \prod_{i=1}^{W_n} \prod_{m=1}^{|I|} \theta_{km}^{[x_{ni} = m][z_n = k]} \\
&= \mathrm{Dir}(\boldsymbol{\theta}_k \mid \boldsymbol{\gamma}) \prod_{m=1}^{|I|} \theta_{km}^{\sum_{n=1}^{N} [z_n = k] \sum_{i=1}^{W_n} [x_{ni} = m]} \\
&= \prod_{m=1}^{|I|} \theta_{km}^{\gamma_m + c_{km} - 1}, \quad \text{where, } c_{km} = \sum_{n=1}^{N} [z_n = k] \sum_{i=1}^{W_n} [x_{ni} = m] \\
&\propto \mathrm{Dir}(\boldsymbol{\theta}_k \mid \boldsymbol{\gamma} + \boldsymbol{c}_k)
\end{aligned}
$$

$$\Rightarrow p(\boldsymbol{\theta}_k \mid \boldsymbol{X}, \boldsymbol{Z}) = \mathrm{Dir}(\boldsymbol{\theta}_k \mid \boldsymbol{\gamma} + \boldsymbol{c}_k)$$

2. **Algorithm 2:** Use the (collapsed) Gibbs sampling algorithm to approximate (using samples) the posterior distribution $p(\boldsymbol{Z} \mid \boldsymbol{X})$. Derive the posterior.

$\underline{p(z_n = k \mid \mathbf{x}_n, \mathbf{Z}_{-n}) :}$

$$\begin{aligned} p(z_n = k \mid \mathbf{x}_n, \mathbf{Z}_{-n}) &\propto p(z_n = k, \mathbf{x}_n \mid \mathbf{X}_{-n}, \mathbf{Z}_{-n}) \\ &= p(z_n = k \mid \mathbf{Z}_{-n}) p(\mathbf{x}_n \mid z_n = k, \mathbf{X}_{-n}, \mathbf{Z}_{-n}) \end{aligned}$$

$\underline{1 \leq k \leq K^+}$:

From Algorithm 1:

$$p(z_n = k \mid \mathbf{Z}_{-n}) = \frac{m_k}{N - 1 + \alpha} \quad \text{, where } m_k = \sum_{i=1}^{n-1} [z_i = k]$$

From Tutorial 3:

$$p(\mathbf{x}_n \mid z_n = k, \mathbf{X}_{-n}, \mathbf{Z}_{-n}) = \frac{\prod_{m=1}^{|I|} \prod_{i=0}^{s_m - 1} (\gamma'_m + i)}{\prod_{i=0}^{W_n - 1} (i + \sum_{m'=1}^{|I|} \gamma'_{m'})}, \text{ where } \gamma'_m = \gamma_m + \sum_{i \neq n} \sum_{j=1}^{W_n} [x_{ij} = m][z_i = k]$$

$$\Rightarrow p(z_n = k \mid \mathbf{x}_n, \mathbf{Z}_{-n}) \propto \frac{m_k}{N - 1 + \alpha} \cdot \frac{\prod_{m=1}^{|I|} \prod_{i=0}^{s_m - 1} (\gamma'_m + i)}{\prod_{i=0}^{W_n - 1} (i + \sum_{m'=1}^{|I|} \gamma'_{m'})}$$

$\underline{k = K_{new} :}$

From Algorithm 1:

$$p(z_n = K_{new} \mid \mathbf{Z}_{-n}) = \frac{\alpha}{N - 1 + \alpha}$$

$$\begin{aligned} p(\mathbf{x}_n \mid z_n = K_{new}, \mathbf{X}_{-n}, \mathbf{Z}_{-n}) &= p(\mathbf{x}_n \mid z_n = K_{new}), \\ &\qquad \text{since } \mathbf{Z}_{-n} \text{ does not contain assignments to } K_{new} \\ &= \frac{\prod_{m=1}^{|I|} \prod_{i=0}^{s_m - 1} (\gamma_m + i)}{\prod_{j=0}^{W_n - 1} \left(i + \sum_{m'=1}^{|I|} \gamma_{m'}\right)} \end{aligned}$$

$$\Rightarrow p(z_n = k \mid \mathbf{x}_n, \mathbf{Z}_{-n}) \propto \frac{\alpha}{N - 1 + \alpha} \cdot \frac{\prod_{m=1}^{|I|} \prod_{i=0}^{s_m - 1} (\gamma_m + i)}{\prod_{j=0}^{W_n - 1} \left(i + \sum_{m'=1}^{|I|} \gamma_{m'}\right)}$$