General Supplement

In this supplement, we will show the step by step computation of every expectation from the lower bound. Then, we give the derivation of the standard updates equations in Algorithm 1 that are not described in the paper. Recall the lower bound:

$$
\begin{aligned}
&\mathscr{L} = \mathbb{E}[\log\ p(\mathbf{w}|c,z,\phi)p(c|\beta')p(z|\pi')p(\phi|\eta)p(\beta'|\gamma)p(\pi|\alpha_0)p(\mathbf{y}|\mathbf{z},\mathbf{c},\mu)]\,\underset{q}{\mathbb{E}}[\log\ q(\beta',\pi',c,z,\phi)] \\
&= \sum_{j}^{M} \big(\underset{q}{\mathbb{E}}[\log p(\mathbf{w_j}|\mathbf{c_j},\mathbf{z_j},\phi)] + \underset{q}{\mathbb{E}}[\log p(\mathbf{c_j}|\beta')] + \underset{q}{\mathbb{E}}[\log\ p(\mathbf{z_j}|\pi_\mathbf{j})] + \underset{q}{\mathbb{E}}[\log\ p(\pi'_j|\alpha_0)] \\
&+ \underset{q}{\mathbb{E}}[\log p(y_j|\mathbf{z}_j,\mathbf{c}_j,\mu)]\big) + \underset{q}{\mathbb{E}}[\log\ p(\phi|\eta)] + \underset{q}{\mathbb{E}}[\log\ p(\beta'|\gamma)] - \sum_{j=1}^{M} \big(\underset{q}{\mathbb{E}}[\log\ q(\pi_j)] + \underset{q}{\mathbb{E}}[\log\ q(c_j)] \\
&+ \underset{q}{\mathbb{E}}[\log\ q(z_j)]\big) - \underset{q}{\mathbb{E}}[\log\ q(\beta')] - \underset{q}{\mathbb{E}}[\log\ q(\phi)].
\end{aligned}
\tag{1}
$$

We compute every term in the same order as they show in Equation (1):

1st term:

$$
\begin{aligned}
&\underset{q}{\mathbb{E}}[\log p(\mathbf{w_j}|\mathbf{c_j},\mathbf{z_j},\phi)] \\
&= \sum_{n=1}^{N} \underset{q}{\mathbb{E}}[\log p(\mathbf{w_{jn}}|\mathbf{c_j},\mathbf{z_{jn}},\phi)] \\
&= \sum_{n=1}^{N} \underset{q}{\mathbb{E}}[\log \prod_{t=1}^{T} p(\mathbf{w_{jn}}|\mathbf{c_{jt}},\phi)^{[z_{jn}=t]}] \\
&= \sum_{n=1}^{N} \sum_{t=1}^{T} \underset{q}{\mathbb{E}}[[z_{jn}=t] \log p(\mathbf{w_{jn}}|\mathbf{c_{jt}},\phi)] \\
&= \sum_{n=1}^{N} \sum_{t=1}^{T} \underset{q}{\mathbb{E}}[[z_{jn}=t] \log \prod_{k=1}^{K} p(\mathbf{w_{jn}}|\phi_\mathbf{k})^{[c_{jt}=k]}] \\
&= \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K} \underset{q}{\mathbb{E}}[[z_{jn}=t][c_{jt}=k] \log p(\mathbf{w_{jn}}|\phi_\mathbf{k})] \\
&= \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K} \zeta_{jnt}\rho_{jtk} \sum_{i=1}^{V} (\Psi(\lambda_{ki}) - \Psi(\sum_{p}\lambda_{kp}))[w_{jn}=i] \\
&= \sum_{n=1}^{N} \sum_{i=1}^{V} \sum_{t=1}^{T} \sum_{k=1}^{K} (\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V}\lambda_{kp}))\zeta_{jnt}\rho_{jtk}[w_{jn}=i],
\end{aligned}
\tag{2}
$$

where Ψ is the Digamma function, which is the first order derivative of log Gamma

function.
2nd term:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{q}[\log p(\mathbf{c_j}|\beta')] \\
&= \sum_{t=1}^{T} \mathop{\mathbb{E}}_{q}[\log p(c_{jt}|\beta')] \\
&= \sum_{t=1}^{T} \mathop{\mathbb{E}}_{q}[\log \prod_{k=1}^{K} \big(\beta'_k \prod_{l=1}^{k-1}(1-\beta'_l)\big)^{1[c_{jt}=k]}] \\
&\textit{since } q(c_{jt}=k|\rho_{jt}) = \rho_{jtk} \\
&= \sum_{t=1}^{T}\sum_{k=1}^{K} \rho_{jtk}\Big(\mathop{\mathbb{E}}_{q}[\log \beta'] + \sum_{l=1}^{k-1}\mathop{\mathbb{E}}_{q}[\log(1-\beta'_l)]\Big) \\
&= \sum_{t=1}^{T}\sum_{k=1}^{K} \rho_{jtk}\Big((\Psi(v_{1k}) - \Psi(v_{1k}+v_{2k})) + \sum_{l=1}^{k-1}(\Psi(v_{2l}) - \Psi(v_{1l}+v_{2l}))\Big)
\end{aligned} \tag{3}
$$

Theorem: For any two K sized vectors $\mathbf{a}$ and $\mathbf{b}$ with real value, $\sum_{k=1}^{K} a_k(\sum_{j=1}^{k-1} b_j) = \sum_{k=1}^{K} b_k(\sum_{j=k+1}^{K} a_j)$.

Proof:

$$\sum_{k=1}^{K} a_k(\sum_{j=1}^{k-1} b_j) = \sum_{k=1}^{K}\sum_{j=1}^{K} b_j a_k 1[b_j < a_k] \tag{4}$$

$$\sum_{k=1}^{K} b_k(\sum_{j=k+1}^{K} a_j) = \sum_{k=1}^{K}\sum_{j=1}^{K} b_k a_j 1[b_k < a_j] \tag{5}$$

The right hand side of the Equations (5) and (4) are identical. □

Using the Theorem above, let $a_k = \Psi(v_{1k}) - \Psi(v_{1k}+v_{2k})$ and $b_k = \rho_{jtk}$ we can rewrite the result of Equation (3) as:

$$\mathop{\mathbb{E}}_{q}[\log p(\mathbf{c_j}|\beta')] = \sum_{t=1}^{T}\sum_{k=1}^{K}\Big(\sum_{i=k+1}^{K} \rho_{jti}(\Psi(v_{2k}) - \Psi(v_{1k}+v_{2k})) + \rho_{jtk}(\Psi(v_{1k}) - \Psi(v_{1k}+v_{2k}))\Big) \tag{6}$$

3rd term in analog to the 2nd term:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{q}[\log\ p(\mathbf{z_j}|\pi_\mathbf{j})] = \sum_{n=1}^{N}\sum_{t=1}^{T}\Big(\sum_{i=t+1}^{T} \zeta_{jni}(\Psi(a_{2jt}) - \Psi(a_{1jt}+a_{2jt})) + \zeta_{jnt}(\Psi(a_{1jt}) - \Psi(a_{1jt}+a_{2jt}))\Big) \\
&\text{rewrite} \\
&= \sum_{n=1}^{N}\sum_{t=1}^{T} \zeta_{jnt}\Big((\Psi(a_{1jt}) - \Psi(a_{1jt}+a_{2jt})) + \sum_{\mathrm{t}=1}^{t-1}(\Psi(a_{2j\mathrm{t}}) - \Psi(a_{1j\mathrm{t}}+a_{2j\mathrm{t}}))\Big)
\end{aligned} \tag{7}
$$

4th term:

$$
\begin{aligned}
&\mathbb{E}_q[\log p(\pi'_{\mathbf{j}}|\alpha_0)] \\
&= \sum_{t=1}^{T} \mathbb{E}_q[\log p(\pi'_{jt}|\alpha_0)] \\
&= \sum_{t=1}^{T} \mathbb{E}_q[\log \frac{\Gamma(\alpha_0+1)}{\Gamma(\alpha_0)\Gamma(1)} {\pi'_{jt}}^{1-1}(1-\pi'_{jt})^{\alpha_0-1}] \\
&= \sum_{t=1}^{T} \mathbb{E}_q[\log \alpha_0 (1-\pi'_{jt})^{\alpha_0-1}] \\
&= \sum_{t=1}^{T} [\log \alpha_0 + (\alpha_0 - 1)\mathbb{E}_q[\log(1-\pi'_{jt})]] \\
&= \sum_{t=1}^{T} [\log \alpha_0 + (\alpha_0 - 1)(\Psi(a_{2jt}) - \Psi(a_{1jt}+a_{2jt}))]
\end{aligned} \tag{8}
$$

5th term is described in the paper.

6th term:

$$
\begin{aligned}
&\mathbb{E}_q[\log p(\phi|\eta)] \\
&= \sum_{k=1}^{K} \mathbb{E}_q[\log p(\phi_k|\eta)] \\
&= \sum_{k=1}^{K} \mathbb{E}_q[\log \frac{\Gamma(V\eta)}{\prod_{i=1}^{V}\Gamma(\eta)} \prod_{i=1}^{V} \phi_{kw}^{\eta-1}] \\
&= \sum_{k=1}^{K} \Big(\log\Gamma(V\eta) - \sum_{i=1}^{V} \log\Gamma(\eta) + \sum_{i=1}^{V} (\eta-1)\mathbb{E}_q[\log \phi_{kw}] \Big) \\
&= \sum_{k=1}^{K} \Big(\log\Gamma(V\eta) - \sum_{i=1}^{V} \log\Gamma(\eta) + \sum_{i=1}^{V} (\eta-1)\big(\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V} \lambda_{kp})\big) \Big)
\end{aligned} \tag{9}
$$

7th term follows the same distribution as the 4th term:

$$
\mathbb{E}_q[\log \ p(\beta'|\gamma)] = \sum_{k=1}^{K} \Big(\log\gamma + (\gamma-1)(\Psi(v_{2k}) - \Psi(v_{1k}+v_{2k})) \Big) \tag{10}
$$

The rest follows from standard entropy computation. We just state the result below:

$$
\begin{aligned}
-\mathbb{E}_q[\log \ q(\pi_j)] = \sum_{t=1}^{T} \Big(&\log \frac{\Gamma(a_{1jt})\Gamma(a_{2jt})}{\Gamma(a_{1jt}+a_{2jt})} - (a_{1jt}-1)\Psi(a_{1jt}) \\
&- (a_{2jt}-1)\Psi(a_{2jt}) + (a_{1jt}+a_{2jt}-2)\Psi(a_{1jt}+a_{2jt}) \Big)
\end{aligned} \tag{11}
$$

$$
-\mathbb{E}_q[\log \ q(c_j)] = -\sum_{t=1}^{T} \rho_{jt} \log \rho_{jt} \tag{12}
$$

$$-\mathbb{E}_q[\log\ q(z_j)] = -\sum_{n=1}^{N} \zeta_{jn} \log \zeta_{jn} \tag{13}$$

$$\begin{aligned}-\mathbb{E}_q[\log\ q(\beta')] = \sum_{k=1}^{K} \{\log \frac{\Gamma(v_{1k})\Gamma(v_{2k})}{\Gamma(v_{1k}+v_{2k})} - (v_{1k}-1)\Psi(v_{1k}) - (v_{2k}-1)\Psi(v_{2k}) \\ + (v_{1k}+v_{2k}-2)\Psi(v_{1k}+v_{2k})\}\end{aligned} \tag{14}$$

$$-\mathbb{E}_q[log\ q(\phi)] = \sum_{k=1}^{K} \big(-\log\Gamma(\sum_{p=1}^{V}\lambda_{kp}) + \sum_{i=1}^{V}\log\Gamma(\lambda_{ki}) - \sum_{i=1}^{V}(\lambda_{ki}-1)(\Psi(\lambda_{ki}) - \Psi(\sum_{p=1}^{V}\lambda_{kp}))\big) \tag{15}$$

Next, we will derive the update equations not described in the paper, for a_{1jt}, a_{2jt}, v_{1k}, v_{2k}, λ. The update for a_{1jt}, a_{2jt}, v_{1k}, v_{2k} are quite similar, we will just describe a_{1jt} and a_{2jt}as an example.

The part of the lower bound that changes with a_{1jt} and a_{2jt} is:

$$\begin{aligned}&\mathscr{L}_{[a_{1jt},a_{2jt}]} \\ &= \sum_{n=1}^{N}\Bigg([\sum_{i=t+1}^{T}\zeta_{jni}(\Psi(a_{2jt}) - \Psi(a_{1jt}+a_{2jt}))] + \zeta_{jnt}(\Psi(a_{1jt}) - \Psi(a_{1jt}+a_{2jt}))\Bigg) \\ &+ (\alpha_0 - 1)(\Psi(a_{2jt}) - \Psi(a_{1jt}+a_{2jt})) \\ &+ \log\frac{\Gamma(a_{1jt})\Gamma(a_{2jt})}{\Gamma(a_{1jt}+a_{2jt})} - (a_{1jt}-1)\Psi(a_{1jt}) - (a_{2jt}-1)\Psi(a_{2jt}) \\ &+ (a_{1jt}-1)\Psi(a_{1jt}+a_{2jt}) + (a_{2jt}-1)\Psi(a_{1jt}+a_{2jt}) \\ &= \Big(\sum_{n=1}^{N}\zeta_{jnt} - a_{1jt} + 1\Big)\Psi(a_{1jt}) \\ &+ \Big(\sum_{n=1}^{N}\sum_{i=t+1}^{T}\zeta_{jni} + \alpha_0 - a_{2jt}\Big)\Psi(a_{1jt}) \\ &- \Big(\sum_{n=1}^{N}\sum_{i=t+1}^{T}\zeta_{jni} + \sum_{n=1}^{N}\zeta_{jnt} + \alpha_0 - a_{1jt} - a_{2jt} + 1\Big)\Psi(a_{1jt}+a_{2jt}) \\ &+ \log\Gamma(a_{1jt}) + \log\Gamma(a_{2jt}) - \log\Gamma(a_{1jt}+a_{2jt})\end{aligned} \tag{16}$$

Compute the derivative respect to a_{1jt} and a_{2jt}:

$$\begin{aligned}&\frac{\partial\mathscr{L}_{[a_{1jt}]}}{\partial a_{1jt}} \\ =&(\sum_{n=1}^{N}\zeta_{jnt} - a_{1jt} + 1)\Psi'(a_{1jt}) \\ &-(\sum_{n=1}^{N}\sum_{i=t+1}^{T}\zeta_{jni} + \sum_{n=1}^{N}\zeta_{jnt} + \alpha_0 - a_{1jt} - a_{2jt} + 1)\Psi'(a_{1jt}+a_{2jt})\end{aligned} \tag{17}$$

$$
\begin{aligned}
&\frac{\partial \mathscr{L}_{[a_{2jt}]}}{\partial a_{1jt}} \\
=&(\sum_{n=1}^{N}\sum_{i=t+1}^{T}\zeta_{jni}+\alpha_0-a_{2jt})\Psi'(a_{2jt}) \\
&-(\sum_{n=1}^{N}\sum_{i=t+1}^{T}\zeta_{jni}+\sum_{n=1}^{N}\zeta_{jnt}+\alpha_0-a_{1jt}-a_{2jt}+1)\Psi'(a_{1jt}+a_{2jt})
\end{aligned} \tag{18}
$$

Set both derivative to 0, we get:

$$
a_{1jt}=1+\sum_{n}^{N}\zeta_{jnt} \tag{19}
$$

$$
a_{2jt}=\alpha_0+\sum_{n=1}^{N}\sum_{i=t+1}^{T}\zeta_{jni} \tag{20}
$$

The part of the lower bound that changes with λ is:

$$
\begin{aligned}
&\mathscr{L}_{\lambda} \\
&=\sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{i=1}^{V}\sum_{t=1}^{T}\sum_{k=1}^{K}(\Psi(\lambda_{ki})-\Psi(\sum_{p}\lambda_{kp}))\zeta_{jnt}\rho_{jtk}[w_{jn}=i]+\sum_{k=1}^{K}\big(\sum_{i=1}^{V}(\eta-1)(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}^{V}\lambda_{kp}))\big) \\
&+\sum_{k=1}^{K}\big(-\log\Gamma(\sum_{p=1}^{V}\lambda_{kp})+\sum_{i=1}^{V}\log\Gamma(\lambda_{ki})-\sum_{i=1}^{V}(\lambda_{ki}-1)(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}^{V}\lambda_{kp}))\big) \\
&=\sum_{k=1}^{K}\Big(\sum_{i=1}^{V}\{\sum_{j=1}^{M}\sum_{n=1}^{N}\sum_{t=1}^{T}(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}^{V}\lambda_{kp}))\zeta_{jnt}\rho_{jtk}[w_{jn}=i]+(\eta-1)(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}^{V}\lambda_{kp})) \\
&+\log\Gamma(\lambda_{ki})-(\lambda_{ki}-1)(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}^{V}\lambda_{kp}))\}-\log\Gamma(\sum_{p=1}^{V}\lambda_{kp})\Big) \\
&=\sum_{k=1}^{K}\Big(\sum_{i=1}^{V}\{(\Psi(\lambda_{ki})-\Psi(\sum_{p}\lambda_{kp}))\sum_{j=1}^{M}\sum_{t=1}^{T}\rho_{jtk}(\sum_{n=1}^{N}\zeta_{jnt}[w_{jn}=i])+(\eta-1)(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}^{V}\lambda_{kp})) \\
&+\log\Gamma(\lambda_{ki})-(\lambda_{ki}-1)(\Psi(\lambda_{ki})-\Psi(\sum_{p=1}^{V}\lambda_{kp}))\}-\log\Gamma(\sum_{p=1}^{V}\lambda_{kp})\Big) \\
&=\sum_{k=1}^{K}\Big(\sum_{i=1}^{V}\{(\sum_{j=1}^{M}\sum_{t=1}^{T}\rho_{jtk}(\sum_{n=1}^{N}\zeta_{jnt}[w_{jn}=i])+\eta-\lambda_{ki})\Psi(\lambda_{ki}) \\
&-(\sum_{j=1}^{M}\sum_{t=1}^{T}\rho_{jtk}(\sum_{n=1}^{N}\zeta_{jnt}[w_{jn}=i])+\eta-\lambda_{ki})\Psi(\sum_{p=1}^{V}\lambda_{kp})+\log\Gamma(\lambda_{ki})\}-\log\Gamma(\sum_{p=1}^{V}\lambda_{kp})\Big).
\end{aligned} \tag{21}
$$

$$
\begin{aligned}
&\frac{\partial \mathscr{L}}{\partial \lambda_{ki}} \\
&= \Big(\sum_{j=1}^{M}\sum_{t=1}^{T}\rho_{jtk}(\sum_{n=1}^{N}\zeta_{jnt}[w_{jn}=i]) + \eta - \lambda_{ki}\Big)\Psi'(\lambda_{ki}) - \Psi(\lambda_{ki}) \\
&- \Big(\sum_{p=1}^{V}(\sum_{j=1}^{M}\sum_{t=1}^{T}\rho_{jtk}(\sum_{n=1}^{N}\zeta_{jnt}[w_{jn}=i]) + \eta - \lambda_{ki})\Big)\Psi'(\sum_{i=1}^{V}\lambda_{ki}) + \Psi(\sum_{p=1}^{V}\lambda_{kp}) - \Psi(\sum_{p=1}^{V}\lambda_{kp}) + \Psi(\lambda_{ki}) \\
&= \Big(\sum_{j=1}^{M}\sum_{t=1}^{T}\rho_{jtk}(\sum_{n=1}^{N}\zeta_{jnt}[w_{jn}=i]) + \eta - \lambda_{ki}\Big)\Psi'(\lambda_{ki}) \\
&- \Big(\sum_{j=1}^{M}\sum_{t=1}^{T}\rho_{jtk}(\sum_{n=1}^{N}\zeta_{jnt}[w_{jn}=i]) + \eta - \lambda_{ki}\Big)\Psi'(\sum_{p=1}^{V}\lambda_{kp})
\end{aligned}
\tag{22}
$$

Set the derivative to 0, we get the update equation for λ:

$$
\lambda_{ki} = \sum_{j=1}^{M}\sum_{t=1}^{T}\rho_{jtk}(\sum_{n=1}^{N}\zeta_{jnt}[w_{jn}=i]) + \eta \tag{23}
$$