B Global update for $q(u)$

Here, we derive the results needed to perform numerical optimization of $\hat{\rho}$ and $\hat{\omega}$, the variational beta parameters of the top-level stick-breaking weights $q(u_k) = \text{Beta}(\hat{\rho}_k\hat{\omega}_k, (1-\hat{\rho}_k)\hat{\omega}_k)$. We only show the results with the sticky hyperparameter; optimization for the case of $\kappa = 0$ and for the HDP-aMMSB is identical to the case for the HDP topic model presented in [11]. For all equations from that paper, you need only substitute in $K+1$ or N, respectively, for the number of documents D.

To begin our derivation, we collect terms in $\mathcal{L}_{HDP}$ that depend on $\hat{\rho}$ and $\hat{\omega}$. Note that we have dropped any additive terms constant with respect to $\hat{\rho}, \hat{\omega}$ in this expression, since they have no bearing on our numerical optimization problem.

$$\begin{aligned}
\mathcal{L}_{obj}(\hat{\rho}, \hat{\omega}) = \sum_{k=1}^{K} \Big(& - c_B(\hat{\rho}_k\hat{\omega}_k, (1-\hat{\rho}_k)\hat{\omega}_k) \\
& + \big(K + 1 - \hat{\rho}_k\hat{\omega}_k\big)\big(\psi(\hat{\rho}_k\hat{\omega}_k) - \psi(\hat{\omega}_k)\big) \\
& + \big(K(K+1-k) + 1 + \gamma - (1-\hat{\rho}_k)\hat{\omega}_k\big)\big(\psi((1-\hat{\rho}_k)\hat{\omega}_k) - \psi(\hat{\omega}_k)\big)\Big) \\
+ \sum_{\ell=1}^{K} \mathbb{E}_q[\beta_\ell] & \Big(\log(\alpha+\kappa) - \log\kappa + \sum_{k=0}^{K} \alpha_k P_{k\ell}(\hat{\theta})\Big) \\
+ \mathbb{E}_q[\beta_{K+1}] & \Big(\sum_{k=0}^{K} \alpha_k P_{k,K+1}(\hat{\theta})\Big)
\end{aligned} \tag{44}$$

B.1 Constrained Optimization Problem

Our goal is to find the $\hat{\rho}, \hat{\omega}$ that maximize $\mathcal{L}_{obj}$. Remember that $\hat{\rho}, \hat{\omega}$ parameterize K Beta distributions, and so have certain positivity constraints. Thus, we need to solve a *constrained* optimization problem:

$$\begin{aligned}
\operatorname{argmax}_{\hat{\rho},\hat{\omega}} \quad & \mathcal{L}_{obj}(\hat{\rho}, \hat{\omega}) \\
& \text{subject to } \ 0 < \hat{\rho}_k < 1 \quad \hat{\omega}_k > 0, \quad k = 1, \ldots, K
\end{aligned} \tag{45}$$

We now give the expressions for the gradient of the objective $\nabla\mathcal{L}_{obj}$, with respect to each entry of $\hat{\omega}$ and $\hat{\rho}$.

Gradient for $\hat{\omega}$ Taking the derivative of Eq. 44 with respect to each entry $\hat{\omega}_m$ of $\hat{\omega}$, for $m \in 1, 2, \ldots K$, is:

$$\begin{aligned}
\frac{\partial \mathcal{L}_{obj}}{\partial \hat{\omega}_m} = & \big(K + 1 - \hat{\rho}_m\hat{\omega}_m\big)\big(\hat{\rho}_m\psi_1(\hat{\rho}_m\hat{\omega}_m) - \psi_1(\hat{\omega}_m)\big) \\
& + \big(K(K+1-m) + 1 + \gamma - (1-\hat{\rho}_m)\hat{\omega}_m\big)\big((1-\hat{\rho}_m)\psi_1((1-\hat{\rho}_m)\hat{\omega}_m) - \psi_1(\hat{\omega}_m)\big)
\end{aligned} \tag{46}$$

where $\psi_1 \triangleq \frac{d^2}{dx^2}\log\Gamma(x)$ is the trigamma function.

Gradient for $\hat{\rho}$ We first define Δ as a $K \times K+1$ matrix of partial derivatives of $\mathbb{E}_q[\beta_k]$ with respect to $\hat{\rho}$:

$$\Delta_{mk} \triangleq \frac{\partial}{\partial \hat{\rho}_m}\mathbb{E}_q[\beta_k] = \begin{cases} -\frac{1}{1-\hat{\rho}_m}\mathbb{E}_q[\beta_k] & m < k \\ \frac{1}{\hat{\rho}_m}\mathbb{E}_q[\beta_k] & m = k \\ 0 & m > k \end{cases} \tag{47}$$