

Note the similarity of this expression with part of the updates in Equation 12. By caching the necessary statistics needed to update  $\theta$ , we can calculate our ELBO in an efficient manner.

### A.3 Updates for the global stick-breaking weights $\beta$

The global stick breaking weights  $\beta$  is not conjugate to node membership weights  $\pi$ . In order to obtain point estimates for  $\beta$  we perform a two-metric constrained optimization using its first order gradients. We can write the objective for  $\beta$  w.r.t to our ELBO in the following manner:

$$\mathcal{L}(\beta) = \sum_{k=1}^K (\gamma - 1) \log(1 - v_k) - N \sum_{k=1}^K \log \Gamma(\alpha \beta_k) + \sum_{k=1}^K (\alpha \beta_k - 1) \sum_{i=1}^N \mathbb{E}_q[\log \pi_{ik}] \quad (9)$$

Since  $\beta_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$  and  $\beta_{K+} = 1 - \sum_{k=1}^K \beta_k$ , we redefine the prior over  $\beta$  as a sum over independently distributed beta variables  $v$ . We can obtain point estimates for  $v$  without having to worry about the constrained optimization task for  $\beta$  which is significantly more costly than the two-metric constrained optimization over  $v$ . We now take the derivatives for  $\beta$  with this in mind:

$$\frac{d\mathcal{L}(\beta)}{dv_m} = \frac{-(\gamma-1)}{(1-v_m)} - \alpha N \sum_{k=1}^K \frac{d\beta_k}{dv_m} \psi(\alpha \beta_k) + \alpha \sum_{k=1}^K \frac{d\beta_k}{dv_m} \sum_{i=1}^N \mathbb{E}_q[\log \pi_{ik}] \quad (10)$$

where the derivative  $\frac{d\beta_k}{dv_m}$  will change depending on the value of  $k$ . When  $m > k$ , then  $\frac{d\beta_k}{dv_m} = 0$ . When  $m = k$ , then  $\frac{d\beta_k}{dv_m} = \frac{\beta_k}{v_m}$ . Finally, when  $m < k$ , then  $\frac{d\beta_k}{dv_m} = \frac{-\beta_k}{(1-v_m)}$ .

Our constrained optimization provides us with updates for  $v^*$  at iteration  $t$  which we can then use in our stochastic variational approach by setting  $v_k^t = (1 - \rho_t)v_k^{t-1} + \rho_t(v_k^*)$ . From this we can determine a new set of values for  $\beta^t$  by setting  $\beta_k^t = v_k^t \prod_{\ell=1}^{k-1} (1 - v_\ell^t)$ .