Forward-backward algorithm for the proposed infinite HMM

A variant of the beam sampling algorithm for infinite HMM (Van Gael et al. 2008) is employed to improve the convergence over standard Gibbs sampling. Specifically, we introduce auxiliary variables \( u_t \) for \( t = 0, \ldots, T - 1 \):

\[
\begin{align*}
  u_0 | S_0 = (k, j) & \sim \text{Uniform}(0, \nu_{jk} \eta_j) \\
  u_t | S_t = (k, j), S_{t-1} = (k', j') & \sim \text{Uniform}(0, q_{it}) \quad \text{for } t = 1, \ldots, T - 1
\end{align*}
\]

where

\[
q_{it} = e^{-G_t^i d_t} e^{-g_t^i d_t} I(k = k') I(j = j') + e^{-G_t^i d_t} (1 - e^{-g_t^i d_t}) I(j = j') \pi_{k'k}^t + (1 - e^{-G_t^i d_t}) \nu_{jk} \eta_j
\]

For notational convenience, we omit the notation \( i \). Let the forward probabilities be \( \alpha_t(k, j) = P(S_t = (k, j) | H_{0:t}, u_{0:t}) \). Then

\[
\begin{align*}
  \alpha_0(k, j) & \propto P(S_0 = (k, j), H_0, u_0) \propto P(S_0 = (k, j)) P(u_0 | S_0 = (k, j)) P(H_0 | C_0 = k) \\
  & = I(u_0 < \nu_{jk} \eta_{Z_0}) P(H_0 | C_0 = k) \\
  \alpha_t(k, j) & \propto \sum_{k', j'} P(S_t = (k, j), S_{t-1} = (k', j'), H_t, u_t | H_{0:t-1}, u_{0:t-1}) \\
  & \propto P(H_t | C_t = k) \sum_{k', j'} P(u_t | S_t = (k, j), S_{t-1} = (k', j')) P(S_t = (k, j) | S_{t-1} = (k', j')) \alpha_{t-1}(k', j') \\
  & \propto P(H_t | C_t = k) \sum_{j' = 0}^{\infty} \sum_{k' = 0}^{\infty} I(u_t < P(S_t = (k, j) | S_{t-1} = (k', j'))) \alpha_{t-1}(k', j') \tag{A1}
\end{align*}
\]

Given \( u_0, \ldots, u_{T-1} \), the number of states \( k \) such that \( \alpha_t(k, j) > 0 \) for \( t = 0, \ldots, T - 1 \) is finite: for \( t = 0 \), the number of \( k \) such that \( \nu_{jk} > u_0 \) is finite for any \( j \) since \( \sum_k \nu_{jk} = 1 \) with \( \nu_{jk} \geq 0 \), and recursively, we can see the number of \( k \) with \( \alpha_t(k, j) > 0 \) is finite. Therefore, the infinite sum over the previous states in the calculation of forward probability reduces to
a finite sum.

$C_{T-1}$ and $Z_{T-1}$ can be sampled from $\alpha_{T-1}(k, j)$. Then for $t = T - 2, \ldots, 0$, we sample $C_t$ and $Z_t$ using

$$P(C_t, Z_t \mid H_{0:T-1}, u_{0:T-1}, C_{t+1}, Z_{t+1}) \propto P(C_{t+1}, Z_{t+1} \mid C_t, Z_t)\alpha_t(C_t, Z_t)P(u_{t+1} \mid S_t, S_{t+1})$$

If we reduce the model to the training phase, we can treat the variable $Z$ as observed. Therefore, the forward probabilities are written as follows:

$$\alpha_0(k) \propto P(C_0 = k, H_0, u_0) \propto P(C_0 = k)P(u_0 \mid C_0 = k)P(H_0 \mid C_0 = k)$$

$$= I(u_0 < \nu_{z_0}k\eta_{t})P(H_0 \mid C_0 = k)$$

$$\alpha_t(k) \propto \sum_{k'} P(C_t = k, C_{t-1} = k', H_t, u_t \mid H_{0:t-1}, u_{0:t-1})$$

$$\propto P(H_t \mid C_t = k)\sum_{k'} P(u_t \mid C_t = k, C_{t-1} = k')P(C_t = k \mid C_{t-1} = k')\alpha_{t-1}(k')$$

$$\propto P(H_t \mid C_t = k)\sum_{k'=0}^{\infty} I(u_t < P(C_t = k \mid C_{t-1} = k'))\alpha_{t-1}(k') \quad (A2)$$

Once we get the trained parameters, we restrict the model to a finite state space, so we don’t need to incorporate the auxiliary variables $u$, so the standard form of forward-backward probabilities can be used.