

Given a sample of the state sequence $z_{1:T}$, we can represent the posterior distribution of the global transition distribution β via a set of auxiliary random variables \bar{m}_{jk} , m_{jk} and w_{jt} , which correspond to the j th restaurant-specific set of table counts associated with the CRF with loyal customers described in the Supplemental Material [Fox et al. (2010)]. The Gibbs sampler iterates between sequential sampling of the state z_t for each individual value of t given β and $z_{\setminus t}$; sampling of the auxiliary variables \bar{m}_{jk} , m_{jk} and w_{jt} given $z_{1:T}$ and β ; and sampling of β given these auxiliary variables.

The direct assignment sampler is initialized by sampling the hyperparameters and β from their respective priors and then sequentially sampling each z_t as if the associated y_t was the last observation. That is, we first sample z_1 given y_1 , β , and the hyperparameters. We then sample z_2 given z_1 , $y_{1:2}$, β , and the hyperparameters, and so on. Based on the resulting sample of $z_{1:T}$, we resample β and the hyperparameters. From then on, the sampler continues with the normal procedure of conditioning on $z_{\setminus t}$ when resampling z_t .

5.2. Blocked sampling of state sequences. The HDP-HMM sequential, direct assignment sampler of Section 5.1 can exhibit slow mixing rates since global state sequence changes are forced to occur coordinate by coordinate. This phenomenon is explored in Scott (2002) for the finite HMM. Although the sticky HDP-HMM reduces the posterior uncertainty caused by fast state-switching explanations of the data, the self-transition bias can cause two continuous and temporally separated sets of observations of a given state to be grouped into two states. See Figure 6(b) for an example. If this occurs, the high probability of self-transition makes it challenging for the sequential sampler to group those two examples into a single state.

We thus propose using a variant of the HMM forward–backward procedure [Rabiner (1989)] to harness the Markovian structure and jointly sample the state sequence $z_{1:T}$ given the observations $y_{1:T}$, transition probabilities π_k , and parameters θ_k . There are two main mechanisms for sampling in an uncollapsed HDP model (i.e., one that instantiates the parameters π_k and θ_k): one is to employ slice sampling while the other is to consider a truncated approximation to the HDP. For the HDP-HMM, a slice sampler, referred to as *beam sampling*, was recently developed [Van Gael et al. (2008)]. This sampler harnesses the efficiencies of the forward–backward algorithm without having to fix a truncation level for the HDP. However, as we elaborate upon in Section 6.1, this sampler suffers from slower mixing rates than the block sampler we propose, which utilizes a fixed-order truncation of the HDP-HMM. Although a fixed truncation reduces our model to a parametric Bayesian HMM, the specific hierarchical prior induced by a truncation of the fully nonparametric HDP significantly improves upon classical parametric Bayesian HMMs. Specifically, a fixed degree L truncation encourages each transition distribution to be sparse over the set of L possible HMM states, and simultaneously encourages transitions from different states to have similar sparsity structures. That is, the truncated HDP prior leads to a *shared* sparse subset of