Material [Fox et al. (2010)]; it roughly follows that of the original HDP-HMM [Teh et al. (2006)]. A key step which simplifies our inference procedure is to note that since we have the deterministic relationships

$$\begin{aligned} \alpha &= (1-\rho)(\alpha+\kappa), \\ \kappa &= \rho(\alpha+\kappa), \end{aligned} \tag{5.7}$$

we can treat ρ and $\alpha+\kappa$ as our hyperparameters and sample these values instead of sampling α and κ directly.

6. Experiments with synthetic data. In this section we explore the performance of the sticky HDP-HMM relative to the original model (i.e., the model with $\kappa = 0$) in a series of experiments with synthetic data. We judge performance according to two metrics: our ability to accurately segment the data according to the underlying state sequence, and the predictive likelihood of held-out data under the inferred model. We additionally assess the improvements in mixing rate achieved by using the blocked sampler of Section 5.2.

6.1. *Gaussian emissions.* We begin our analysis of the sticky HDP-HMM performance by examining a set of simulated data generated from an HMM with Gaussian emissions. The first data set is generated from an HMM with a high probability of self-transition. Here, we aim to show that the original HDP-HMM inadequately captures state persistence. The second data set is from an HMM with a high probability of leaving the current state. In this scenario, our goal is to demonstrate that the sticky HDP-HMM is still able to capture rapid dynamics by inferring a small probability of self-transition.

For all of the experiments with simulated data, we used weakly informative hyperpriors. We placed a Gamma$(1, 0.01)$ prior on the concentration parameters γ and $(\alpha+\kappa)$. The self-transition proportion parameter ρ was given a Beta$(10, 1)$ prior. The parameters of the base measure were set from the data, as will be described for each scenario.

State persistence. The data for the high persistence case were generated from a three-state HMM with a 0.98 probability of self-transition and equal probability of transitions to the other two states. The observation and true state sequences for the state persistence scenario are shown in Figure 6(a). We placed a normal inverse-Wishart prior on the space of mean and variance parameters and set the hyperparameters as follows: 0.01 pseudocounts, mean equal to the empirical mean, three degrees of freedom, and scale matrix equal to 0.75 times the empirical variance. We used this conjugate base measure so that we may directly compare the performance of the blocked and direct assignment samplers. For the blocked sampler, we used a truncation level of $L = 20$.

In Figure 6(d)–(h), we plot the 10th, 50th and 90th quantiles of the Hamming distance between the true and estimated state sequences over the 1000 Gibbs iterations using the direct assignment and blocked samplers on the sticky and original