Supplementary Material for
The Segmented iHMM: A Simple, Efficient Hierarchical Infinite HMM

A. Details for the experiments

A.1. Synthetic data

For the hyperparameters in the feature-independent model, we set $a_0 = 1$ and $b_0 = 1$ for the beta prior; for the feature-based model, we randomly initialize the feature weights from $\mathcal{N}(0, 1)$ and only use the hidden state as a feature. We do a grid search over combinations of values for $\alpha = \{1, 5\}$, $\gamma = \{1, 5\}$, and $K = \{20, 30\}$. We place an $\mathcal{N}(0, 10)$ prior on the mean of the observation distributions. Their variance prior is $\Gamma^{-1}(\alpha_g, \beta_g)$ where α_g and β_g are coming from $\{1, 10\}$ and $\{0.1, 1\}$.

In addition to these hyperparameters, the sub-iHMM also requires the truncation level of the superstates and the substates, which we set to 10 and $\{5, 10\}$, correspondingly. For all hyperparameters, shared with the siHMM, we use the same set of settings for the baselines. For the self-transition bias for sticky HDP-HMM we try $\kappa \in \{1, 10, 100\}$.

A.2. Segmenting user behavior traces

The possible hyperparameter settings that we consider are $K \in \{60, 80\}$, $\alpha \in \{1, 3\}$, $\gamma \in \{1, 3\}$ and finally $\alpha_0 \in \{1, 10\}$, the parameter for the Dirichlet prior, which we place on the parameters of the observation likelihood. We run SVI with 10 different seeds for each setting and report the result for the best setting with the highest VLB.

A.3. Segmenting fruit fly behavior

We use the empirical mean and variance of the dataset as the mean and variance of the Gaussian prior. The parameters of the inverse-Wishart are chosen from the possible combinations of $\nu \in \{47, 50\}$ and $\kappa \in \{1, 0.1, 0.01, 0.001\}$. We set the truncation level for the states to 20. Finally, we have the following possible settings for the transition matrix prior $\alpha \in \{1, 10\}$ and $\gamma \in \{1, 10\}$.

B. Runtime comparison for siHMM and sub-iHMM

We compare our model with sub-iHMM in terms of the evolution of the predictive log-likelihood against time. We compute the predictive log-likelihood over a held-out set from the synthetic dataset. We see in Fig. 1 that SVI for siHMM converges faster than inference in the sub-iHMM.

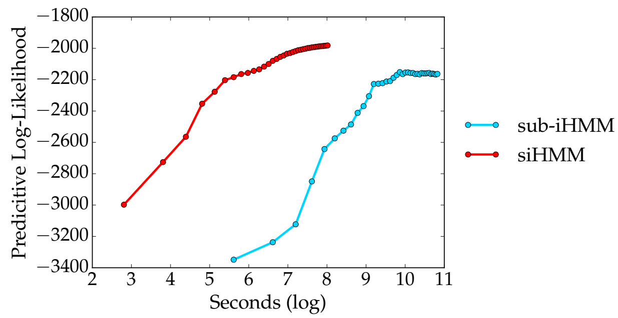

Figure 1. Predictive log-likelihood for the held-out set from the synthetic dataset.