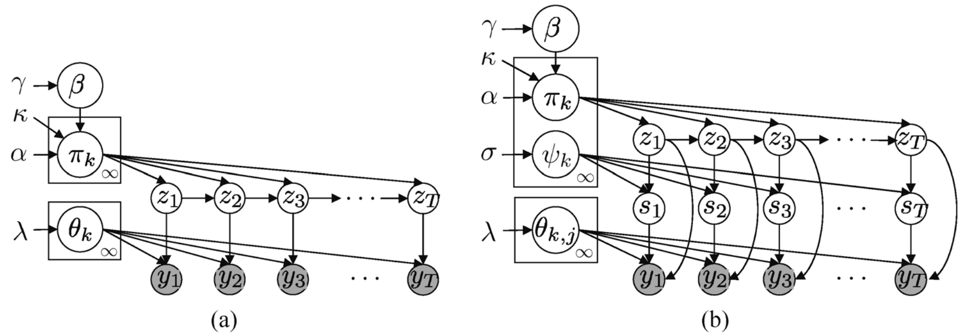

FIG. 5. (a) *Graphical representation of the sticky HDP-HMM. The state evolves as* $z_{t+1}|\{\pi_k\}_{k=1}^{\infty}, z_t \sim \pi_{z_t}$, *where* $\pi_k|\alpha,\kappa,\beta \sim \mathrm{DP}(\alpha+\kappa, (\alpha\beta+\kappa\delta_k)/(\alpha+\kappa))$ *and* $\beta|\gamma \sim \mathrm{GEM}(\gamma)$, *and observations are generated as* $y_t|\{\theta_k\}_{k=1}^{\infty}, z_t \sim F(\theta_{z_t})$. *The original HDP-HMM has* $\kappa = 0$. (b) *Sticky HDP-HMM with DP emissions, where* s_t *indexes the state-specific mixture component generating observation* y_t. *The DP prior dictates that* $s_t|\{\psi_k\}_{k=1}^{\infty}, z_t \sim \psi_{z_t}$ *for* $\psi_k|\sigma \sim \mathrm{GEM}(\sigma)$. *The jth Gaussian component of the kth mixture density is parameterized by* $\theta_{k,j}$ *so* $y_t|\{\theta_{k,j}\}_{k,j=1}^{\infty}, z_t, s_t \sim F(\theta_{z_t,s_t})$.

persistence, the flexible nature of the HDP-HMM prior allows for state sequences with unrealistically fast dynamics to have large posterior probability. For example, with multinomial emissions, a good explanation of the data is to divide different observation values into unique states and then rapidly switch between them (see Figure 1). In such cases, many models with redundant states may have large posterior probability, thus impeding our ability to identify a compact dynamical model which best explains the observations. The problem is compounded by the fact that once this alternating pattern has been instantiated by the sampler, its persistence is then reinforced by the properties of the Chinese restaurant franchise, thus slowing mixing rates. Furthermore, this fragmentation of data into redundant states can reduce predictive performance, as is discussed in Section 6. In many applications, one would like to be able to incorporate prior knowledge that slow, smoothly varying dynamics are more likely.

To address these issues, we propose to instead model the transition distributions π_j as follows:

$$\begin{aligned} &\beta|\gamma \sim \mathrm{GEM}(\gamma), \\ &\pi_j|\alpha,\kappa,\beta \sim \mathrm{DP}\biggl(\alpha+\kappa, \frac{\alpha\beta+\kappa\delta_j}{\alpha+\kappa}\biggr). \end{aligned} \tag{5.1}$$

Here, $(\alpha\beta+\kappa\delta_j)$ indicates that an amount $\kappa > 0$ is added to the jth component of $\alpha\beta$. Informally, what we are doing is increasing the expected probability of self-transition by an amount proportional to κ:

$$E[\pi_{jk}|\beta,\kappa] = \frac{\alpha\beta_k + \kappa\delta(j,k)}{\alpha+\kappa}. \tag{5.2}$$