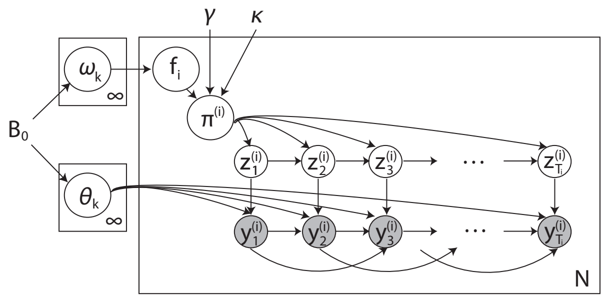

Figure 1: Graphical model of the BP-AR-HMM. The beta process distributed measure $B \mid B_0 \sim \text{BP}(1, B_0)$ is represented by its masses ω_k and locations θ_k, as in Eq. (2). The features are then conditionally independent draws $f_{ik} \mid \omega_k \sim \text{Bernoulli}(\omega_k)$, and are used to define feature-constrained transition distributions $\pi_j^{(i)} \mid \boldsymbol{f}_i, \gamma, \kappa \sim \text{Dir}([\gamma, \ldots, \gamma, \gamma+\kappa, \gamma, \ldots] \otimes \boldsymbol{f}_i)$. The switching VAR dynamics are as in Eq. (6).

these behaviors are indexed by $\{1, \ldots, K_+^{-i}\}$. Given the i^{th} object's observation sequence $\mathbf{y}_{1:T_i}^{(i)}$, transition variables $\boldsymbol{\eta}^{(i)} = \eta^{(i)}_{1:K_+^{-i}, 1:K_+^{-i}}$, and shared dynamic parameters $\theta_{1:K_+^{-i}}$, feature indicators f_{ik} for currently used features $k \in \{1, \ldots, K_+^{-i}\}$ have the following posterior distribution:

$$p(f_{ik} \mid \boldsymbol{F}^{-ik}, \mathbf{y}_{1:T_i}^{(i)}, \boldsymbol{\eta}^{(i)}, \theta_{1:K_+^{-i}}, \alpha) \propto p(f_{ik} \mid \boldsymbol{F}^{-ik}, \alpha) p(\mathbf{y}_{1:T_i}^{(i)} \mid \boldsymbol{f}_i, \boldsymbol{\eta}^{(i)}, \theta_{1:K_+^{-i}}). \tag{10}$$

Here, the IBP prior implies that $p(f_{ik} = 1 \mid \boldsymbol{F}^{-ik}, \alpha) = m_k^{-i}/N$, where m_k^{-i} denotes the number of objects *other* than object i that exhibit behavior k. In evaluating this expression, we have exploited the exchangeability of the IBP [9], which follows directly from the beta process construction [22].

For binary random variables, MH proposals can mix faster [6] and have greater statistical efficiency [14] than standard Gibbs samplers. To update f_{ik} given $\boldsymbol{F}^{-ik}$, we thus use the posterior of Eq. (10) to evaluate a MH proposal which flips f_{ik} to the complement $\bar{f}$ of its current value f:

$$\begin{aligned} f_{ik} &\sim \rho(\bar{f} \mid f)\delta(f_{ik}, \bar{f}) + (1 - \rho(\bar{f} \mid f))\delta(f_{ik}, f) \\ \rho(\bar{f} \mid f) &= \min\left\{ \frac{p(f_{ik} = \bar{f} \mid \boldsymbol{F}^{-ik}, \mathbf{y}_{1:T_i}^{(i)}, \boldsymbol{\eta}^{(i)}, \theta_{1:K_+^{-i}}, \alpha)}{p(f_{ik} = f \mid \boldsymbol{F}^{-ik}, \mathbf{y}_{1:T_i}^{(i)}, \boldsymbol{\eta}^{(i)}, \theta_{1:K_+^{-i}}, \alpha)}, 1 \right\}. \end{aligned} \tag{11}$$

To compute likelihoods, we combine $\boldsymbol{f}_i$ and $\boldsymbol{\eta}^{(i)}$ to construct feature-constrained transition distributions $\pi_j^{(i)}$ as in Eq. (8), and apply the sum-product message passing algorithm [19].

An alternative approach is needed to resample the Poisson(α/N) "unique" features associated only with object i. Let $K_+ = K_+^{-i} + n_i$, where n_i is the number of features unique to object i, and define $\boldsymbol{f}_{-i} = f_{i,1:K_+^{-i}}$ and $\boldsymbol{f}_{+i} = f_{i,K_+^{-i}+1:K_+}$. The posterior distribution over n_i is then given by

$$\begin{aligned} & p(n_i \mid \boldsymbol{f}_i, \mathbf{y}_{1:T_i}^{(i)}, \boldsymbol{\eta}^{(i)}, \theta_{1:K_+^{-i}}, \alpha) \\ & \quad \propto \frac{(\frac{\alpha}{N})^{n_i} e^{-\frac{\alpha}{N}}}{n_i!} \iint p(\mathbf{y}_{1:T_i}^{(i)} \mid \boldsymbol{f}_{-i}, \boldsymbol{f}_{+i} = \mathbf{1}, \boldsymbol{\eta}^{(i)}, \boldsymbol{\eta}_+, \theta_{1:K_+^{-i}}, \boldsymbol{\theta}_+) \, dB_0(\boldsymbol{\theta}_+) dH(\boldsymbol{\eta}_+), \end{aligned} \tag{12}$$

where H is the gamma prior on transition variables, $\boldsymbol{\theta}_+ = \theta_{K_+^{-i}+1:K_+}$ are the parameters of unique features, and $\boldsymbol{\eta}_+$ are transition parameters $\eta_{jk}^{(i)}$ to or from unique features $j, k \in \{K_+^{-i} + 1 : K_+\}$. Exact evaluation of this integral is intractable due to dependencies induced by the AR-HMMs.

One early approach to approximate Gibbs sampling in non-conjugate IBP models relies on a finite truncation [7]. Meeds et al. [15] instead consider independent Metropolis proposals which replace the existing unique features by $n_i' \sim \text{Poisson}(\alpha/N)$ new features, with corresponding parameters $\boldsymbol{\theta}_+'$ drawn from the prior. For high-dimensional models like that considered in this paper, however, moves proposing large numbers of unique features have low acceptance rates. Thus, mixing rates are greatly affected by the beta process hyperparameter α. We instead develop a "birth and death" reversible jump MCMC (RJMCMC) sampler [8], which proposes to either add a single new feature,