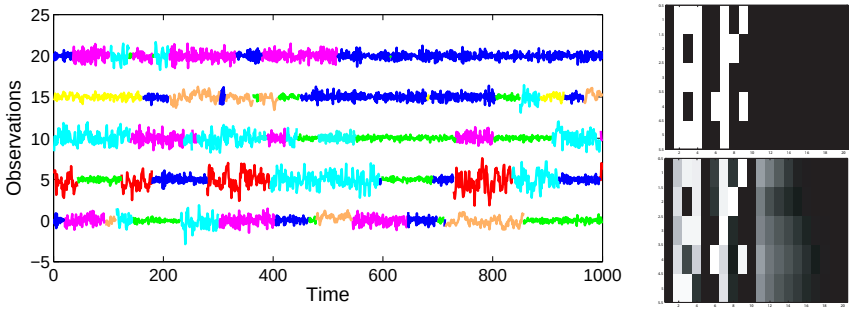

(a) (b)

Figure 2: (a) Observation sequences for each of 5 switching AR(1) time series colored by true mode sequence, and offset for clarity. (b) True feature matrix (top) of the five objects and estimated feature matrix (bottom) averaged over 10,000 MCMC samples taken from 100 trials every 10th sample. White indicates active features. The estimated feature matrices are produced from mode sequences mapped to the ground truth labels according to the minimum Hamming distance metric, and selecting modes with more than 2% of the object's observations.

4.3 Sampling the beta process and Dirichlet transition hyperparameters

We additionally place priors on the Dirichlet hyperparameters γ and κ, as well as the beta process parameter α. Let $\boldsymbol{F} = \{\boldsymbol{f}_i\}$. As derived in [9], $p(\boldsymbol{F} \mid \alpha)$ can be expressed as

$$p(\boldsymbol{F} \mid \alpha) \propto \alpha^{K_+} \exp\left(-\alpha \sum_{n=1}^{N} \frac{1}{n} \right), \tag{20}$$

where, as before, K_+ is the number of unique features activated in $\boldsymbol{F}$. As in [7], we place a conjugate Gamma(a_α, b_α) prior on α, which leads to the following posterior distribution:

$$p(\alpha \mid \boldsymbol{F}, a_\alpha, b_\alpha) \propto p(\boldsymbol{F} \mid \alpha) p(\alpha \mid a_\alpha, b_\alpha) \propto \text{Gamma}\left(a_\alpha + K_+, b_\alpha + \sum_{n=1}^{N} \frac{1}{n} \right). \tag{21}$$

Transition hyperparameters are assigned similar priors $\gamma \sim \text{Gamma}(a_\gamma, b_\gamma)$, $\kappa \sim \text{Gamma}(a_\kappa, b_\kappa)$. Because the generative process of Eq. (7) is non-conjugate, we rely on MH steps which iteratively resample γ given κ, and κ given γ. Each sub-step uses a gamma proposal distribution $q(\cdot \mid \cdot)$ with fixed variance σ_γ^2 or σ_κ^2, and mean equal to the current hyperparameter value. To update γ given κ, the acceptance probability is $\min\{r(\gamma' \mid \gamma), 1\}$, where $r(\gamma' \mid \gamma)$ is defined to equal

$$\frac{p(\gamma' \mid \kappa, \boldsymbol{\pi}, \boldsymbol{F}) q(\gamma \mid \gamma')}{p(\gamma \mid \kappa, \boldsymbol{\pi}, \boldsymbol{F}) q(\gamma' \mid \gamma)} = \frac{p(\boldsymbol{\pi} \mid \gamma', \kappa, \boldsymbol{F}) p(\gamma') q(\gamma \mid \gamma')}{p(\boldsymbol{\pi} \mid \gamma, \kappa, \boldsymbol{F}) p(\gamma) q(\gamma' \mid \gamma)} = \frac{f(\gamma') \Gamma(\vartheta) e^{-\gamma' b_\gamma} \gamma^{\vartheta' - \vartheta - a_\gamma} \sigma_\gamma^{2\vartheta}}{f(\gamma) \Gamma(\vartheta') e^{-\gamma b_\gamma} \gamma'^{\vartheta - \vartheta' - a_\gamma} \sigma_\gamma^{2\vartheta'}}.$$

Here, $\vartheta = \gamma^2/\sigma_\gamma^2$, $\vartheta' = \gamma'^2/\sigma_\gamma^2$, and $f(\gamma) = \prod_i \frac{\Gamma(\gamma K_i + \kappa)^{K_i}}{\Gamma(\gamma)^{K_i^2 - K_i} \Gamma(\gamma + \kappa)^{K_i}} \prod_{(j,k)=1}^{K_i} \pi_{kj}^{(i)^{\gamma + \kappa \delta(k,j) - 1}}$. The MH sub-step for resampling κ given γ is similar, but with an appropriately redefined $f(\kappa)$.

5 Synthetic Experiments

To test the ability of BP-AR-HMM to discover shared dynamics, we generated five time series that switched between AR(1) models

$$y_t^{(i)} = a_{z_t^{(i)}} y_{t-1}^{(i)} + e_t^{(i)}(z_t^{(i)}) \tag{22}$$

with $a_k \in \{-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8\}$ and process noise covariance Σ_k drawn from an IW$(0.5, 3)$ prior. The object-specific features, shown in Fig. 2(b), were sampled from a truncated IBP [9] using $\alpha = 10$ and then used to generate the observation sequences of Fig. 2(a). The resulting feature matrix estimated over 10,000 MCMC samples is shown in Fig. 2. Comparing to the true feature matrix, we see that our model is indeed able to discover most of the underlying latent structure of the time series despite the challenging setting defined by the close AR coefficients.

One might propose, as an alternative to the BP-AR-HMM, using an architecture based on the hierarchical Dirichlet process of [21]; specifically we could use the HDP-AR-HMMs of [5] tied together with a shared set of transition and dynamic parameters. To demonstrate the difference between these models, we generated data for three switching AR(1) processes. The first two objects, with four times the data points of the third, switched between dynamical modes defined