where $\otimes$ denotes the element-wise, or Hadamard, vector product. Let $\mathbf{y}_t^{(i)}$ represent the observation vector of the ith object at time t, and $z_t^{(i)}$ the latent behavior mode. Assuming an order r switching VAR process, the dynamics of the ith object are described by the following generative process:

$$z_t^{(i)} \mid z_{t-1}^{(i)} \sim \pi_{z_{t-1}^{(i)}}^{(i)} \quad (20)$$

$$\mathbf{y}_t^{(i)} = \sum_{j=1}^{r} A_{j, z_t^{(i)}} \mathbf{y}_{t-j}^{(i)} + \mathbf{e}_t^{(i)}(z_t^{(i)}) \triangleq A_{z_t^{(i)}} \tilde{\mathbf{y}}_t^{(i)} + \mathbf{e}_t^{(i)}(z_t^{(i)}), \quad (21)$$

where $\mathbf{e}_t^{(i)}(k) \sim \mathcal{N}(0, \Sigma_k)$, $A_k = [A_{1,k} \ \ \ldots \ \ A_{r,k}]$, and $\tilde{\mathbf{y}}_t^{(i)} = [\mathbf{y}_{t-1}^{(i)T} \ \ \ldots \ \ \mathbf{y}_{t-r}^{(i)T}]^T$. The standard HMM with Gaussian emissions arises as a special case of this model when $A_k = 0$ for all k.

A BAYESIAN NONPARAMETRIC FEATURAL MODEL UTILIZING BETA AND BERNOULLI PROCESSES

Following the theme of sections "Sticky HDP-HMM" and "HDP-AR-HMM and HDP-SLDS," it is often desirable to consider a Bayesian nonparametric featural model that relaxes the assumption that the number of features is known or bounded. Such a featural model seeks to allow for infinitely many features, while encouraging a sparse, finite representation. Just as the DP provides a useful Bayesian nonparametric prior in clustering applications (i.e., when each observation is associated with a single parameter θ_k), it has been shown that a stochastic process known as the beta process is useful in Bayesian nonparametric featural models (i.e., when each observation is associated with a subset of parameters) [30].

JUST AS THE DP PROVIDES A USEFUL BAYESIAN NONPARAMETRIC PRIOR IN CLUSTERING APPLICATIONS, IT HAS BEEN SHOWN THAT A STOCHASTIC PROCESS KNOWN AS THE BETA PROCESS IS USEFUL IN BAYESIAN NONPARAMETRIC FEATURAL MODELS.

The beta process is a special case of a general class of stochastic processes known as completely random measures [32]. A completely random measure G is defined such that for any disjoint sets A_1 and A_2, the corresponding random measures $G(A_1)$ and $G(A_2)$ are independent. This idea generalizes the family of independent increments processes on the real line. All completely random measures (up to a deterministic component) can be constructed from realizations of a nonhomogenous Poisson process [32]. Specifically, a Poisson rate measure η is defined on a product space $\Theta \otimes \mathbb{R}$, and a draw from the specified Poisson process yields a collection of points $\{\theta_j, \omega_j\}$ that can be used to define a completely random measure

$$G = \sum_{j=1}^{\infty} \omega_j \delta_{\theta_j}. \quad (22)$$

This construction assumes η has infinite mass, yielding the countably infinite collection of points from the Poisson process. From (22), we see that completely random measures are discrete. Letting the rate measure be defined as a product of a base measure G_0 and an improper gamma distribution

$$\eta(d\theta, d\omega) = c\omega^{-1} e^{-c\omega} d\omega G_0(d\theta) \quad (23)$$

with $c > 0$, gives rise to completely random measures $G \sim \mathrm{GP}(c, G_0)$, where GP denotes a gamma process. Normalizing G yields draws from a $\mathrm{DP}(\alpha, G_0/\alpha)$, with $\alpha = G_0(\Theta)$. Random probability measures G are necessarily not completely random since the random variables $G(A_1)$ and $G(A_2)$ for disjoint sets A_1 and A_2 are dependent due to the normalization constraint.

Now consider a rate measure defined as the product of a base measure B_0, with total mass $B_0(\Theta) = \alpha$, and an improper beta distribution on the product space $\Theta \otimes [0, 1]$

$$\nu(d\omega, d\theta) = c\omega^{-1}(1-\omega)^{c-1} d\omega B_0(d\theta), \quad (24)$$

where, once again, $c > 0$. The resulting completely random measure is known as the beta process with draws denoted by $B \sim \mathrm{BP}(c, B_0)$. Note that using this construction, the weights ω_k of the atoms in B lie in the interval $(0, 1)$. Since η is σ-finite, Campbell's theorem [33] guarantees that for α finite, B has finite expected measure. The characteristics of this process define desirable traits for a Bayesian nonparametric featural model: we have a countably infinite collection of coin-tossing probabilities (one for each of our infinite number of features), but only a sparse, finite subset are active in any realization.

The beta process is conjugate to a class of Bernoulli processes [30], denoted by $\mathrm{BeP}(B)$, which provide our sought-for featural representation. A realization $X_i \sim \mathrm{BeP}(B)$, with B an atomic measure, is a collection of unit mass atoms on Θ located at some subset of the atoms in B. In particular,

$$f_{ik} \sim \mathrm{Bernoulli}(\omega_k) \quad (25)$$

is sampled independently for each atom θ_k in B, and then $X_i = \sum_k f_{ik} \delta_{\theta_k}$. In many applications, we interpret the atom locations θ_k as a shared set of global features. A Bernoulli process realization X_i then determines the subset of features allocated to object i

$$B \mid B_0, c \sim \mathrm{BP}(c, B_0)$$

$$X_i \mid B \sim \mathrm{BeP}(B), \quad i = 1, \ldots, N. \quad (26)$$

Computationally, Bernoulli process realizations X_i are often summarized by an infinite vector of binary indicator variables $f_i = [f_{i1}, f_{i2}, \ldots]$, where $f_{ik} = 1$ if and only if object i exhibits feature k. Using the beta process measure B to tie together the feature vectors encourages them to share similar features while allowing object-specific variability.

As shown by Thibaux and Jordan [30], marginalizing over the latent beta process measure B, and taking $c = 1$, induces a predictive distribution on feature indicators known as the Indian buffet process (IBP) [31]. The IBP is a culinary