The previous work of Fox et al. [8] considered a related, yet simpler formulation for modeling a maneuvering target as a fixed LDS driven by a switching exogenous input. Since the number of maneuver modes was assumed unknown, the exogenous input was taken to be the emissions of a HDP-HMM. This work can be viewed as an extension of the work by Caron et al. [25] in which the exogenous input was an independent noise process generated from a DP mixture model. The HDP-SLDS of [19] is a departure from these works since the dynamic parameters themselves change with the mode, providing a much more expressive model.

In [19], the utility of the HDP-SLDS and HDP-AR-HMM was demonstrated on two different problems: 1) detecting changes in the volatility of the IBOVESPA stock index and 2) segmenting sequences of honey bee dances. The dynamics underlying both of these data sets appear to be quite complex, yet can be described by repeated returns to simpler dynamical models, and as such have been modeled with Markov switching processes [26], [27]. Without prespecifying domain-specific knowledge, and instead simply relying on a set of observations along with weakly informative hyperprior settings, the HDP-SLDS and HDP-AR-HMM were able to discover the underlying structure of the data with performance competitive with these alternative methods, and consistent with domain expert analysis.

[OUR BAYESIAN NONPARAMETRIC APPROACH ENVISIONS A LARGE LIBRARY OF BEHAVIORS, WITH EACH TIME SERIES OR OBJECT EXHIBITING A SUBSET OF THESE BEHAVIORS.]

MULTIPLE RELATED TIME SERIES

In many applications, one would like to discover and model dynamical behaviors which are shared among several related time series. By jointly modeling such time series, one can improve parameter estimates, especially in the case of limited data, and find interesting structure in the relationships between the time series. Assuming that each of these time series is modeled via a Markov switching process, our Bayesian nonparametric approach envisions a large library of behaviors, with each time series or object exhibiting a subset of these behaviors. We aim to allow flexibility in the number of total and sequence-specific behaviors, while encouraging objects to share similar subsets of the behavior library. Additionally, a key aspect of a flexible model for relating time series is to allow the objects to switch between behaviors in different manners (e.g., even if two people both exhibit running and walking behaviors, they might alternate between these dynamical modes at different frequencies).

One could imagine a Bayesian nonparametric approach based on tying together multiple time series under the HDP prior outlined in the section "Sticky HDP-HMM." However, such a formulation assumes that all time series share the same set of behaviors, and switch among them in exactly the same manner. Alternatively, Fox et al. [28] consider a featural representation, and show the utility of an alternative family of priors based on the beta process [29], [30].

FINITE FEATURE MODELS OF MARKOV SWITCHING PROCESSES

Assume we have a finite collection of behaviors $\{\theta_1, \ldots, \theta_K\}$ that are shared in an unknown manner among N objects. One can represent the set of behaviors each object exhibits via an associated list of features. A standard featural representation for describing the N objects employs an $N \times K$ binary matrix $F = \{f_{ik}\}$. Setting $f_{ik} = 1$ implies that object i exhibits feature k for some $t \in \{1, \ldots, T_i\}$, where T_i is the length of the ith time series. To discover the structure of behavior sharing (i.e., the feature matrix), one takes the feature vector $f_i = [f_{i1}, \ldots, f_{iK}]$ to be random. Assuming each feature is treated independently, this necessitates defining a feature inclusion probability ω_k for each feature k. Within a Bayesian framework, these probabilities are given a prior that is then informed by the data to provide a posterior distribution on feature inclusion probabilities. For example, one could consider the finite Bayesian feature model of [31] that assumes

$$\omega_k \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right)$$
$$f_{ik} \mid \omega_k \sim \text{Bernoulli}(\omega_k). \tag{18}$$

Beta random variables $\omega_k \in (0, 1)$, and can thus be thought of as defining coin-tossing probabilities. The resulting biased coin is then tossed to define whether f_{ik} is 0 or 1 (i.e., the outcome of a Bernoulli trial). Because each feature is generated independently, and a Beta(a, b) random variable has mean $a/(a+b)$, the expected number of active features in an $N \times K$ matrix is $N\alpha/(\alpha/K + 1) < N\alpha$.

A hierarchical Bayesian featural model also requires priors for behavior parameters θ_k, and the process by which each object switches among its selected behaviors. In the case of Markov switching processes, this switching mechanism is governed by the transition distributions of object i, $\pi_j^{(i)}$. As an example of such a model, imagine that each of the N objects is described by a switching VAR process (see the section "Markov Jump Linear Systems") that moves among some subset of K possible dynamical modes. Each of the VAR process parameters $\theta_k = \{A_k, \Sigma_k\}$ describes a unique behavior. The feature vector f_i constrains the transitions of object i to solely be between the selected subset of the K possible VAR processes by forcing $\pi_{jk}^{(i)} = 0$ for all k such that $f_{ik} = 0$. One natural construction places Dirichlet priors on the transition distributions, and some prior H (e.g., an MNIW) on the behavior parameters. Then,

$$\pi_j^{(i)} \mid f_i, \gamma, \kappa \sim \text{Dir}([\gamma, \ldots, \gamma, \gamma + \kappa, \gamma, \ldots, \gamma] \otimes f_i) \qquad \theta_k \sim H, \tag{19}$$