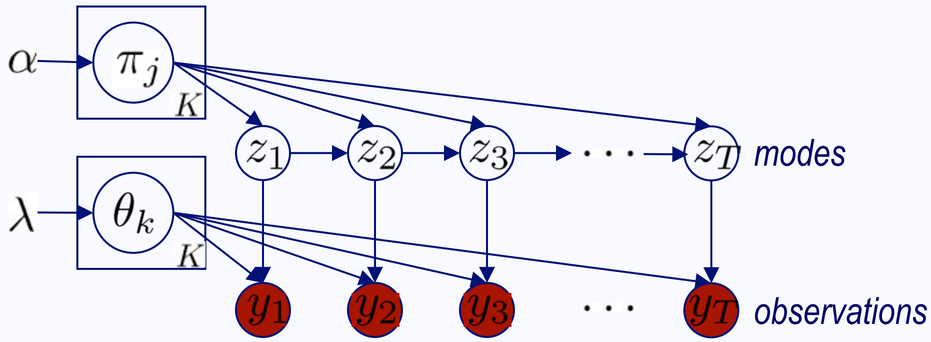
Applied Bayesian Nonparametrics

Special Topics in Machine Learning
Brown University CSCI 2950-P, Fall 2011

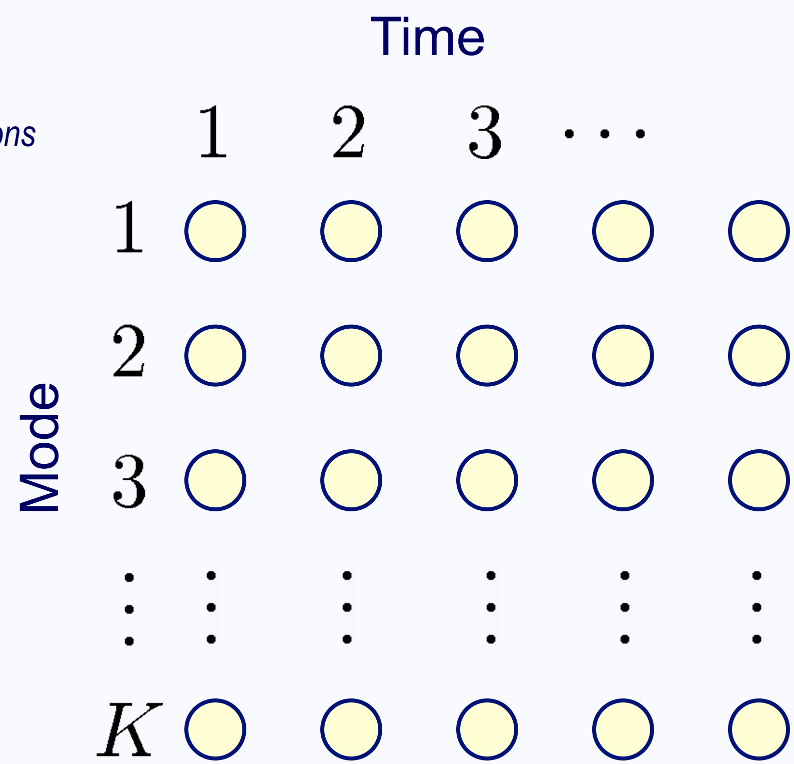
November 29: Beta Process Hidden Markov Models

Hidden Markov Models

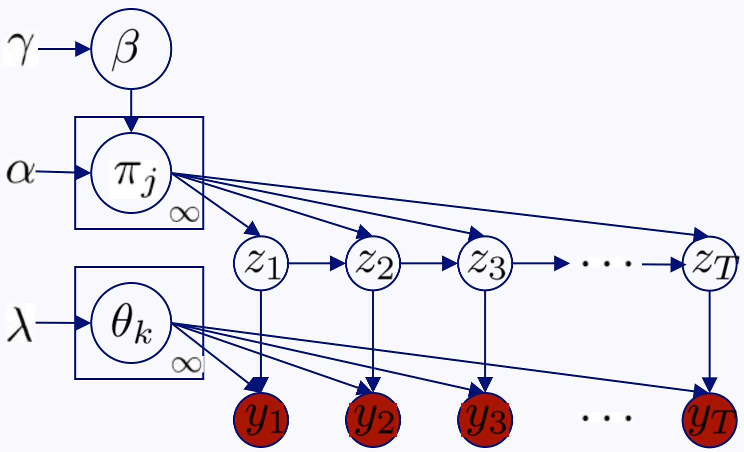


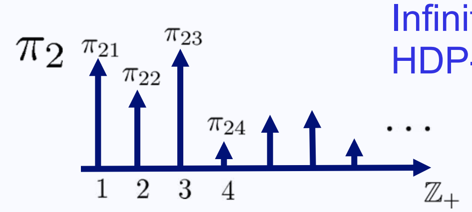
$$z_t \sim \pi_{z_{t-1}}$$
$$y_t \sim F(\theta_{z_t})$$

$$P = \begin{bmatrix} \text{——} \ \pi_1 \ \text{——} \\ \text{——} \ \pi_2 \ \text{——} \\ \vdots \\ \text{——} \ \pi_K \ \text{——} \end{bmatrix}$$

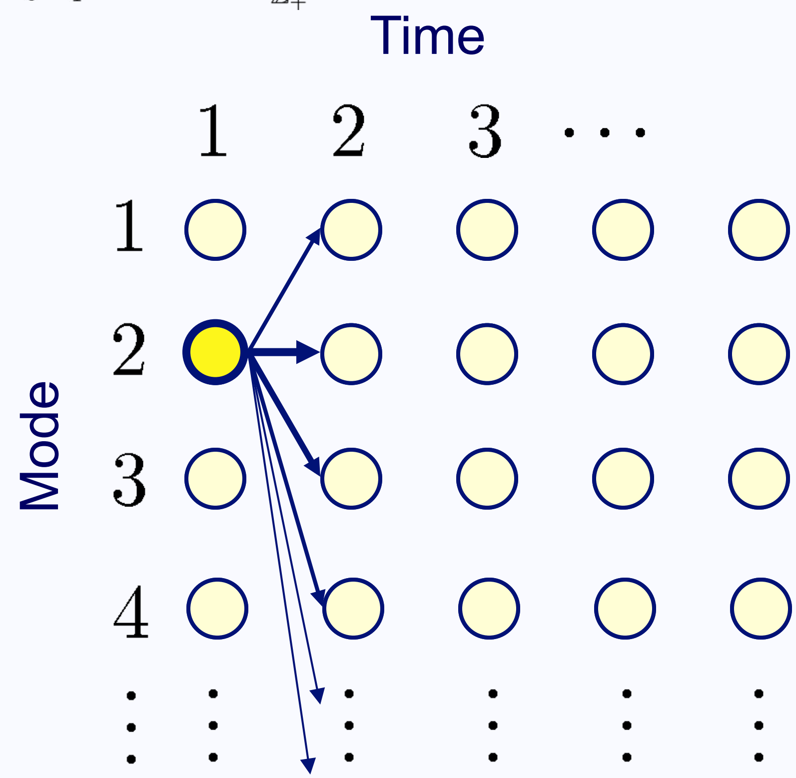


Issue 1: How many modes?





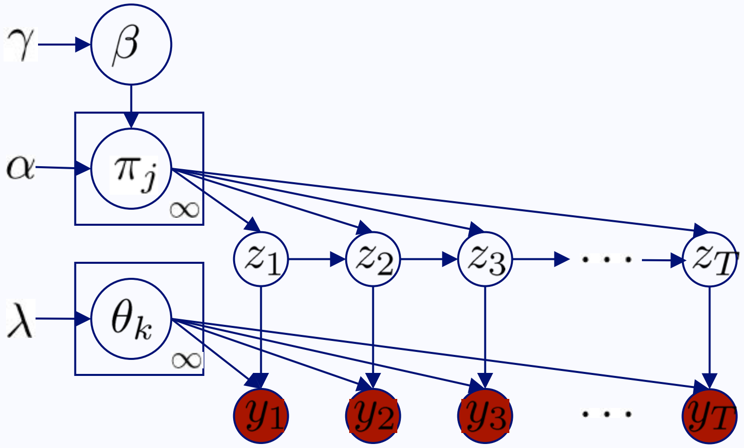
Infinite HMM: Beal, et.al., *NIPS* 2002
HDP-HMM: Teh, et. al., *JASA* 2006

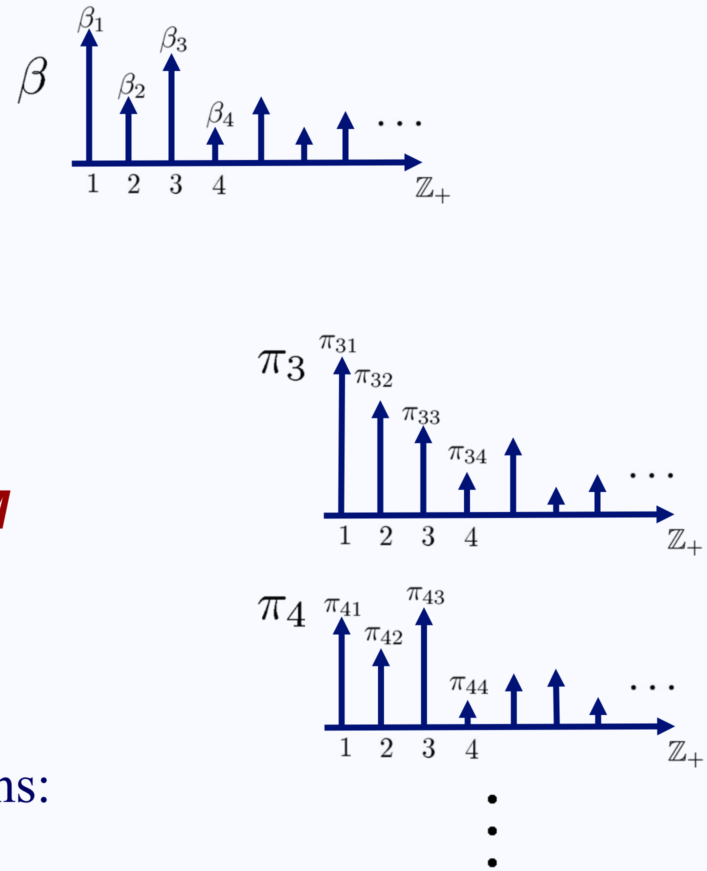


Hierarchical Dirichlet Process HMM

- Dirichlet process (DP):
  - Mode space of unbounded size
  - Model complexity adapts to observations
- Hierarchical:
  - Ties mode transition distributions
  - *Shared* sparsity

HDP-HMM





Hierarchical Dirichlet Process HMM

- Global transition distribution:

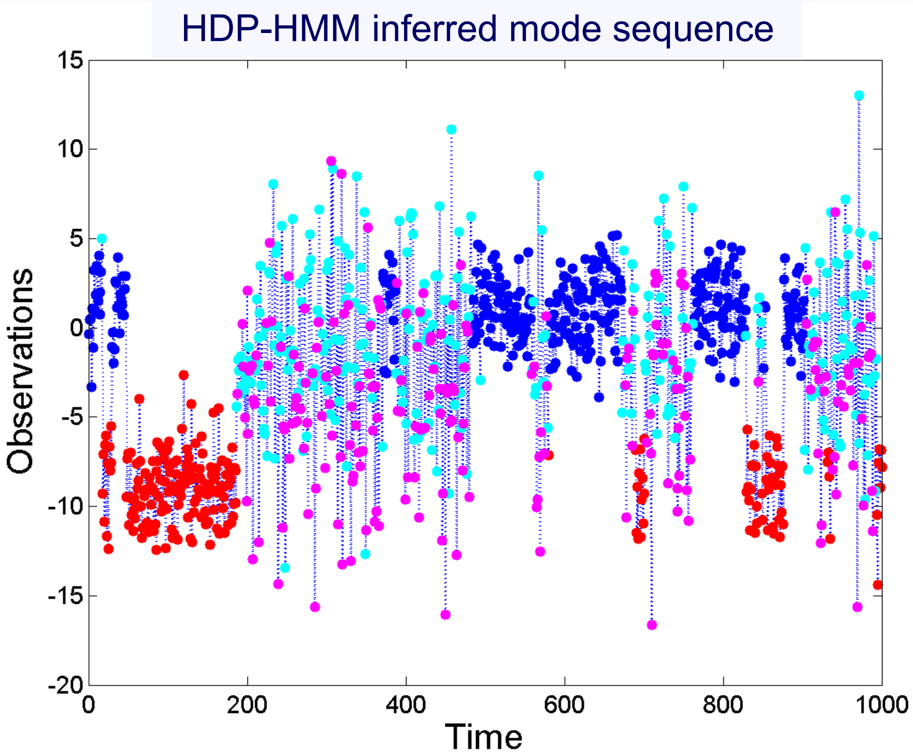
$$\beta \sim \text{Stick}(\gamma)$$

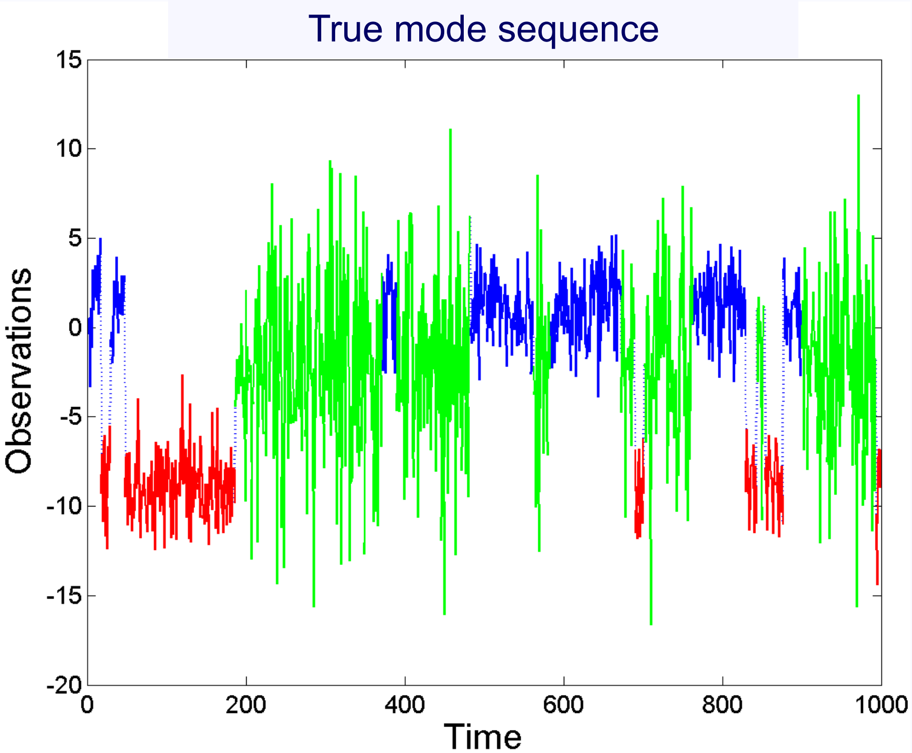
- Mode-specific transition distributions:

$$\pi_j \sim \text{DP}(\alpha\beta) \quad j = 1, 2, 3, \ldots$$

sparsity of β is shared → $E[\pi_{jk}] = \beta_k$

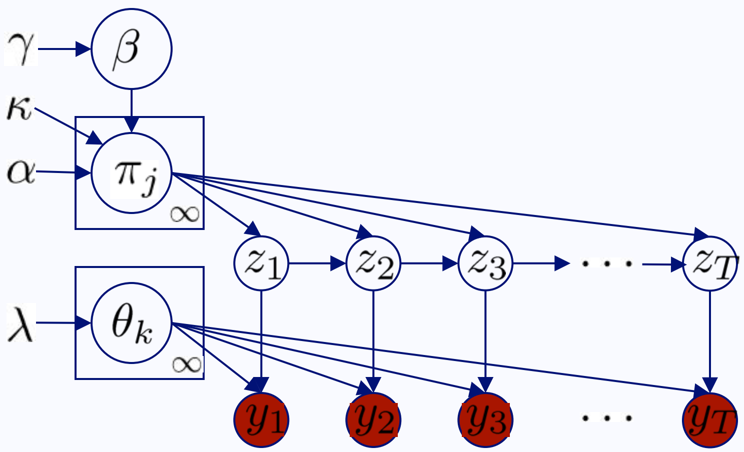
Issue 2: Temporal Persistence

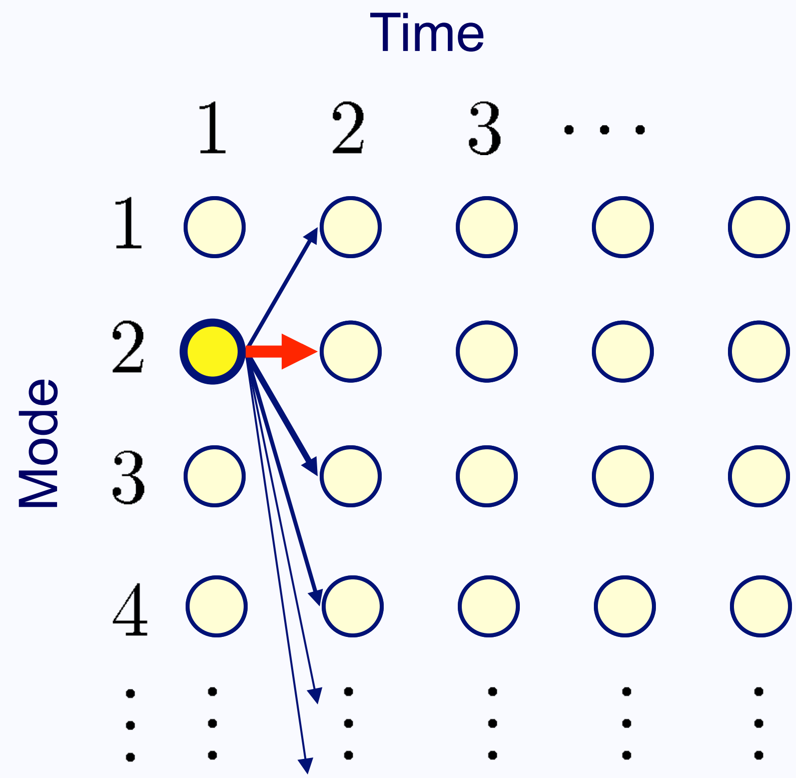




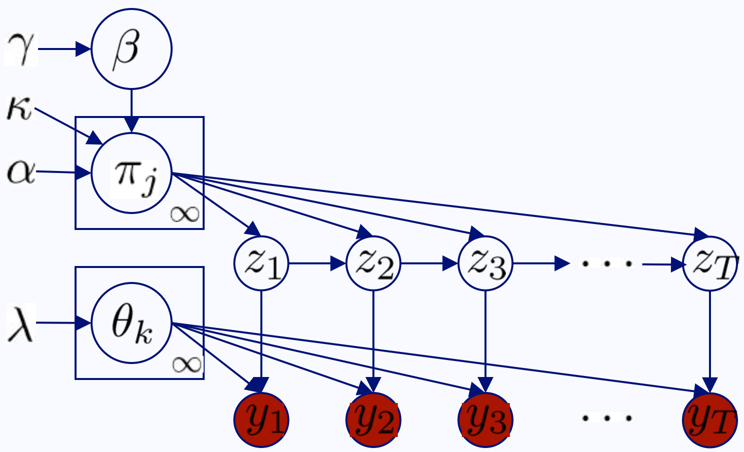
Hidden Markov Model

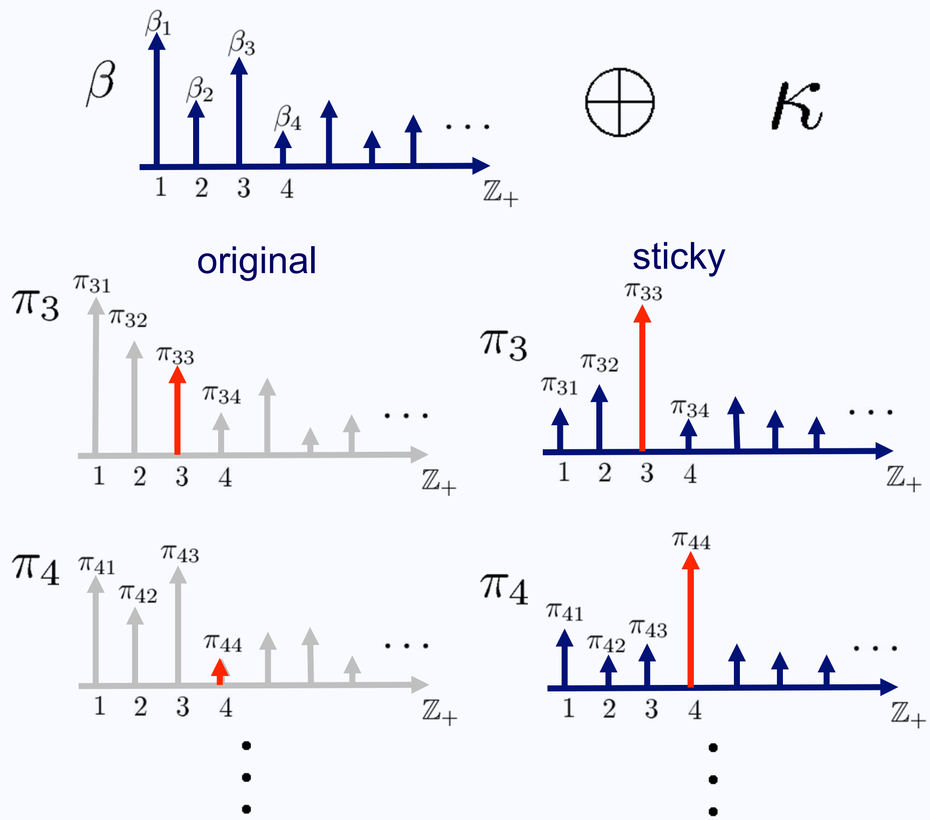
“Sticky” HDP-HMM





"Sticky" HDP-HMM





$$\beta \sim \text{Stick}(\gamma)$$
$$\pi_j \sim \text{DP}\left(\alpha\beta + \kappa\delta_j\right)$$

mode-specific base measure

$$E[\pi_{jk}] = \beta_k$$

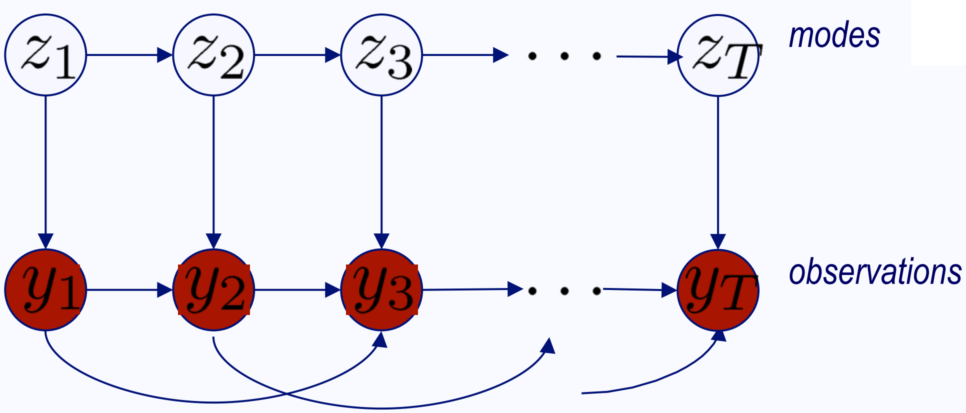
Increased probability of self-transition

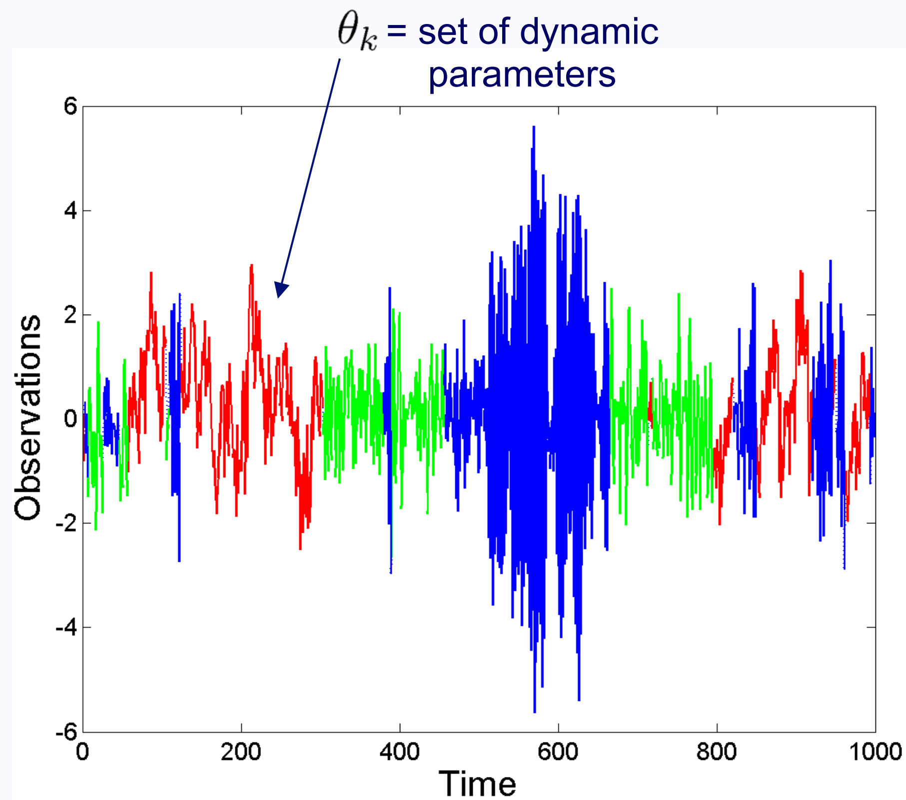
$$E[\pi_{jk}] = \frac{\alpha\beta_k + \kappa\delta(j,k)}{\alpha + \kappa}$$

Infinite HMM: Beal, et.al., *NIPS* 2002

Issue 3: Complex Local Dynamics

- Discrete clusters may not accurately capture high-dimensional data
- Autoregressive HMM: Discrete-mode switching of *smooth* observation dynamics





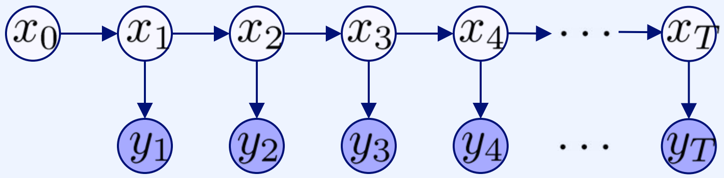
Switching Dynamical Processes

Linear Dynamical Systems

- State space LTI model:

$$x_t = Ax_{t-1} + e_t$$
$$y_t = Cx_t + w_t$$

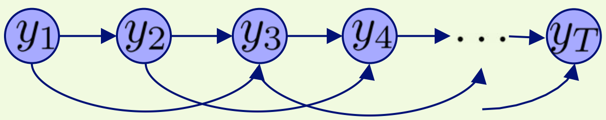
$$e_t \sim \mathcal{N}(0, \Sigma) \quad w_t \sim \mathcal{N}(0, R)$$



- Vector autoregressive (VAR) process:

$$y_t = \sum_{i=1}^{r} A_i y_{t-i} + e_t$$

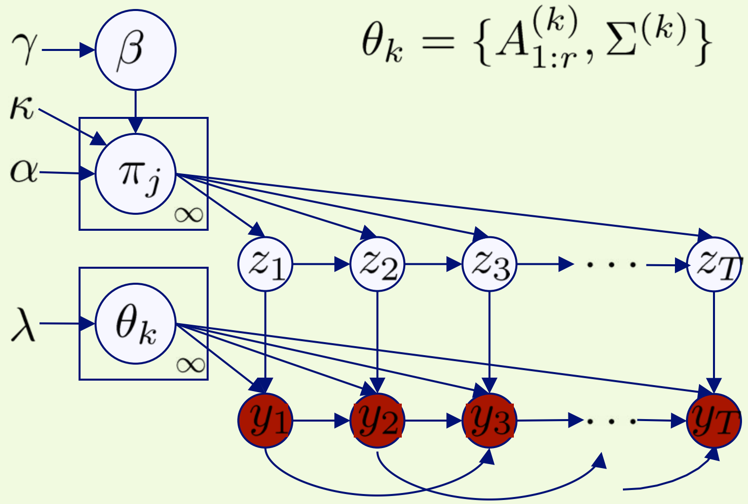
$$e_t \sim \mathcal{N}(0, \Sigma)$$



HDP-AR-HMM and HDP-SLDS

HDP-AR-HMM

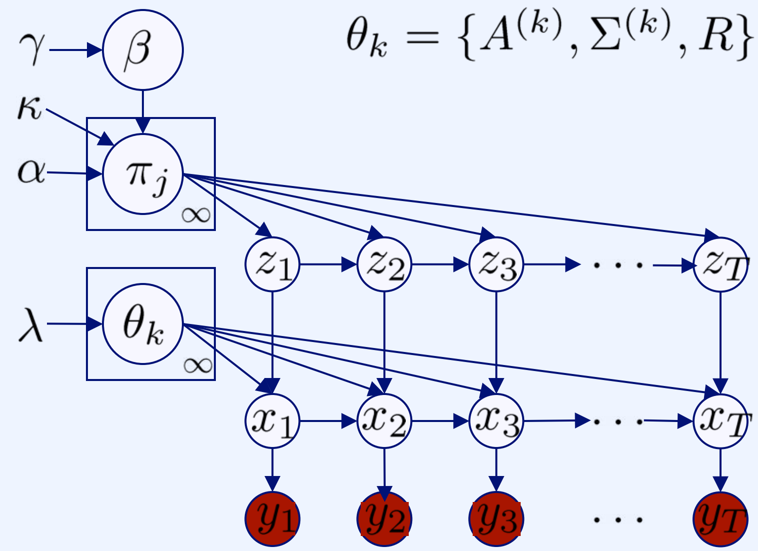
$$\theta_k = \{A_{1:r}^{(k)}, \Sigma^{(k)}\}$$



$$\begin{aligned} z_t &\sim \pi_{z_{t-1}} \\ y_t &= \sum_{i=1}^{r} A_i^{(z_t)} y_{t-i} + e_t(z_t) \end{aligned}$$

HDP-SLDS

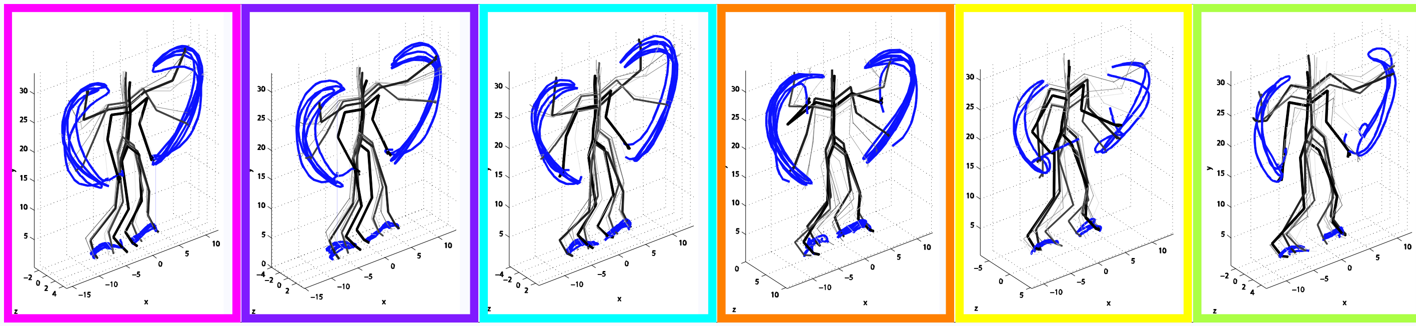
$$\theta_k = \{A^{(k)}, \Sigma^{(k)}, R\}$$



$$\begin{aligned} z_t &\sim \pi_{z_{t-1}} \\ x_t &= A^{(z_t)} x_{t-1} + e_t(z_t) \\ y_t &= C x_t + w_t \end{aligned}$$

$$C = [I \ 0]$$

Issue 4: Multiple Time Series

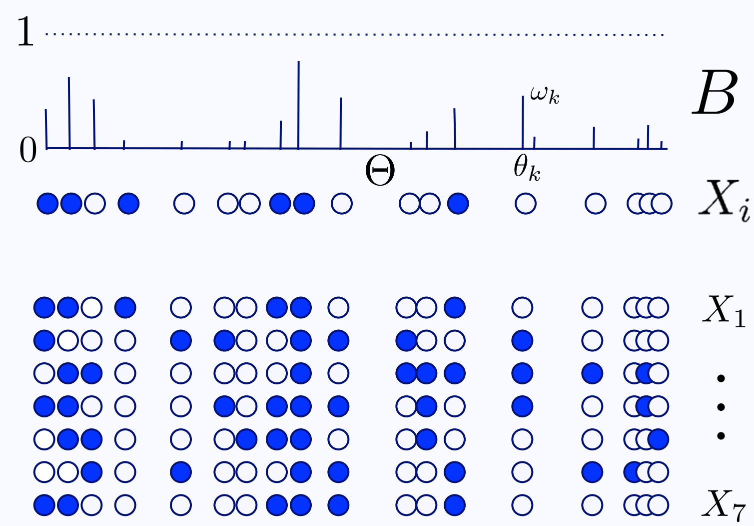


- Goal:
  - Transfer knowledge between related time series
  - Allow each system to switch between an arbitrarily large set of dynamical modes
- Method:
  - Beta process prior
  - Predictive distribution: Indian buffet process

Beta Processes & Featural Models

- Beta process-Bernoulli Process

$$
\begin{aligned}
B \mid B_0, c &\sim \mathrm{BP}(c, B_0) \\
X_i \mid B &\sim \mathrm{BeP}(B), \quad i = 1, \ldots, N
\end{aligned}
$$



- Equivalently

Points generated from Poisson process on: $\Theta \otimes [0,1]$

$$
\begin{aligned}
B &= \sum_k \omega_k \delta_{\theta_k} \\
X_i &= \sum_k f_{ik} \delta_{\theta_k}
\end{aligned}
$$

$f_{ik} \in \{0,1\}$ *result of coin flip w.p.* ω_k

- *Why is the beta process view helpful?*
  - ➢ *Stick-breaking representation useful for inference*
  - ➢ *Extensions to dependent and hierarchical featural models*
  - ➢ *Conceptual connections to other BNP models*

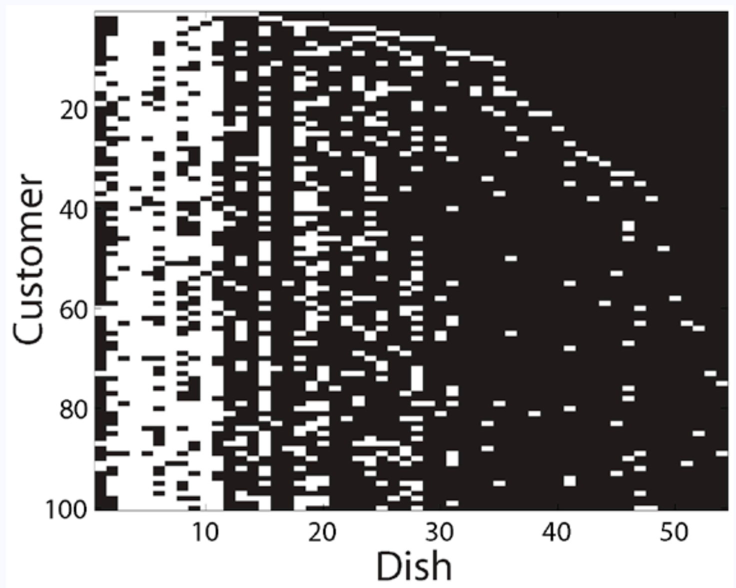
Indian Buffet Process (IBP)

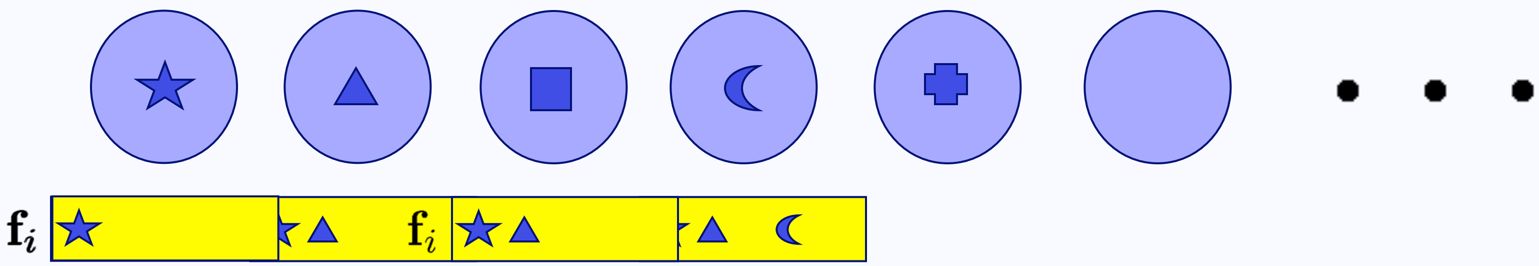
- Marginalize beta process measure
  → Indian buffet process (IBP)
  - Shared features:

$$p(f_{ik} \mid \mathbf{f}_1, \ldots, \mathbf{f}_{i-1}, \alpha) \propto \frac{m_k^{-i}}{i}$$

  - Unique features:

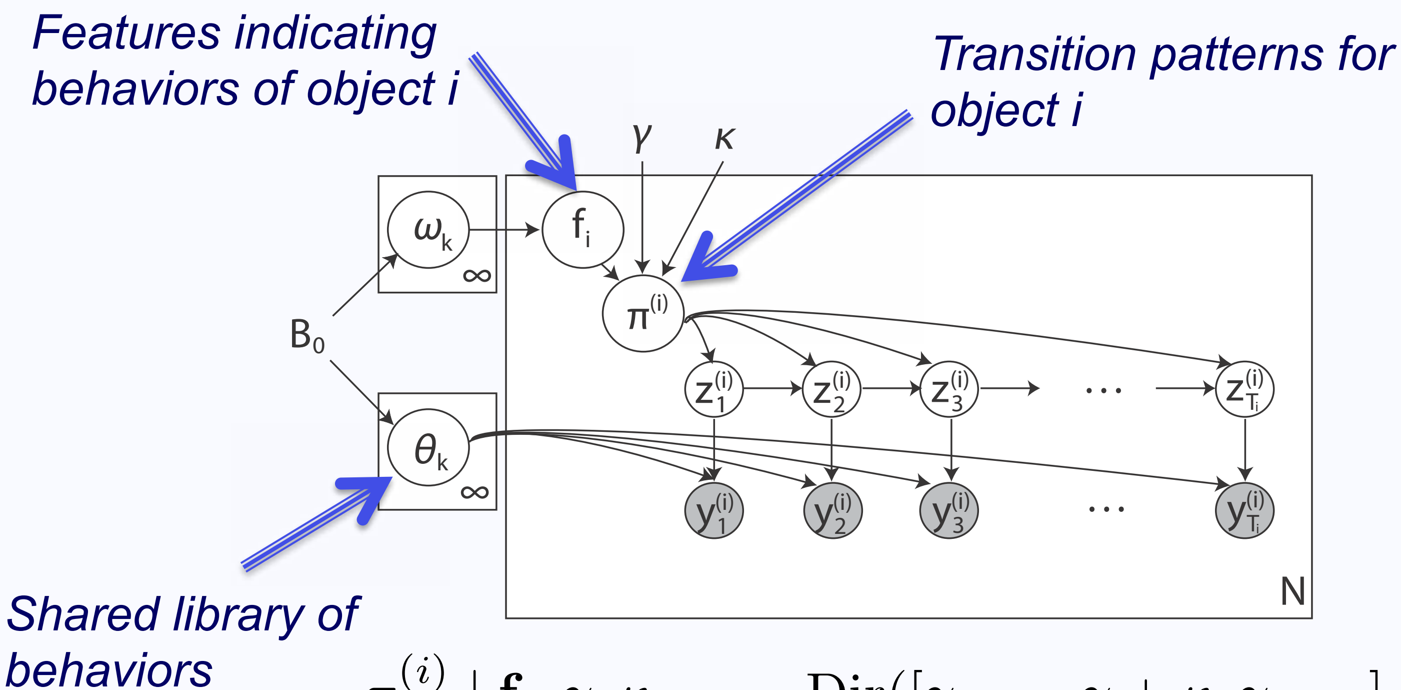
$$n_i \mid \alpha \sim \text{Poisson}\left(\frac{\alpha}{i}\right)$$





Griffiths and Ghahramani, TR, 2005 *Thibaux and Jordan, AISTATS, 2007*

BP-HMM



$$\pi_j^{(i)} \mid \mathbf{f}_i, \gamma, \kappa \sim \mathrm{Dir}([\gamma, \ldots, \gamma + \kappa, \gamma, \ldots] \otimes \mathbf{f}_i)$$

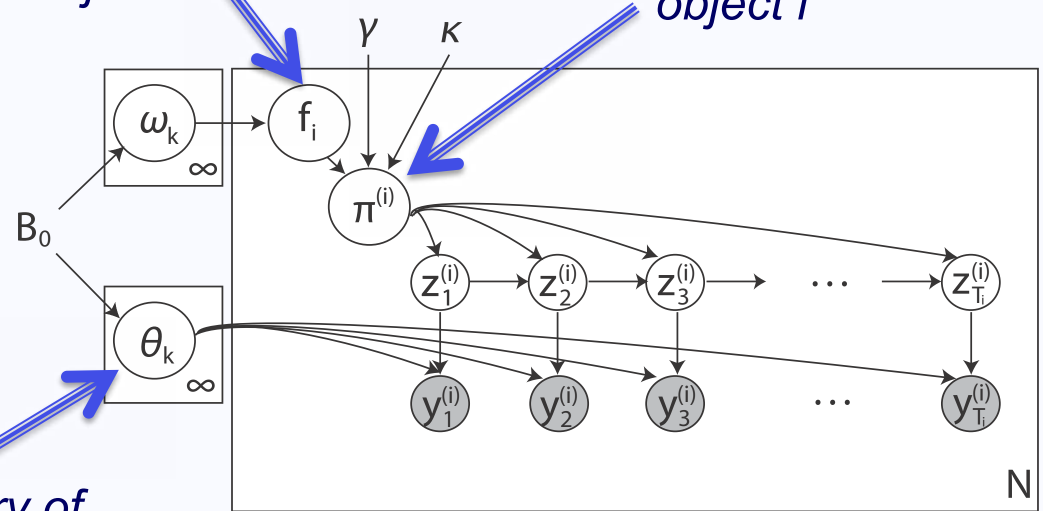
$$z_t^{(i)} \sim \pi^{(i)}_{z_{t-1}^{(i)}}$$

$$\mathbf{y}_t^{(i)} \mid z_t^{(i)} \sim \mathcal{N}\left(\mu_{z_t^{(i)}}, \Sigma_{z_t^{(i)}}\right)$$

BP-HMM

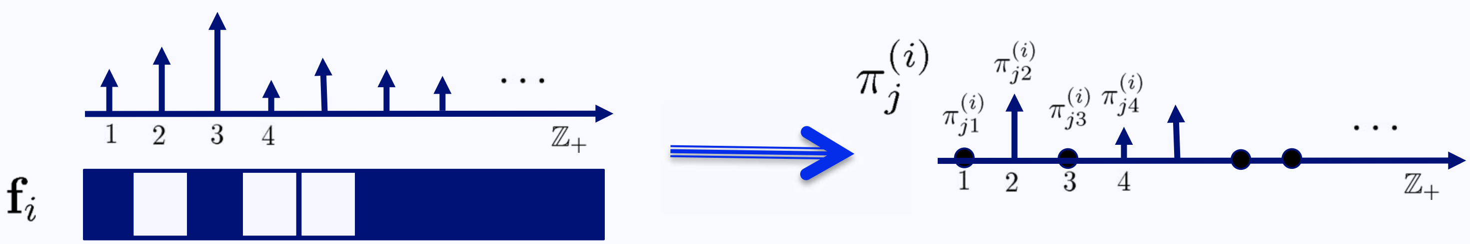
Features indicating behaviors of object i

Transition patterns for object i



Shared library of behaviors

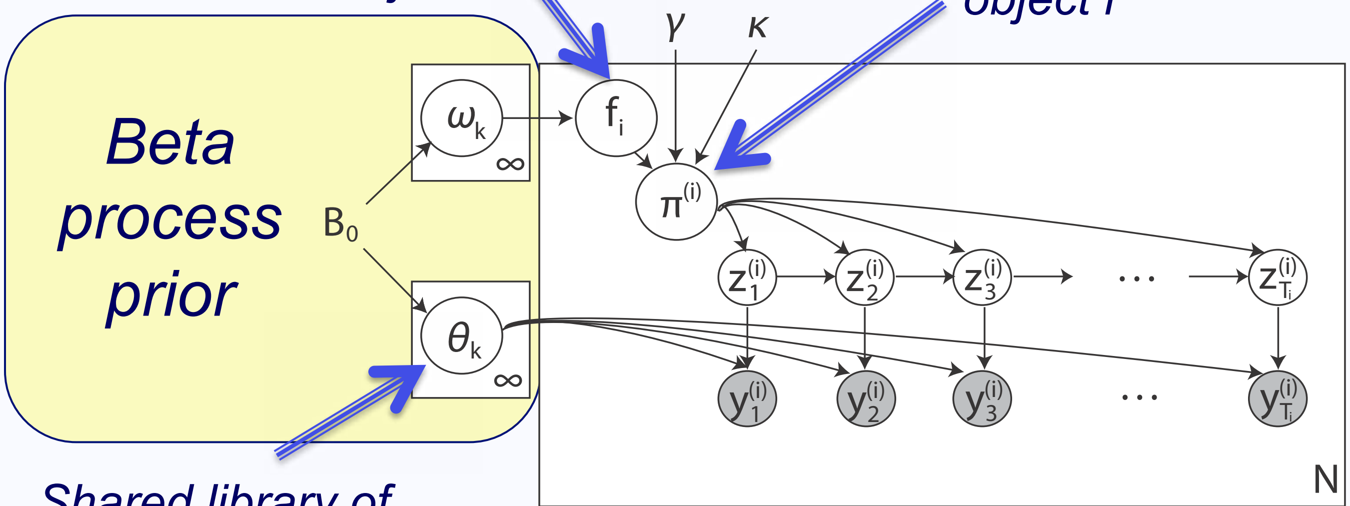
Features constrain dynamics to chosen behaviors



BP-HMM

Features indicating behaviors of object i

Transition patterns for object i



Shared library of behaviors

$$B \sim \text{Beta}(1, B_0) \qquad B = \sum_{k=1}^{\infty} \omega_k \delta_{\theta_k}$$

$$X_i \sim \text{Bernoulli}(B) \qquad X_i = \sum_{k=1}^{\infty} f_{ik} \delta_{\theta_k}$$

Beta process prior: Encourages sharing + allows variability

Switching VAR Process

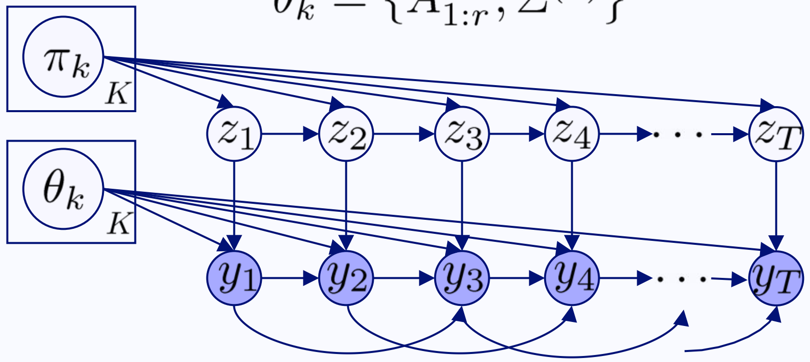
- *Alternative Markov switching process*
- *Captures more complex temporal dependencies*
- *Gaussian HMM is a special case*

$$z_t \sim \pi_{z_{t-1}}$$

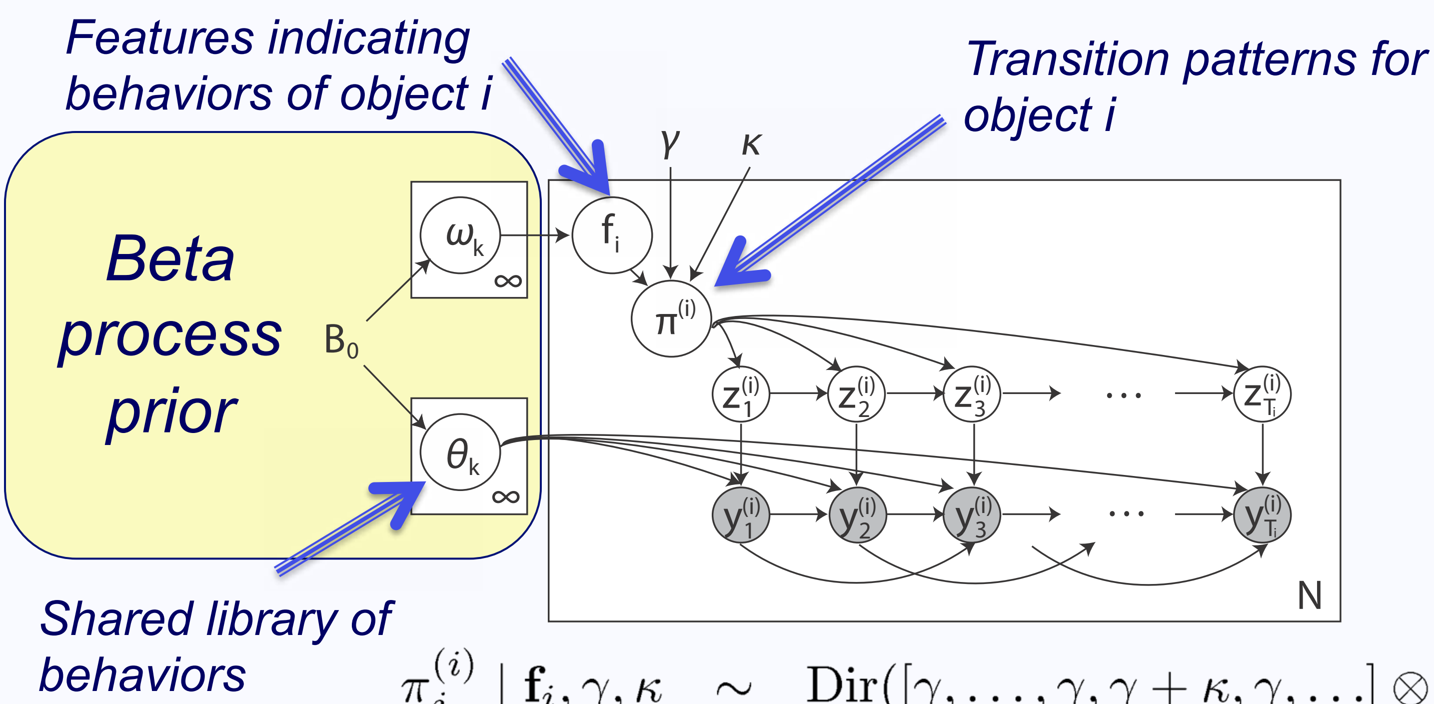
$$y_t = \sum_{i=1}^{r} A_i^{(z_t)} y_{t-i} + e_t(z_t)$$

$$e_t \sim \mathcal{N}(0, \Sigma^{(z_t)})$$

$$\theta_k = \{A_{1:r}^{(k)}, \Sigma^{(k)}\}$$



BP-AR-HMM



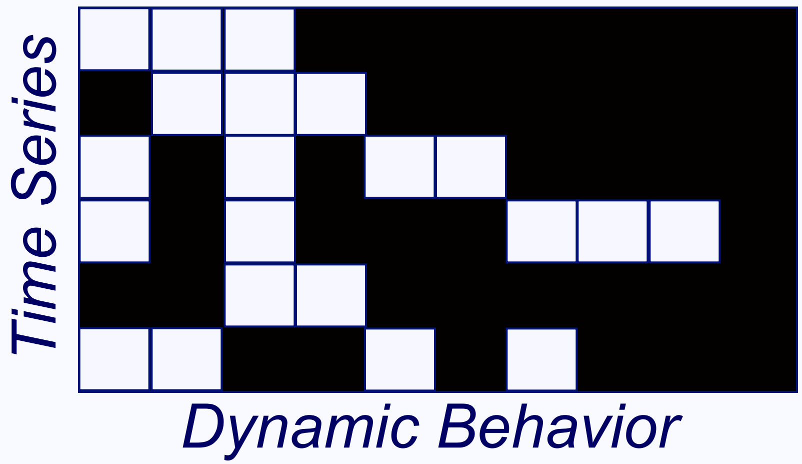
$$\pi_j^{(i)} \mid \mathbf{f}_i, \gamma, \kappa \sim \mathrm{Dir}([\gamma, \ldots, \gamma, \gamma + \kappa, \gamma, \ldots] \otimes \mathbf{f}_i)$$

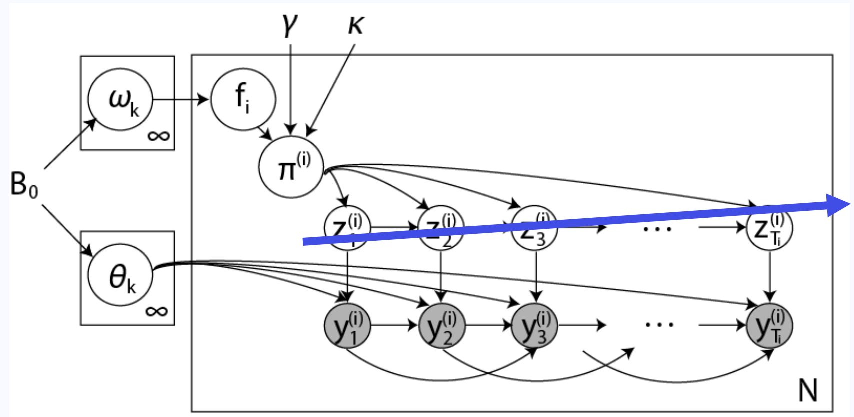
$$z_t^{(i)} \sim \pi^{(i)}_{z_{t-1}^{(i)}}$$

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

BP-AR-HMM Inference

- Metropolis-Hastings feature sampling:





$$p(f_{ik} \mid \mathbf{F}^{-ik}, \mathbf{y}_{1:T_i}^{(i)}, \pi^{(i)}, \theta_{1:K_+^{-i}}, \alpha) \propto \underbrace{p(f_{ik} \mid \mathbf{F}^{-ik}, \alpha)}_{\text{IBP prior}} \underbrace{p(\mathbf{y}_{1:T_i}^{(i)} \mid \mathbf{f}_i, \pi^{(i)}, \theta_{1:K_+^{-i}})}_{\text{Likelihood}}$$

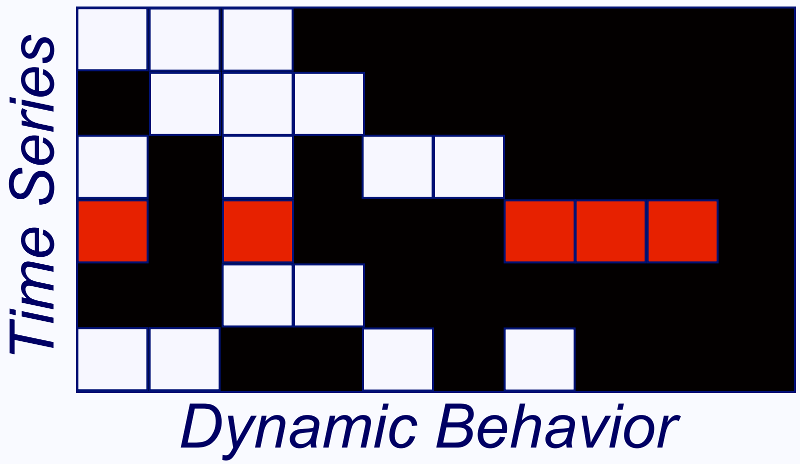
IBP prior

Likelihood of observations given feature-constrained transition distributions

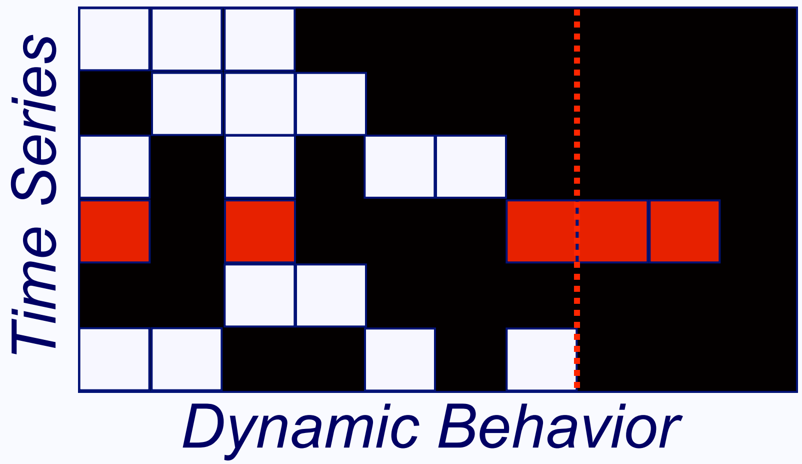
Forward-Backward Algorithm

BP-AR-HMM Inference

- Examine i^{th} object

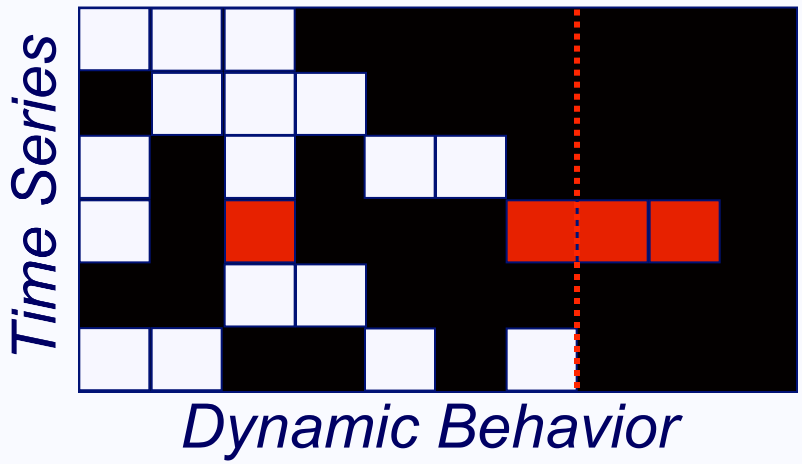


BP-AR-HMM Inference



- *Examine i^{th} object*
- *Consider shared and unique features separately*

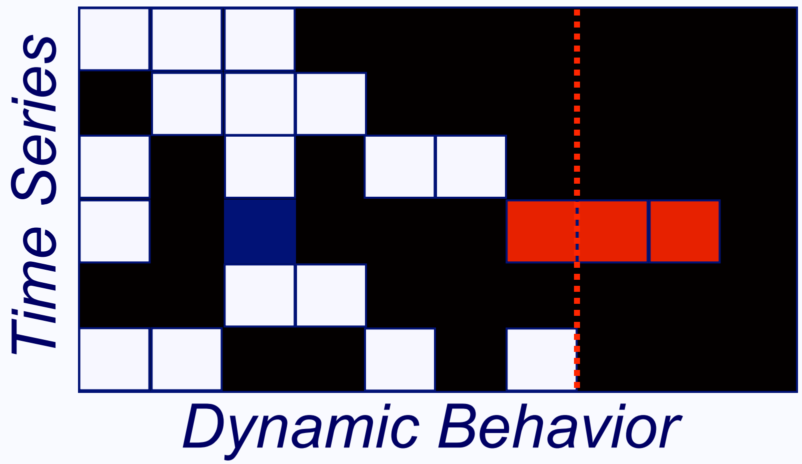
BP-AR-HMM Inference



- *Examine i^{th} object*
- *Consider shared and unique features separately*
- *For each shared feature k, sample f_{ik} using:*

$$p(f_{ik} \mid \mathbf{F}^{-ik}, \alpha) \propto \frac{m_k^{-i}}{N}$$

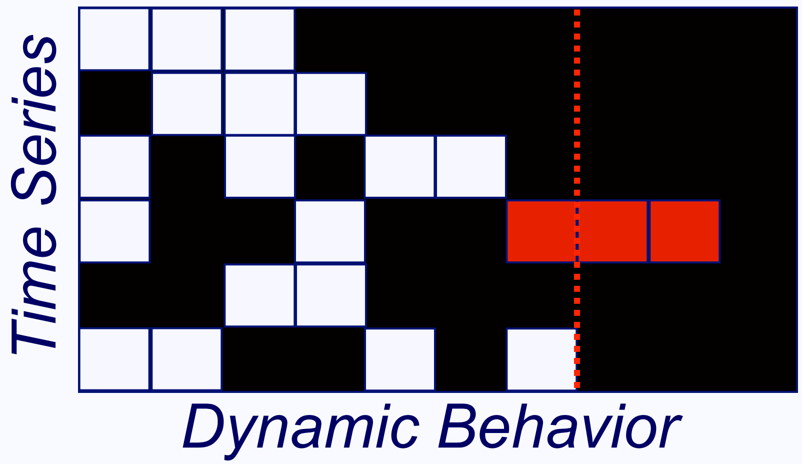
BP-AR-HMM Inference



- *Examine ith object*
- *Consider shared and unique features separately*
- *For each shared feature k, sample f_{ik} using:*

$$p(f_{ik} \mid \mathbf{F}^{-ik}, \alpha) \propto \frac{m_k^{-i}}{N}$$

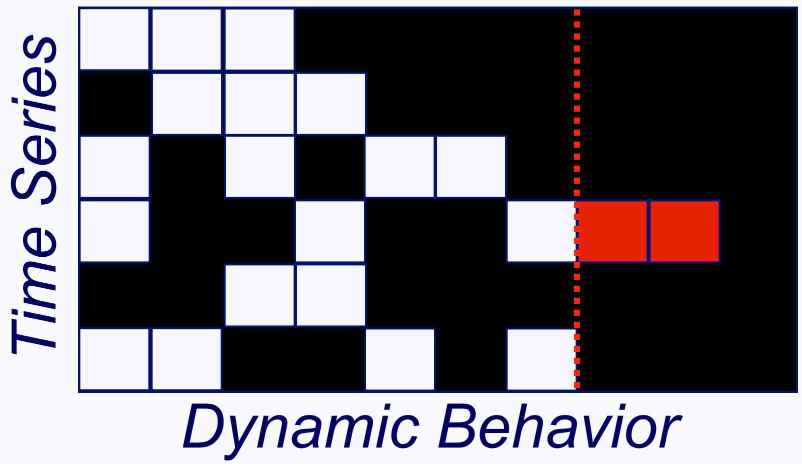
BP-AR-HMM Inference



- *Examine i^{th} object*
- *Consider shared and unique features separately*
- *For each shared feature k, sample f_{ik} using:*

$$p(f_{ik} \mid \mathbf{F}^{-ik}, \alpha) \propto \frac{m_k^{-i}}{N}$$

BP-AR-HMM Inference



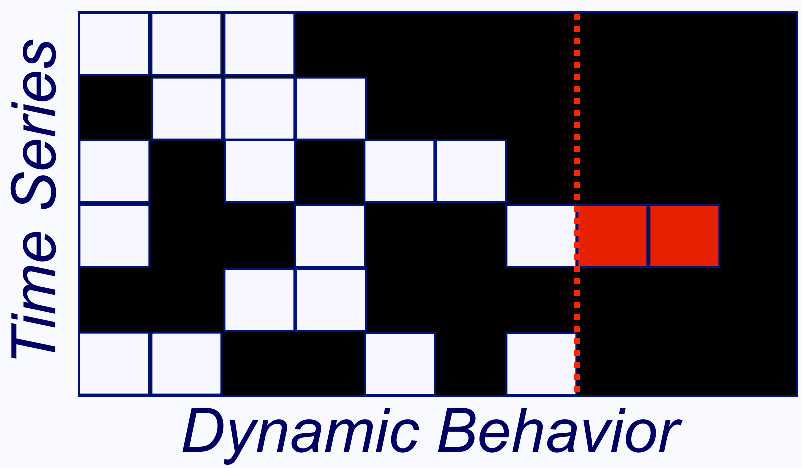
- *Examine i^{th} object*
- *Consider shared and unique features separately*
- *For each shared feature k, sample f_{ik} using:*

$$p(f_{ik} \mid \mathbf{F}^{-ik}, \alpha) \propto \frac{m_k^{-i}}{N}$$

- *Unique features:*

$$n_i \mid \alpha \sim \text{Poisson}\left(\frac{\alpha}{N}\right)$$

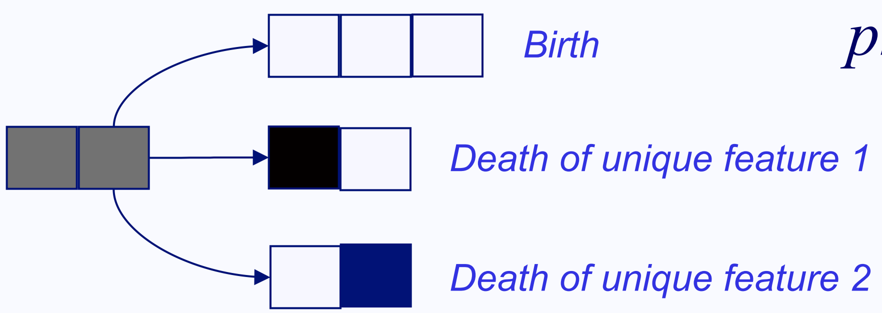
BP-AR-HMM Inference



- *Examine i^th object*
- *Consider shared and unique features separately*
- *For each shared feature k, sample f_{ik} using:*

$$p(f_{ik} \mid \mathbf{F}^{-ik}, \alpha) \propto \frac{m_k^{-i}}{N}$$

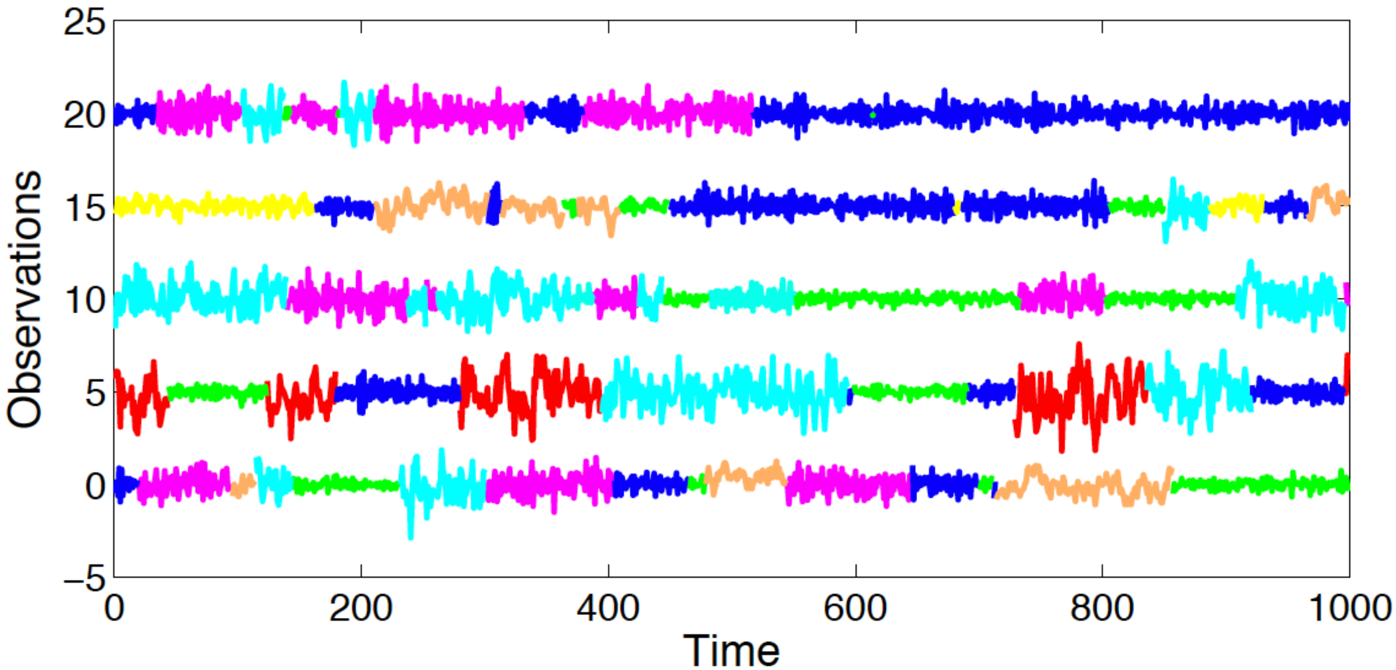
- *Use birth-death RJ-MCMC to propose a new unique features*

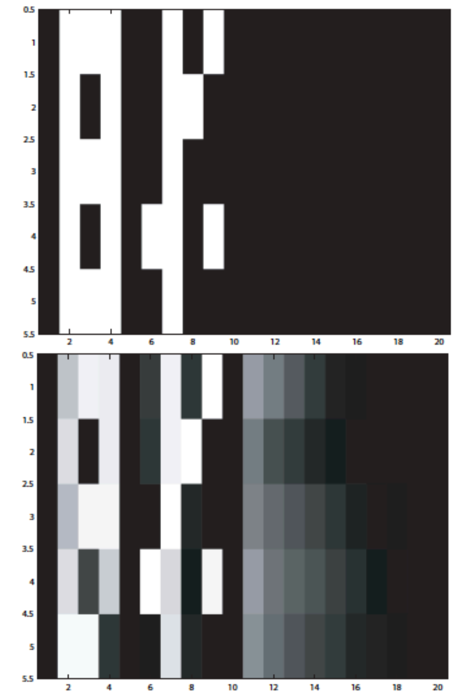


Gorur et. al., ICML, 2006

Meeds et. al., NIPS, 2007

Validation on Toy Data



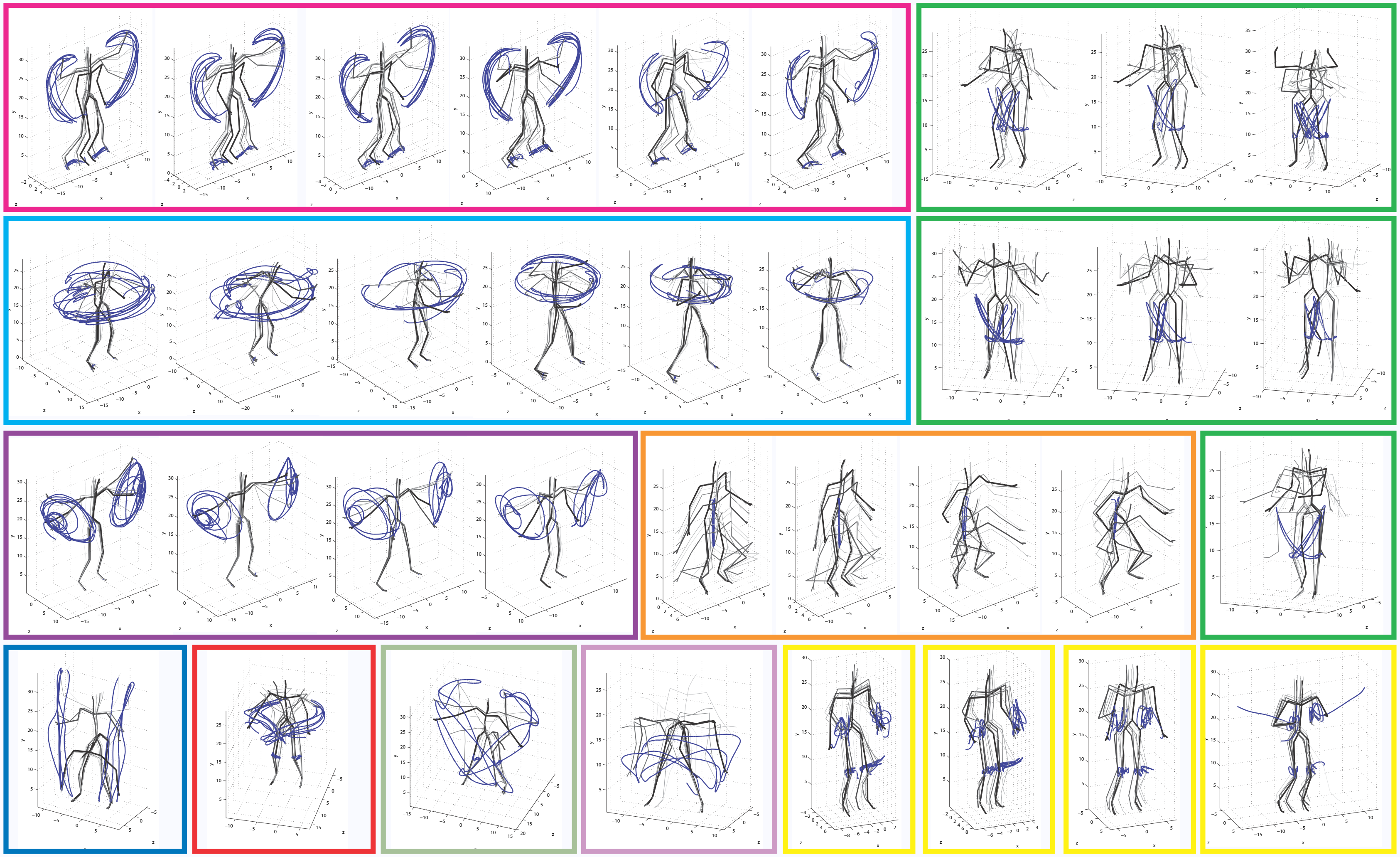


Motion Capture

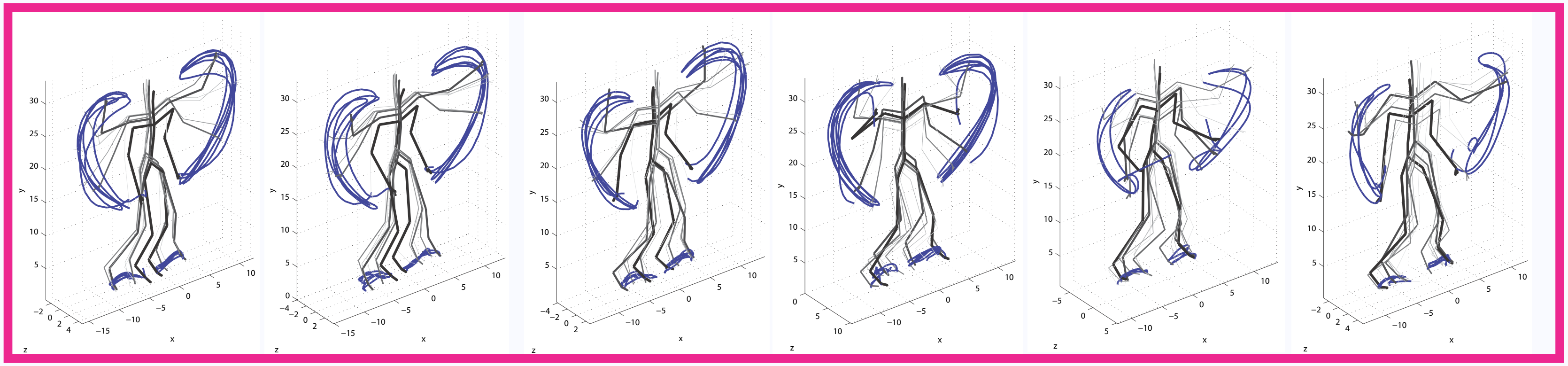
6 videos of exercise routines

CMU MoCap: http://mocap.cs.cmu.edu/

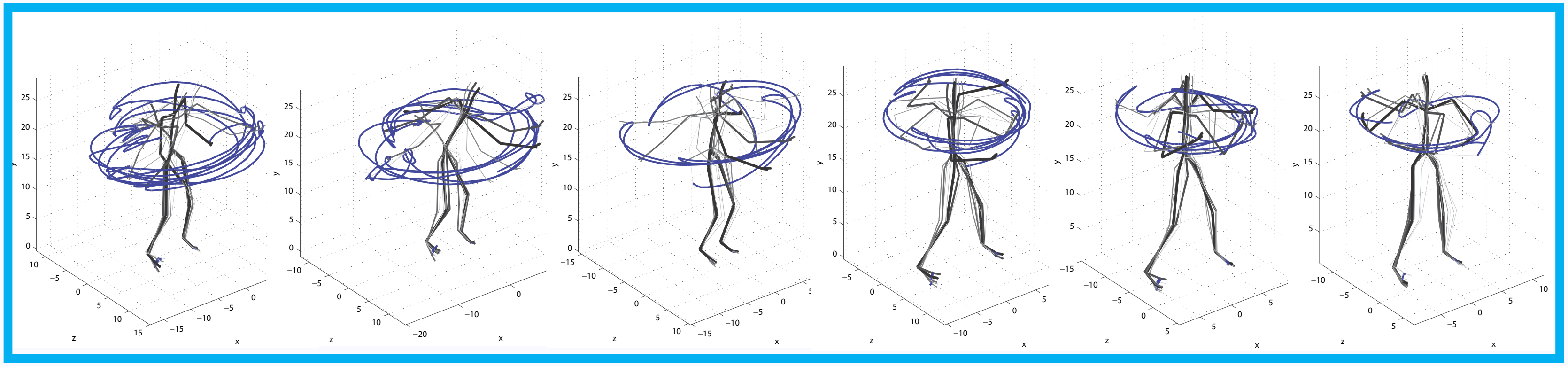
Motion Capture Results - I



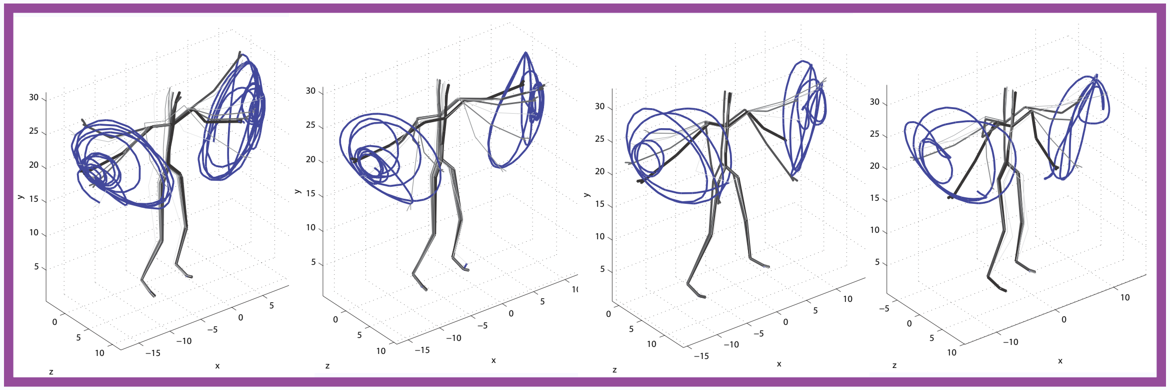
Motion Capture Results - I



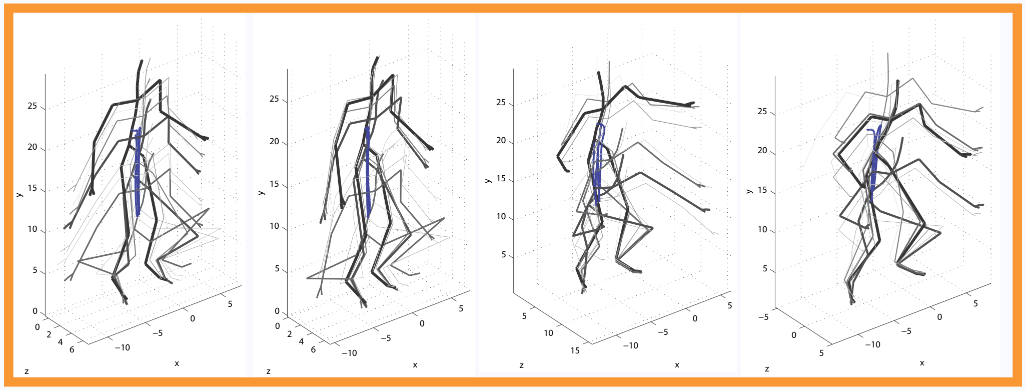
Motion Capture Results - I



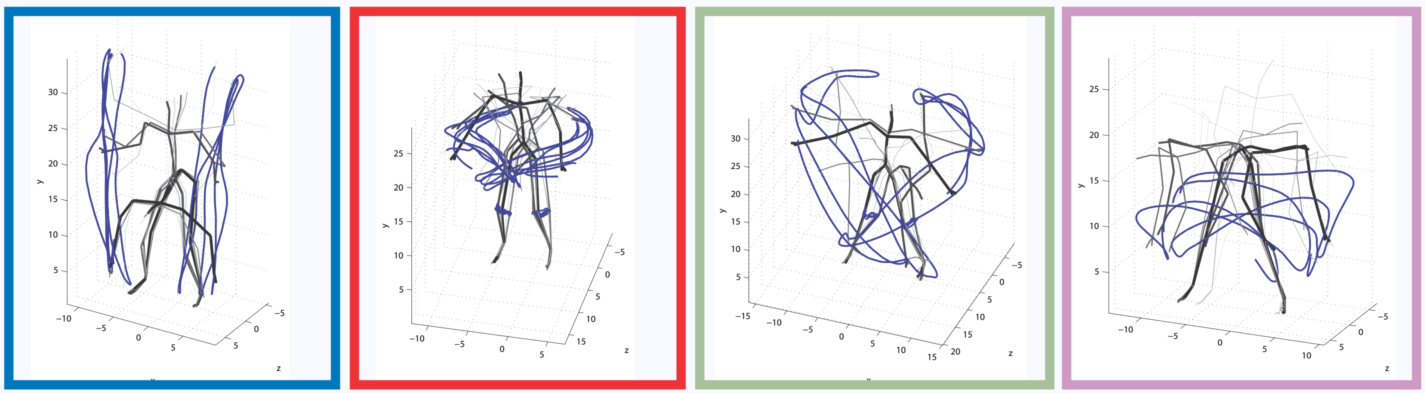
Motion Capture Results - I



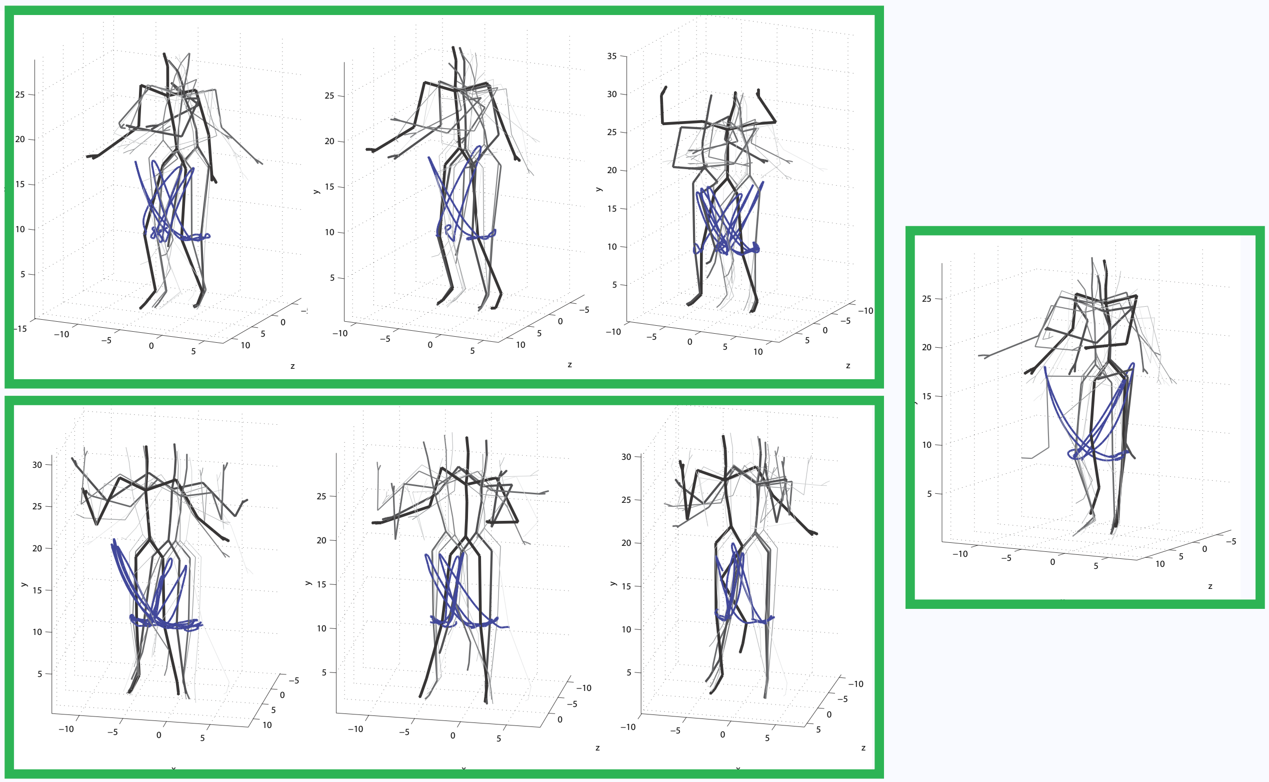
Motion Capture Results - I



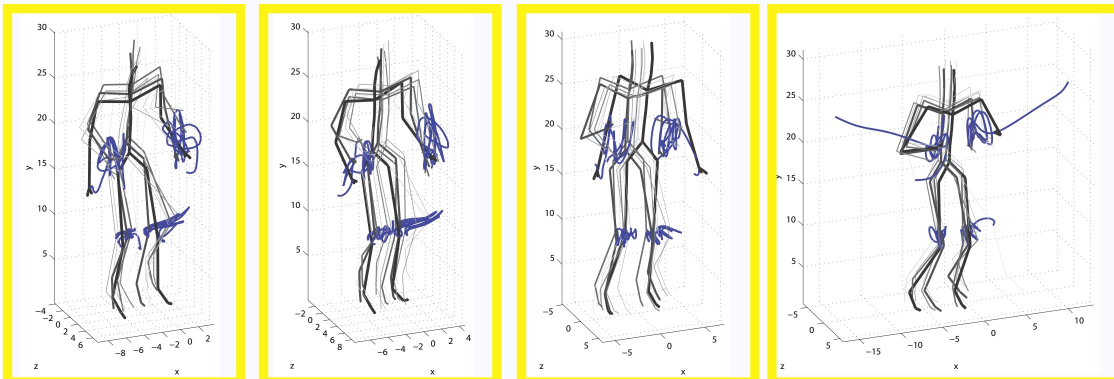
Motion Capture Results - I



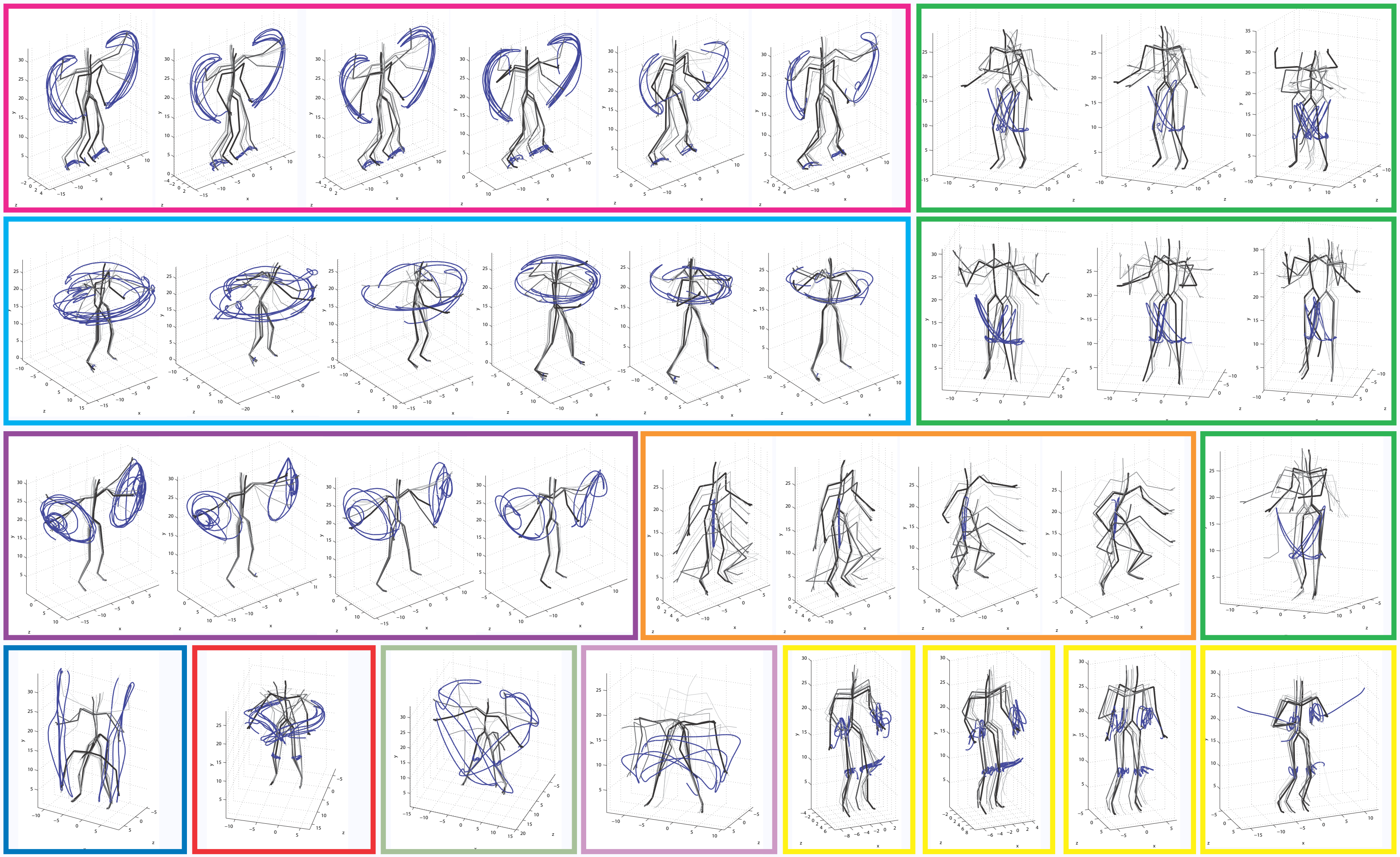
Split Motions



Split Motions

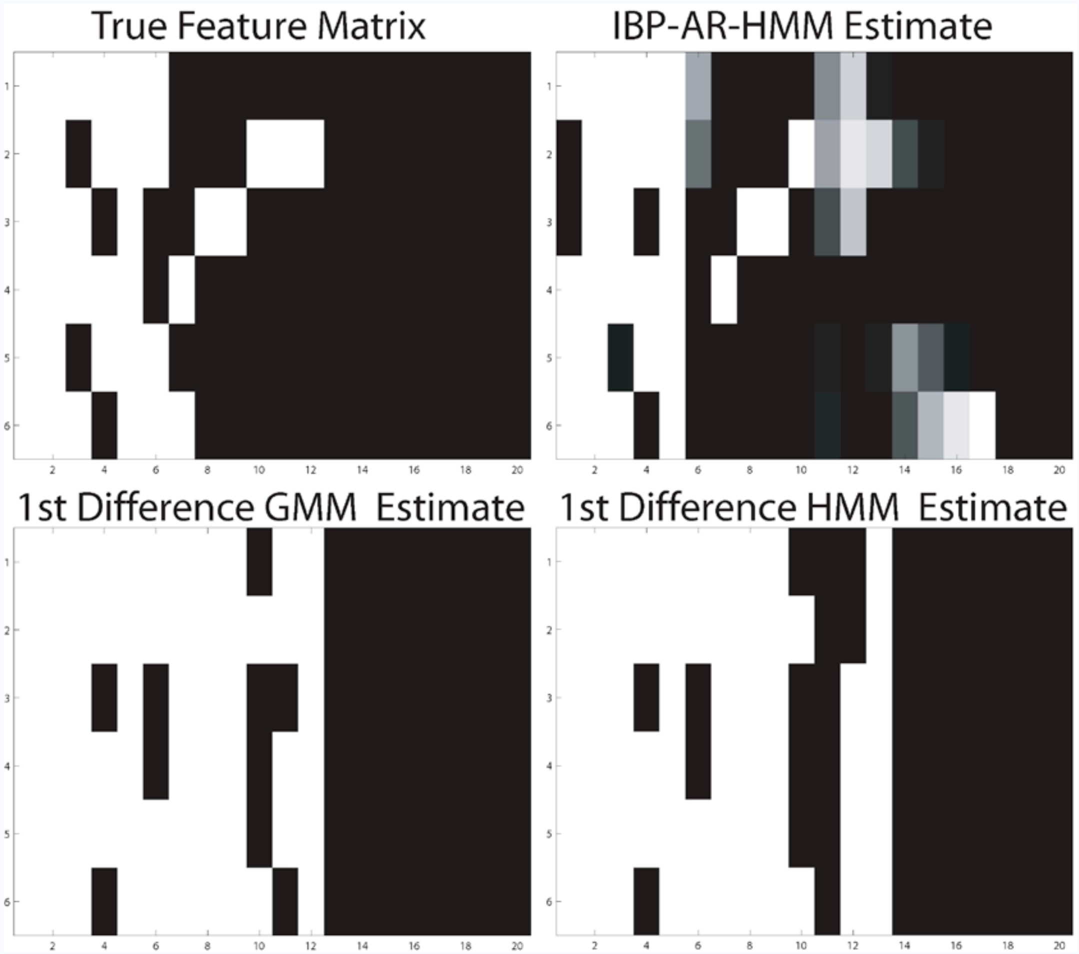


Motion Capture Results - I



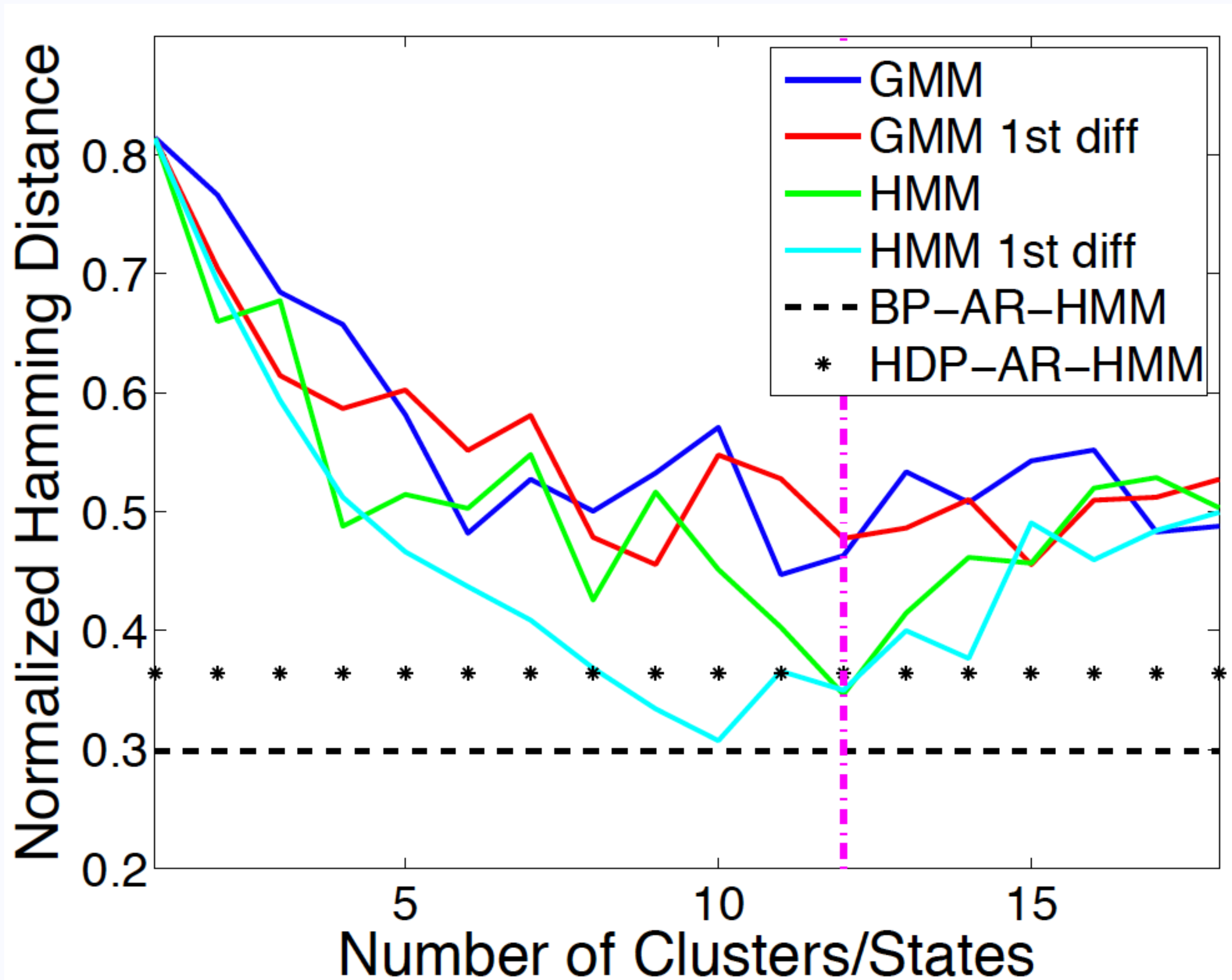
Motion Capture Results - II

- Learned feature matrices:



Motion Capture Results - II

- Comparison to parametric mixtures & HMMs:



Next: Infinite Factorial HMMs

Beta Process HMM

- *Combinatorial structure in the relationships among different time series*
- *Single Markov process within each time series*

Infinite Factorial HMM

- *Combinatorial structure of temporal dynamics within a single time series*
- *No consideration of relationships among time series*

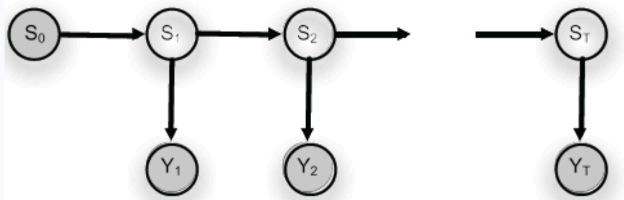


Figure 1: The Hidden Markov Model

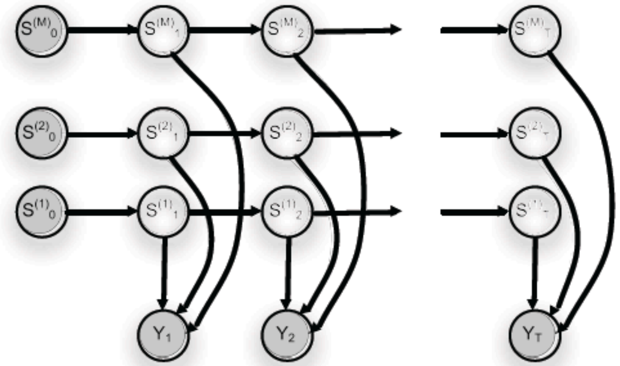


Figure 2: The Factorial Hidden Markov Model

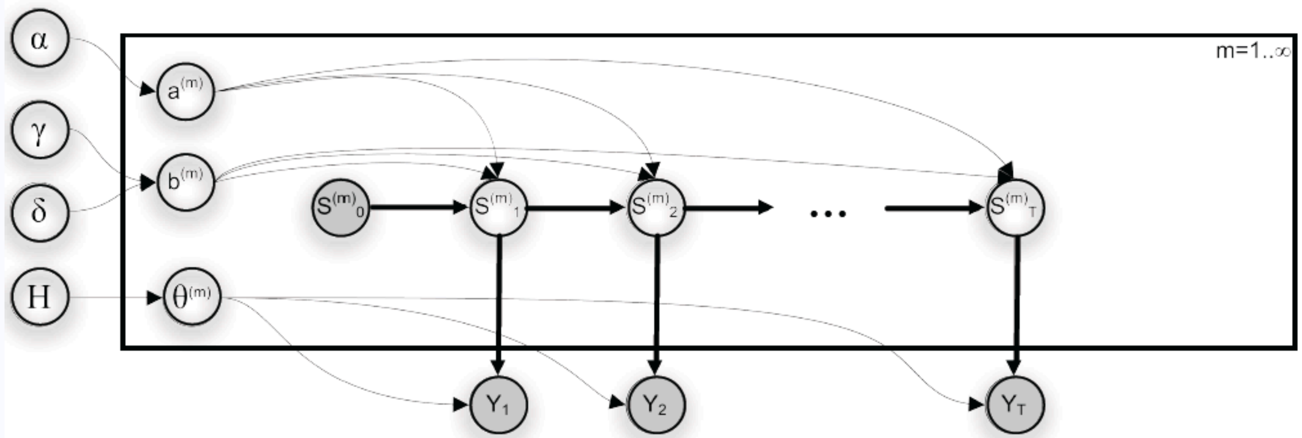
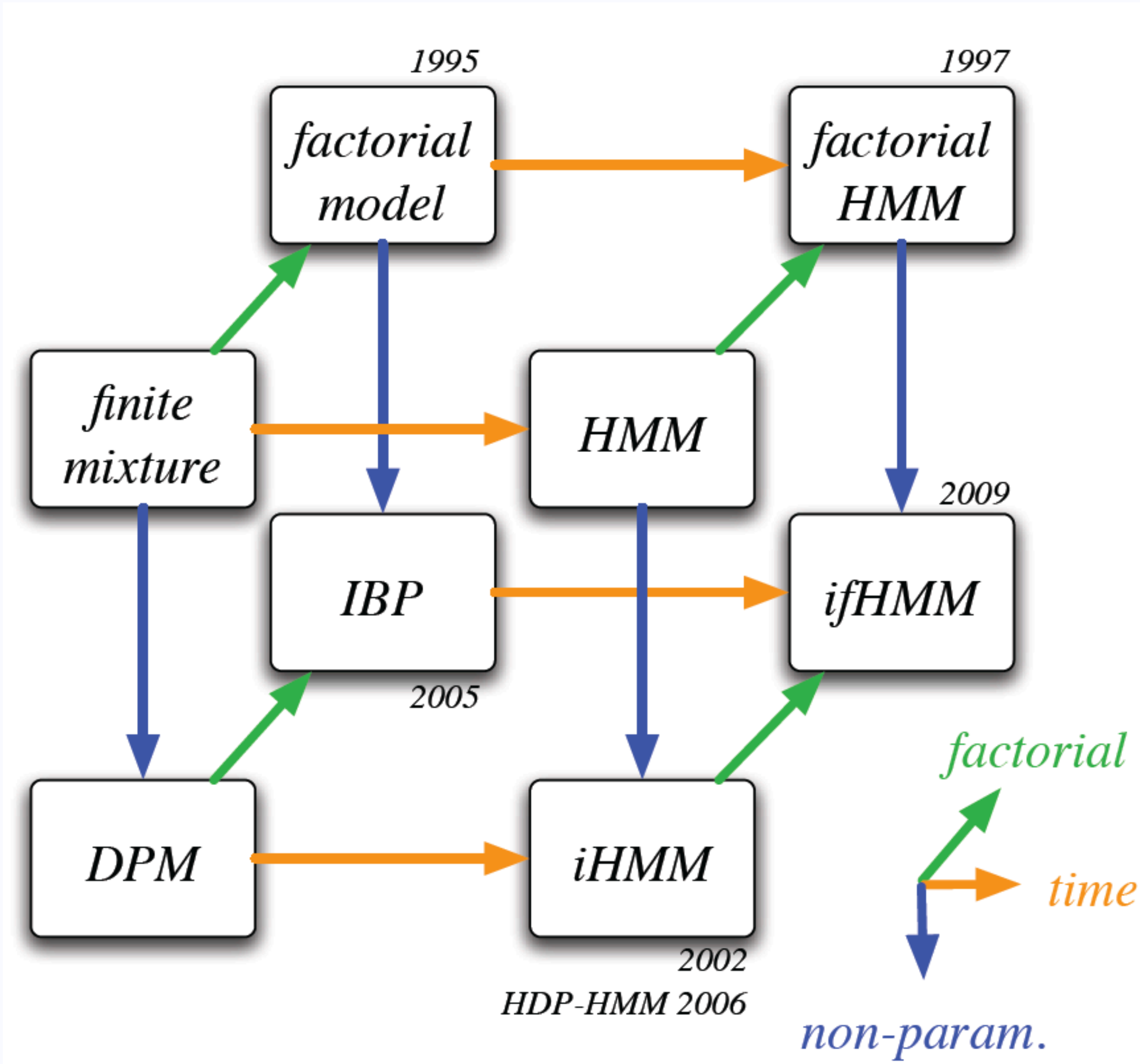


Figure 3: The Infinite Factorial Hidden Markov Model

Ghahramani, BNP Workshop 2009

A Model Evolution Flow Chart



Ghahramani, BNP Workshop 2009