The Infinite Factorial Hidden Markov Model

Jurgen Van Gael

Yee Whye Teh

Zoubin Ghahramani

Presented by
Zachary Kahn

Motivation

- HMMs allow us to use latent variables to describe the emissions of data at discrete steps, but the latent variable can only describe one hidden state
- It might be that multiple hidden states are combining to create your data in a more complicated fashion

Graphical Model

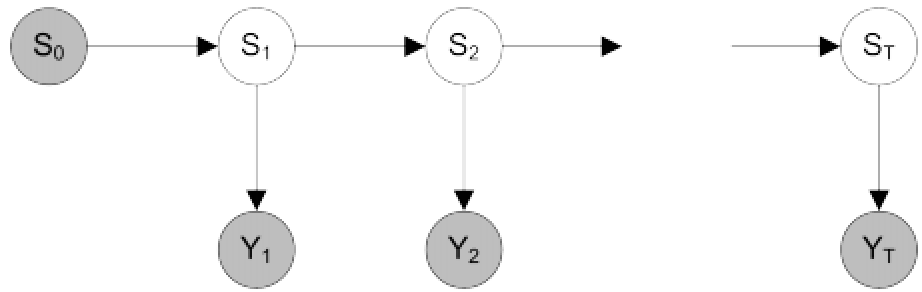


Figure 1: The Hidden Markov Model

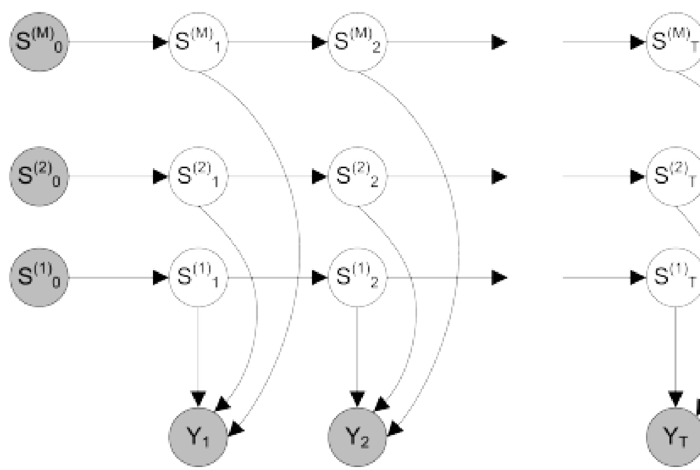


Figure 2: The Factorial Hidden Markov Model

Finite Model

- S = Binary Matrix with T rows (data points) and M columns (features)
- c_m^{ij} = # transitions from state i to j in chain m

$$\boldsymbol{W}^{(m)} = \begin{pmatrix} 1-a_m & a_m \\ 1-b_m & b_m \end{pmatrix} \qquad \boldsymbol{W}_{ij}^{(m)} = p(s_{t+1,m} = j | s_{tm} = i)$$

$$\forall m \in \{1, 2, \cdots, M\} : a_m \sim \text{Beta}\left(\frac{\alpha}{M}, 1\right) \quad , \quad b_m \sim \text{Beta}(\gamma, \delta),$$

$$s_{0m} = 0 \quad , \quad s_{tm} \sim \text{Bernoulli}(a_m^{1-s_{t-1,m}} b_m^{s_{t-1,m}}).$$

$$p(\boldsymbol{S}|\boldsymbol{a}, \boldsymbol{b}) = \prod_{m=1}^{M} (1-a_m)^{c_m^{00}} a_m^{c_m^{01}} (1-b_m)^{c_m^{10}} b_m^{c_m^{11}}.$$

$$p(\boldsymbol{S}|\alpha, \gamma, \delta) = \prod_{m=1}^{M} \frac{\frac{\alpha}{M}\Gamma(\frac{\alpha}{M} + c_m^{01})\Gamma(c_m^{00}+1)\Gamma(\gamma+\delta)\Gamma(\delta+c_m^{10})\Gamma(\gamma+c_m^{11})}{\Gamma(\frac{\alpha}{M}+c_m^{00}+c_m^{01}+1)\Gamma(\gamma)\Gamma(\delta)\Gamma(\gamma+\delta+c_m^{10}+c_m^{11})}$$

Infinite Limit

- Just taking the limit as M ->∞ we get a probability of zero, need to use lof-equivalence classes

$$p([\boldsymbol{S}]) = \sum_{\boldsymbol{S}\in[\boldsymbol{S}]} p(\boldsymbol{S}|\alpha,\gamma,\delta)$$

$$= \frac{M!}{\prod_{h=0}^{2^T-1} M_h!} \prod_{m=1}^{M} \frac{\frac{\alpha}{M}\Gamma(\frac{\alpha}{M}+c_m^{01})\Gamma(c_m^{00}+1)\Gamma(\gamma+\delta)\Gamma(\delta+c_m^{10})\Gamma(\gamma+c_m^{11})}{\Gamma(\frac{\alpha}{M}+c_m^{00}+c_m^{01}+1)\Gamma(\gamma)\Gamma(\delta)\Gamma(\gamma+\delta+c_m^{10}+c_m^{11})}.$$

$$\lim_{M\to\infty} p([\boldsymbol{S}]) = \frac{\alpha^{M_+}}{\prod_{h=0}^{2^T-1} M_h!} \exp\{-\alpha H_T\} \prod_{m=1}^{M_+} \frac{(c_m^{01}-1)!c_m^{00}!\Gamma(\gamma+\delta)\Gamma(\delta+c_m^{10})\Gamma(\gamma+c_m^{11})}{(c_m^{00}+c_m^{01})!\Gamma(\gamma)\Gamma(\delta)\Gamma(\gamma+\delta+c_m^{10}+c_m^{11})}$$

- H_t= Harmonic # t
- M_+= # active Markov chains

A Modified Indian Buffet Process

- First Customer takes Poisson(α) dishes starting from left.
- The t'th customer looks at dish m
  - If t-1'th customer took dish m, t'th takes with prob $(c_m^{11}+\delta)/(\gamma+\delta+c_m^{10}+c_m^{11})$
  - If not, t'th takes with prob $c_m^{00}/(c_m^{00}+c_m^{01})$
- He then takes Poisson(α/t) new dishes

$$p([\mathbf{S}]) = \frac{\alpha^{M_+}}{\prod_{t=1}^{T} M_1^{(t)}!} \exp\{-\alpha H_T\} \prod_{m=1}^{M} \frac{\frac{\alpha}{M}\Gamma(\frac{\alpha}{M}+c_m^{01})\Gamma(c_m^{00}+1)\Gamma(\gamma+\delta)\Gamma(\delta+c_m^{10})\Gamma(\gamma+c_m^{11})}{\Gamma(\frac{\alpha}{M}+c_m^{00}+c_m^{01}+1)\Gamma(\gamma)\Gamma(\delta)\Gamma(\gamma+\delta+c_m^{10}+c_m^{11})}.$$

$M_+ \sim \text{Poisson}(\alpha H_T)$

Stick Breaking Representation

- While theoretically convenient, the previous models are not practical for inference. Instead, stick breaking will be more tractable.

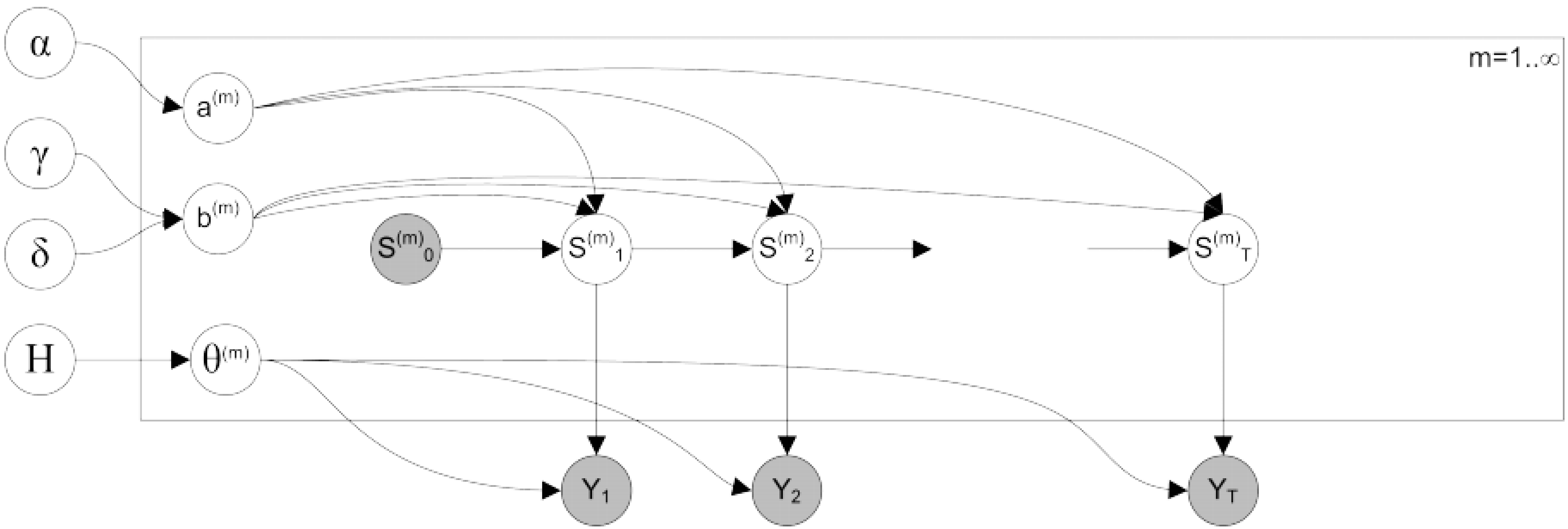
$$
\begin{aligned}
a_{(1)} &\propto \text{Beta}(\alpha, 1), \\
p(a_{(m)} | a_{(m-1)}) &= \alpha a_{(m-1)}^{-\alpha} a_{(m)}^{\alpha-1} \mathbb{I}(0 \leq a_{(m)} \leq a_{(m-1)}).
\end{aligned}
$$

$$b_{(m)} \sim \text{Beta}(\gamma, \delta)$$

The Infinite Factorial HMM

- To use the mIBP as a probabilistic model, we need to add feature properties through $\theta_m \sim H$
- Need to define conditional probability over observations given latent features $F(y_t|\boldsymbol{\theta}, \boldsymbol{s}_{t,\cdot})$
- In order to be valid in the infinite limit, we require the probability to be invariant to permutations of features, and independent of θ_m if $s_{tm} = 0$.

iFHMM Graphical Model



Independent Component Analysis

- Assume that M signals are represented as vectors x_m and $X = [x_1 x_2 \ldots x_M]$.
- Signals are combined using mixing matrix W to generate Y=XW
  - Also assume IID Normal($0,\sigma_Y^2$) ϵ s.t. Y=XW+ϵ
- There exist fast algorithms to extract X from Y (e.g. ICA) but they depend on the number of signals being known in advance.

ICA iFHMM Generative Model

- S ~ mIBP
- X_{ij} ~ Laplace(0, 1) i.i.d.
- W_{ij} ~ Normal(0, σ_W^2) i.i.d
- ϵ ~ Normal(0, σ_Y^2)
- Y = (S $\odot$ X)W + ϵ

Inference Plan

- Nonparametric models are typically inferred using Gibbs sampling with augmented Metropolis Hastings steps.
- Gibbs sampling is known to be bad in time series though due to string coupling in successive steps.
- Can avoid this using a dynamic programming solution with stick breaking.

Auxiliary Slice variable

$$\mu \sim \text{Uniform}(0, \min_{m:\exists t, s_{tm}=1} a_m).$$

$$p(\mu, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{S}) = p(\mu|\boldsymbol{a}, \boldsymbol{S}) p(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{S}).$$

$$p(\boldsymbol{S}|\boldsymbol{Y}, \mu, \boldsymbol{a}, \boldsymbol{b}) \propto p(\boldsymbol{S}|\boldsymbol{Y}, \boldsymbol{a}, \boldsymbol{b}) \frac{\mathbb{I}(0 \le \mu \le \min_{m:\exists t, s_{tm}=1} a_m)}{\min_{m:\exists t, s_{tm}=1} a_m}$$

- The slice variables force all columns of S where $a_m < \mu$ to be 0, allowing us to resample a finite number of columns in S

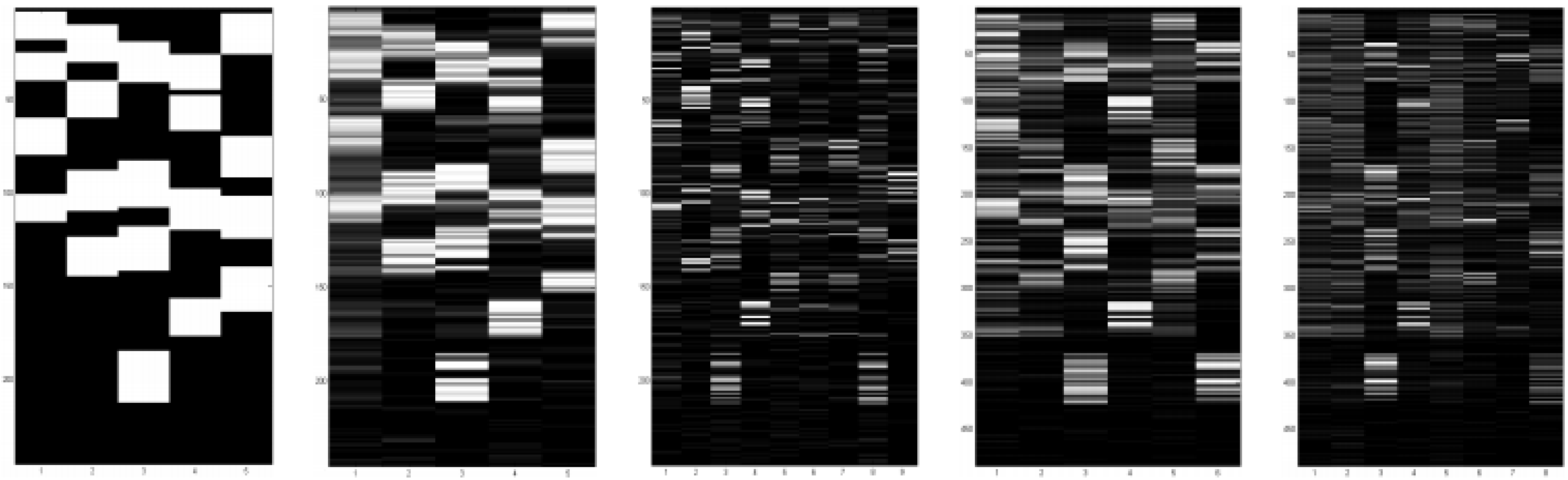
Inference Algorithm

- Start with an initial S and sample a and b. Then sample an initial X and W. Then iterate
  1. Sample the auxiliary variable μ
  2. Sample S, X, and W for all represented features
  3. Resample the hyperparameters.
  4. Remove all unused features.

Sampling methods

- There are multiple ways that S, X, and W can be sampled in step 2
  - A naïve Gibbs sampler performs badly as expected
  - A blocked sampler that fixes all but one column of S and runs a forwards-backwards algorithm.
  - A third sampler runs dynamic programming on multiple chains with the possibility to merge features, but you can’t integrate out X and W.

Results



(a) Ground Truth (b) ICA iFHMM (c) iICA (d) ICA iFHMM (e) iICA

Questions?