The Infinite Factorial Hidden Markov Model

Computational & Biological Learning Lab
University of Cambridge

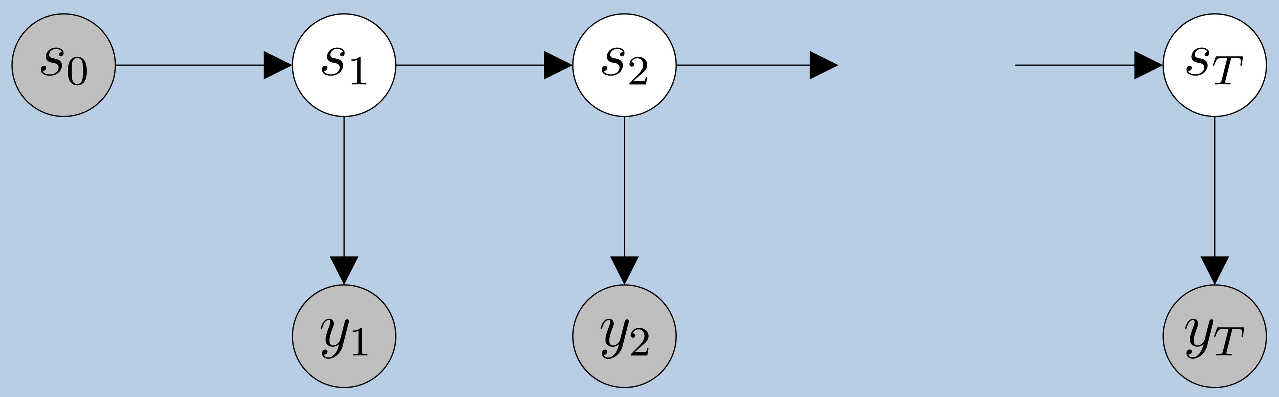
Jurgen Van Gael#, Yee Whye Teh* & Zoubin Ghahramani

1. Abstract

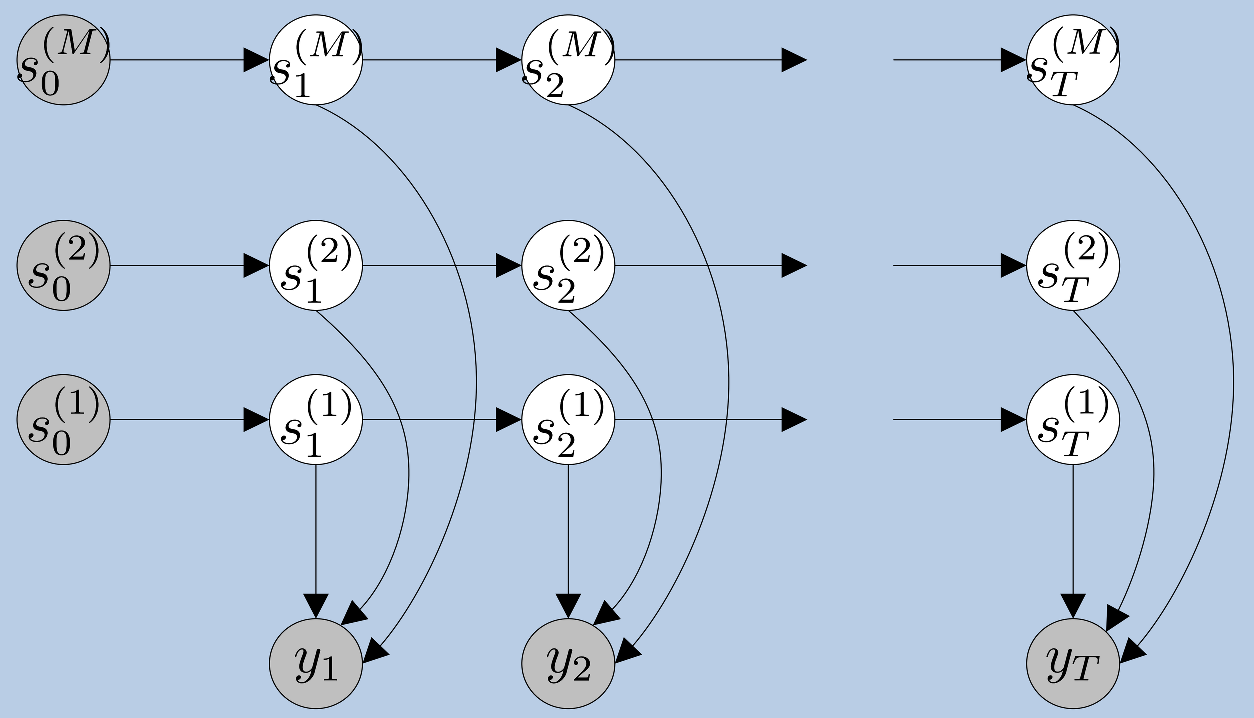
We introduce a new probability distribution over a potentially infinite number of binary Markov chains which we call the Markov Indian Buffet Process (mIBP). This process extends the Indian Buffet Process to allow temporal dependencies in the hidden variables. We use this stochastic process to build a nonparametric extension of the factorial hidden Markov model. After deriving an inference scheme which combines slice sampling and dynamic programming we demonstrate how the infinite factorial hidden Markov model can be used for blind source separation.

2. The Factorial HMM

Consider the problem of recording a meeting and segmenting the continuous speech signal into sections where only one speaker is active at a time. A hidden Markov model (HMM) (graphical model below) could solve this problem by mapping each hidden state to a speaker.



A slightly more complicated problem is to record an audio signal when multiple people are speaking at once and discovering who is speaking when. The Factorial Hidden Markov Model (FHMM) (graphical model below) is a natural representation for this problem: now we identify each hidden chain with a speaker.



3. The Markov IBP

HDP + **Time** → **iHMM**
(Teh et al. 2006) (Beal et al. 2002)

IBP + **Time** → **??????**
(Griffiths & Ghahramani, 2005)

Step 1 – A Finite Model
We define the dynamics of a single binary Markov chain m

Each Markov chain

1. Starts in a dummy state $p(s_0 = 1) = 1$
2. Follows dynamics with transition matrix

$$W_m = \begin{pmatrix} 1-a_m & a_m \\ 1-b_m & b_m \end{pmatrix}$$

$a_m \sim \text{Beta}(\alpha/M, 1)$

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

Step 2 – Marginalizing Out the Parameters
We can marginalize out W_m and compute the probability of the hidden state sequence

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

where c_m^{01} is the number of times sequence S transitions from 0 -> 1, etc.

Step 3 – Taking the Infinite Limit
With a bit of algebra following [Griffiths & Ghahramani, 2005] we find

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

Comments:

- The distribution over S is Markov exchangeable

Inference:

- A stick breaking construction for the mIBP exists and allows us to combine slice sampling with dynamic programming for inference (Van Gael et al. 2008)
- α controls the number of features
- γ,δ control how fast features turn off

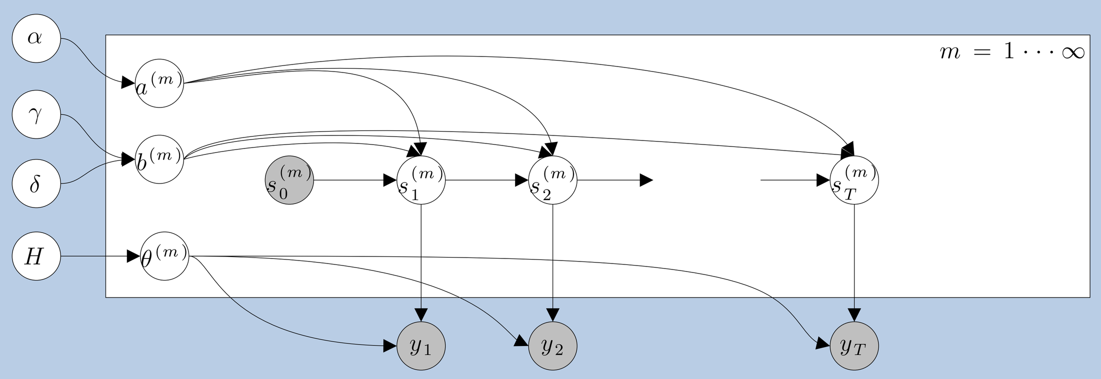
3. The Infinite Factorial HMM

The mIBP distribution defines a prior on a model with a potentially infinite number of Markov chains.

Start with the state sequence from the mIBP

- Add an observation model: $y_t \sim F(\boldsymbol{\theta}, \boldsymbol{s}_t)$
- Add a base distribution H: $\theta_m \sim H$

The iFHMM Graphical Model:



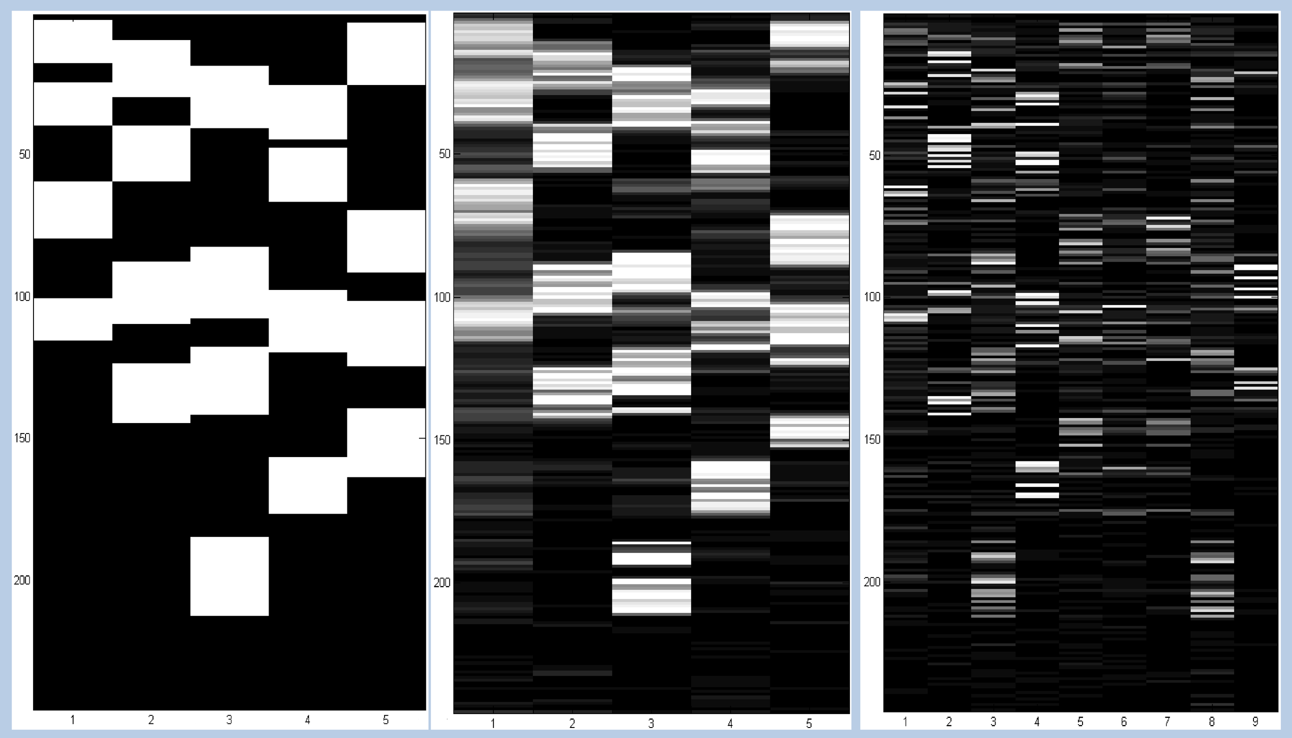
The Independent Component Analysis iFHMM
This model illustrates one particular choice of base measure H and observation model F which allows us to perform a non-parametric version of independent component analysis for time series.

$$\begin{aligned} S &\sim \text{mIBP}(\alpha,\gamma,\delta) \\ X_{tm} &\sim \text{Laplace}(0,1) \\ G_{md} &\sim \text{Normal}(0,\sigma_w^2) \\ Y &\sim \text{Normal}\left((S \odot X)G, \sigma_n^2\right) \end{aligned}$$

This is the time series equivalent of the iICA model of (Knowles et al. 2007).

4. Experiment

We recorded speech from 5 people using 10 microphones. The left plot shows time on the vertical axis, speakers on the horizontal axis. White denotes when a particular speaker is talking, black when he is silent. The middle plot shows a sample from the iFHMM solution: it has discovered both the number of speakers and a reasonably accurate segmentation. The right plot shows the solution found by infinite Independent Component Analysis.



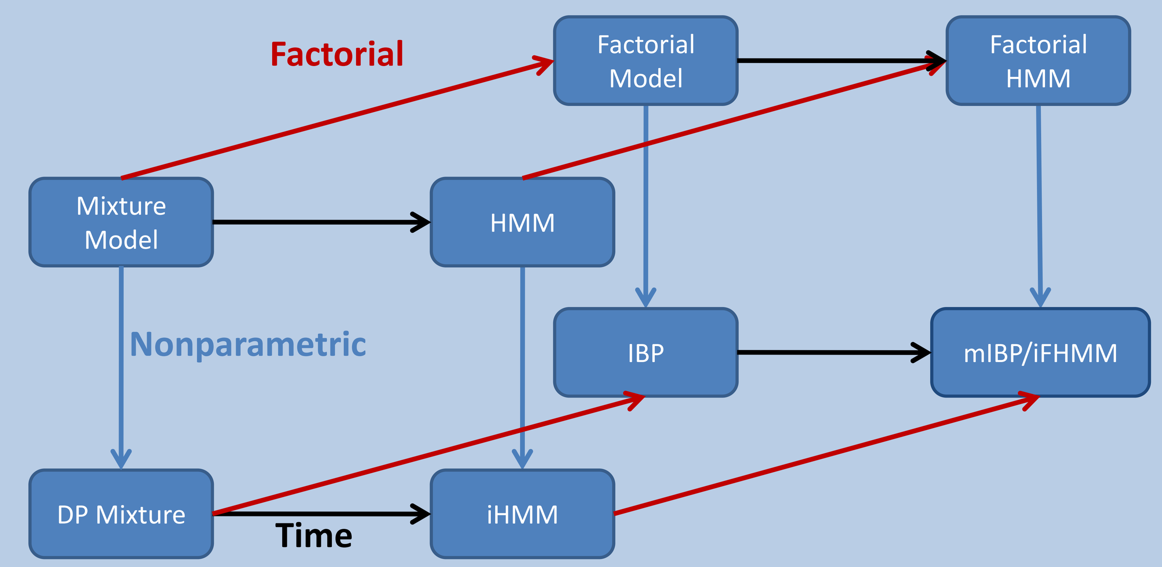
5. Discussion

Contributions:

- Novel nonparametric distribution over binary matrices with a Markov dependency between the rows
- An infinite capacity version of the factorial hidden Markov model
- Nonparametric independent component analysis

Future Work:

- More flexible parameterizations of the Markov dependency (K. Miller, personal communication)
- More than two states in the Markov chains



#Jurgen Van Gael is supported by a Microsoft Research Scholarship, * Yee Whye Teh is at the Gatsby unit, University College London