Infinite Factorial Infinite Hidden Markov Model

Isabel Valera, Francisco J. R. Ruiz, Fernando Perez-Cruz

University Carlos III in Madrid

Outline

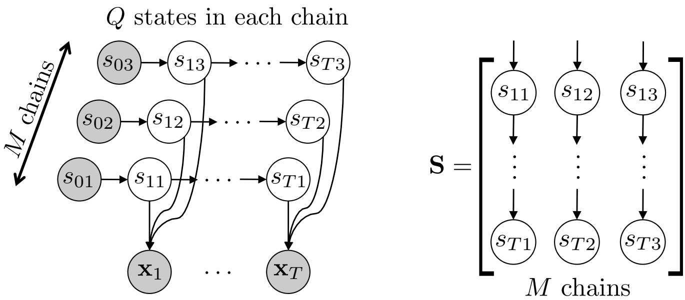
Outline

Introduction (1/2)

- Real-world time series analysis:
  - Trading in financial markets (many agents buying or selling shares, operating at different frequencies).
  - Psyquiatric patients (several mood/personality disorders evolving with time).
- Hidden Markov models (HMM) broadly applied to deal with time series.
- Limitations:
  - Not considering the complexity of the model.
  - Needing to prespecificate the model structure.
- Solutions:
  - Reversible jump MCMC.
  - Bayesian nonparametric (BNP) models, which allow an open-ended number of degrees of freedom.

Introduction (2/2)

- Our objective model takes the form of a factorial HMM.

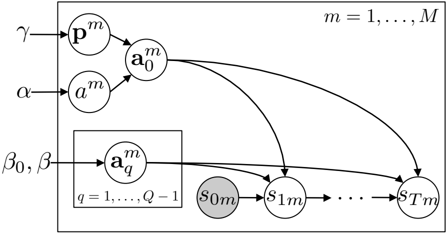


- Both Q and M are unknown.
  - M is inferred from standard BNP inference methods.
  - Q is inferred from techniques similar to reversible jump MCMC methods.

Outline

Nonbinary Infinite Factorial HMM (1/2)

- Finite version of the model (with a fixed number of states Q):



$$s_{tm} | s_{(t-1)m}, \mathbf{A}^m \sim \mathbf{a}^m_{s_{(t-1)m}}$$

$$\mathbf{a}_q^m | Q, \beta_0, \beta \sim \mathrm{Dirichlet}(\beta_0, \beta, \dots, \beta)$$

$$\mathbf{a}_0^m = \begin{bmatrix} a^m & (1-a^m)p_1^m & \dots & (1-a^m)p_{Q-1}^m \end{bmatrix}$$

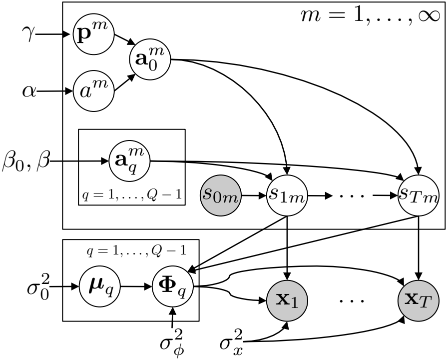
$$a^m | \alpha \sim \mathrm{Beta}\left(1, \frac{\alpha}{M}\right), \qquad \mathbf{p}^m | Q, \gamma \sim \mathrm{Dirichlet}(\gamma)$$

Nonbinary Infinite Factorial HMM (2/2)

- This model is reproducible when we make $M \to \infty$.
- We can compute $p([\mathbf{S}]|Q, \alpha, \beta_0, \beta, \gamma)$.
- Properties:
  - The columns of **S** (Markov chains) are exchangeable.
  - Elements in **S** (labels $\in \{0, 1, \ldots, Q-1\}$) are exchangeable.
  - Markov-exchangeable in the rows.
- Culinary metaphor:
  - T customers enter sequentially a restaurant.
  - Each of them can either take q units from dish m or not take dish m, with probabilities based on the previous customer's choice.
  - After that, the t-th customer tries $M_{new} \sim \mathrm{Poisson}(\alpha/T)$ new dishes.

Outline

Gaussian Observation Model (1/2)



- $\boldsymbol{\mu}_q$ is a Gaussian $1 \times D$ vector.
- $\boldsymbol{\Phi}_q$ is a Gaussian $M_+ \times D$ matrix, each row with mean $\boldsymbol{\mu}_q$.
- Observations $\mathbf{x}_t$ are Gaussian, with mean value depending on the additive contribution of all the active chains at instant t.

Gaussian Observation Model: Inference (2/2)

Gibbs sampling

- For $t = 1, \ldots, T$:
  - Sequentially sample each element s_{tm} from

$$p(s_{tm} = k | \mathbf{X}, \mathbf{S}_{\neg tm}, \mathcal{H}) \propto p(s_{tm} = k | \mathbf{S}_{\neg tm}, \mathcal{H}) p(\mathbf{X} | \mathbf{S}, \mathcal{H})$$

  - Draw M_{new} columns of $\mathbf{S}$ with states s_{tm} from a distribution where the prior is $\mathrm{Poisson}(M_{new} | \frac{\alpha}{T}) \times \frac{1}{(Q-1)^{M_{new}}}$.

Variational inference

- Valid for a finite (and large enough) number of chains M.
- Propose a distribution $q(\cdot)$ to approximate $p_M(\cdot | \mathbf{X}, \mathcal{H})$.
- Minimize the KL divergence from $q(\cdot)$ to $p_M(\cdot | \mathbf{X}, \mathcal{H})$.
- Optimize by iteratively applying a fixed-point set of equations. Involves a forward-backward algorithm for each chain.
- Approximate method, but much faster than sampling.

Outline

Prior on the Number of States

- The number of states Q is Poisson distributed:

$$Q|\lambda \sim 2 + \text{Poisson}(\lambda)$$

Inference

1. Update matrix $\mathbf{S}$ via Gibbs sampling for a fixed value of Q.
2. **Split** a component into two or **merge** two into one, then accept or reject the change.
3. **Birth** of a new state or **death** of an empty one, then accept or reject the change.

- M (number of chains) is updated in step 1.
- Q (number of states) is updated in steps 2-3.

Outline

Toy example

FHMM with 3 states and 2 chains:

- Transition probabilities:

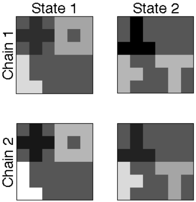
FROM \ TO	State 0	State 1	State 2
State 0	0.5964	0.1530	0.2506
State 1	0.2973	0.6738	0.0289
State 2	0.2463	0.2208	0.5329

(a) Chain 1

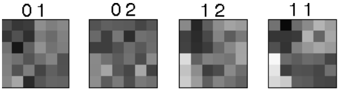
FROM \ TO	State 0	State 1	State 2
State 0	0.6321	0.0466	0.3213
State 1	0.3205	0.4947	0.1848
State 2	0.2413	0.1262	0.6325

(b) Chain 2

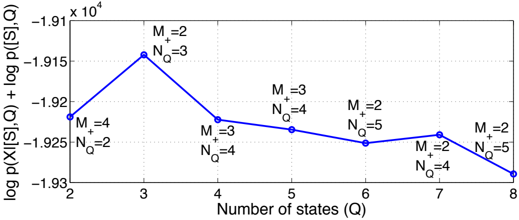
Observation model parameters ($\mathbf{\Phi}_q$):

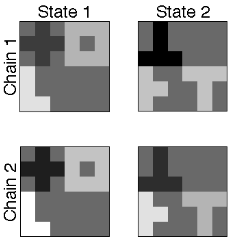


Observation examples:

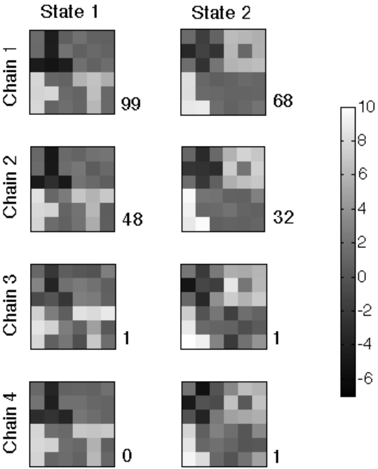


- Model selection based on the Gibbs sampler with fixed number of states (Q).
- The maximum of the log posterior $p(\mathbf{S}, Q|\mathbf{X})$ is the model with $Q = 3$ and $M_+ = 2$.

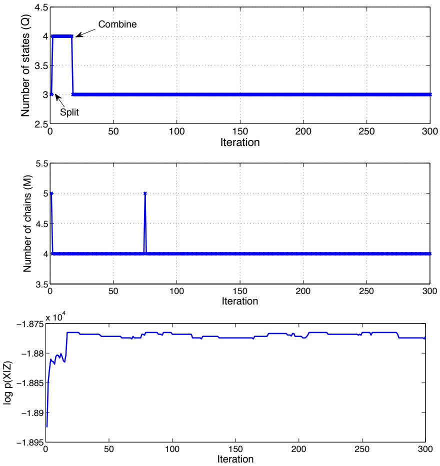




(a) The four base images.



(b) Mean of the base image inferred posterior probability.


Number of states (Q)
4.5
4
3.5
3
2.5
Combine
Split
0 50 100 150 200 250 300
Iteration
Number of chains (M)
5.5
5
4.5
4
3.5
0 50 100 150 200 250 300
Iteration
x 10^4
log p(X|Z)
−1.875
−1.88
−1.885
−1.89
−1.895
0 50 100 150 200 250 300
Iteration

Outline

Conclusions

- The binary infinite factorial HMM has been extended to its nonbinary version.
- We have developed a Gibbs sampling and a variational inference algorithm.
- We have developed an inference algorithm to learn both the number of chains M and the number of states Q of the factorial HMM.
- Two illustrative experiments have been included.