# Temporal Models

Anthony Gitter

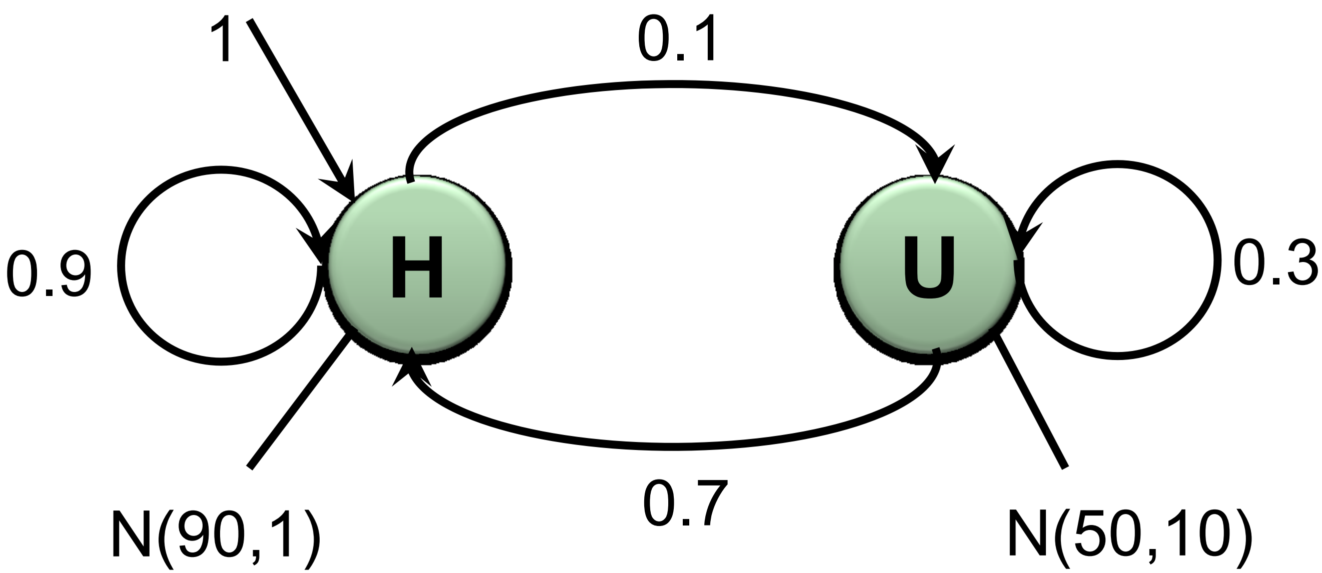
November 15, 2010

# How will we spend our final recitation?

- HMM review
- Generalizing HMMs
- PS5 questions

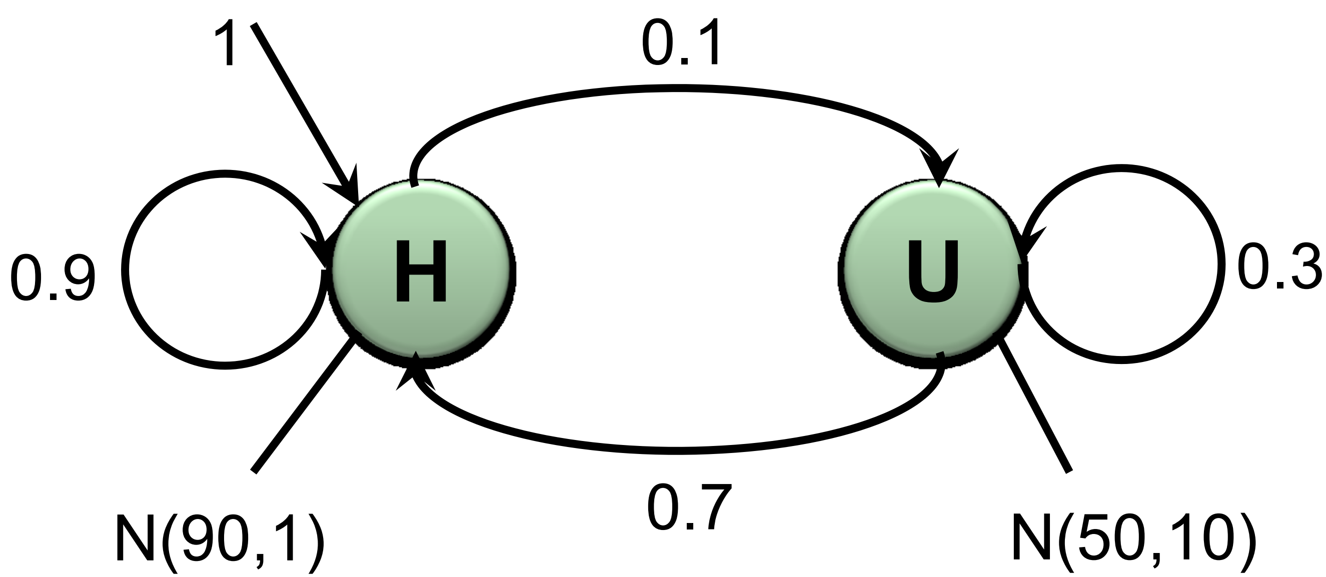
# Hidden Markov Models

- Often represented using state transition diagram
- Example for homework grades
- Parameters:
  - Initial state probabilities
  - Transition probabilities
  - Emission probabilities



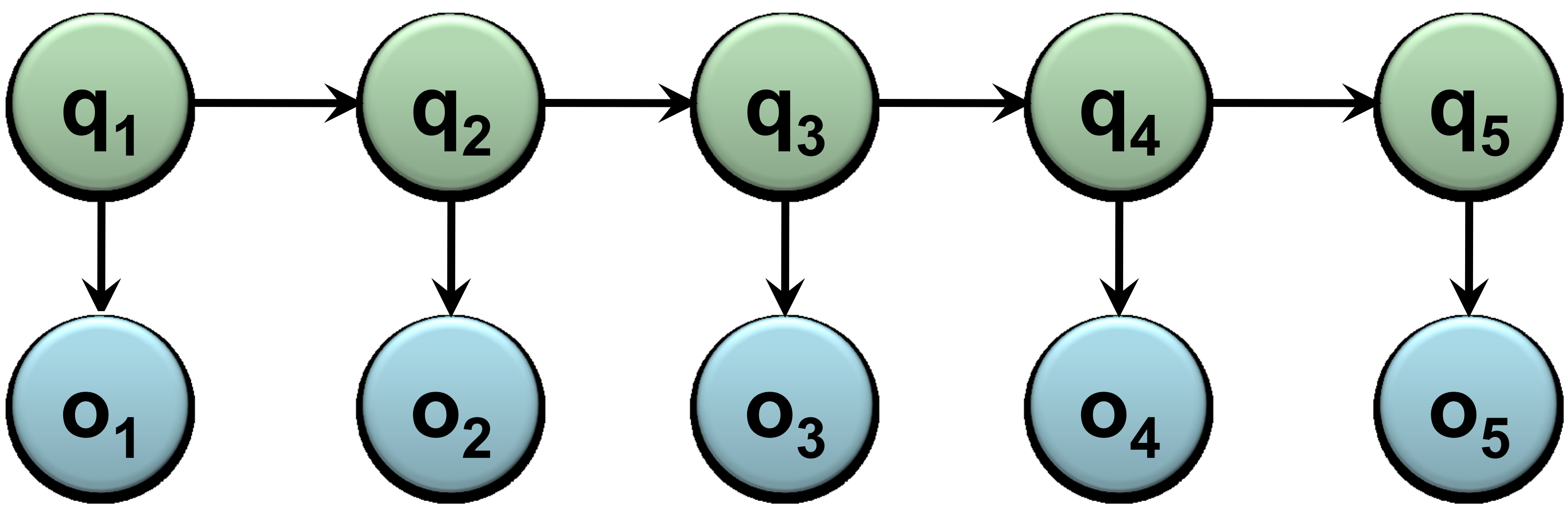
# Hidden Markov Models

- Markov assumption allows us to use this compact representation
- What are the nodes in this diagram?
- How many random variables?



# Unrolling HMMs

- We can "unroll" the HMM and explicitly show the variables
- At each of the 5 time points:
  - One binary variable for state (hidden)
  - One continuous variable for output (observed)

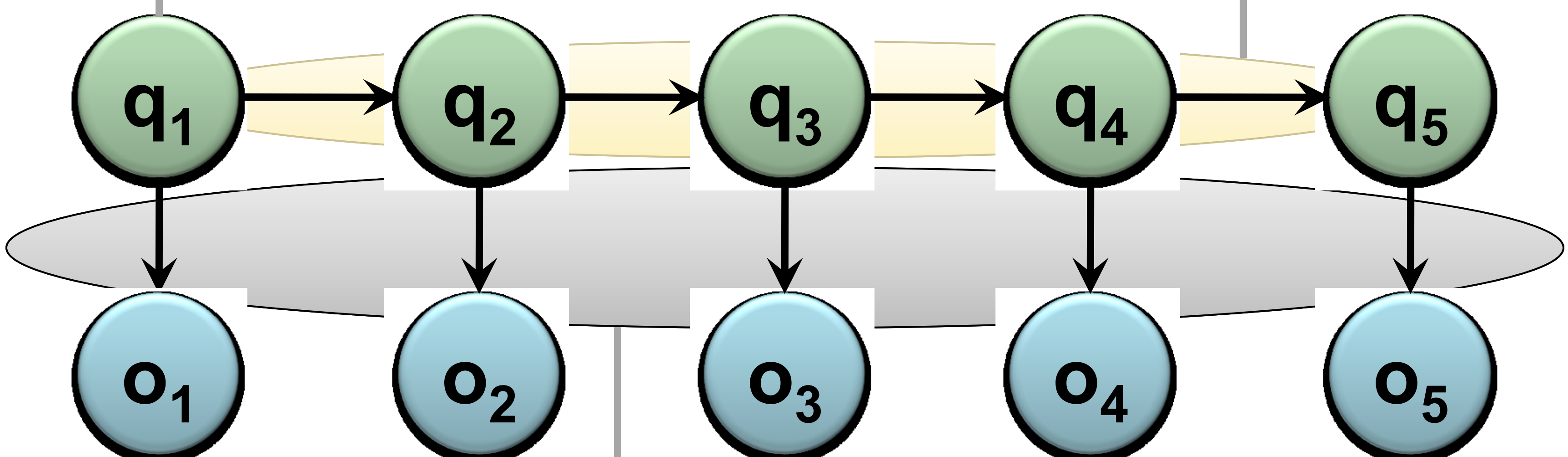


# Unrolling HMMs

- We still have shared transition and emission probabilities
- Use $q_t = 0$ to be state U and $q_t = 1$ to be state H

| $q_1 = 0$ | $q_1 = 1$ |
|---|---|
| 0 | 1 |

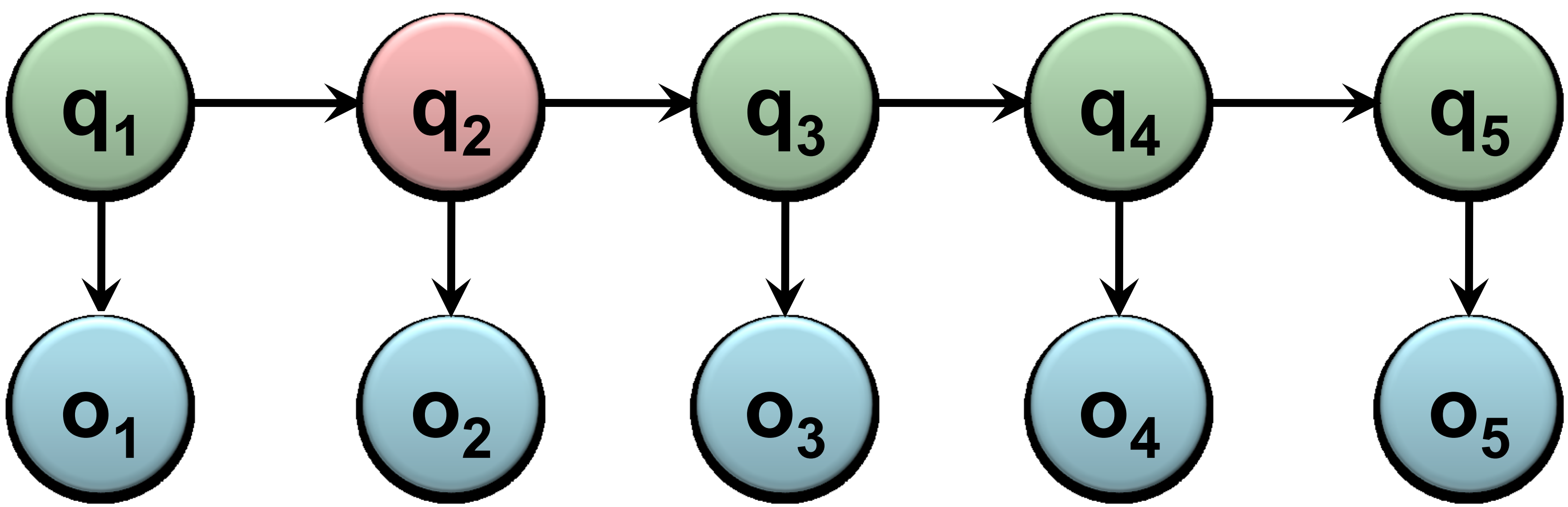
| | $q_{t-1} = 0$ | $q_{t-1} = 1$ |
|---|---|---|
| $q_t = 0$ | 0.3 | 0.1 |
| $q_t = 1$ | 0.7 | 0.9 |




| $q_t = 0$ | $q_t = 1$ |
|---|---|
| $N(o_t \mid 50,10)$ | $N(o_t \mid 90,1)$ |

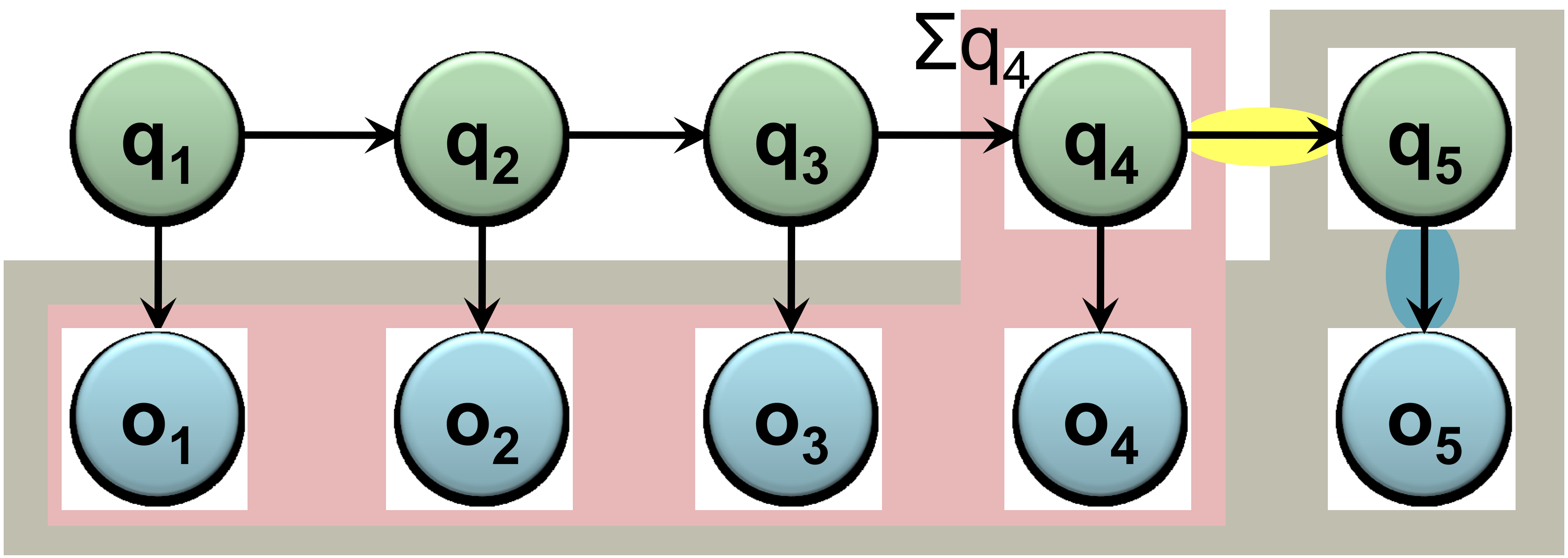
# HMM d-separation

- d-separation can be used to read independence assumptions
- $q_3 \perp q_1 \mid q_2$ (Markov assumption)
- $o_5 \not\perp o_3 \mid q_2$
- $o_5 \perp o_1 \mid q_2$



# HMM inference

- May be helpful to think about the terms we defined for HMM inference using this representation
- $\alpha_5(i=1) = P(o_1=81, o_2=97, o_3=92, o_4=44, o_5=88, q_5=1)$
  $= \Sigma_k\, b_1(o_5=88)\, a_{k,1}\, \alpha_4(k)$
  $= \Sigma_k\, P(o_5=88|q_5=1) P(q_5=1|q_4=k) P(o_1=81, o_2=97, o_3=92, o_4=44, q_4=k)$



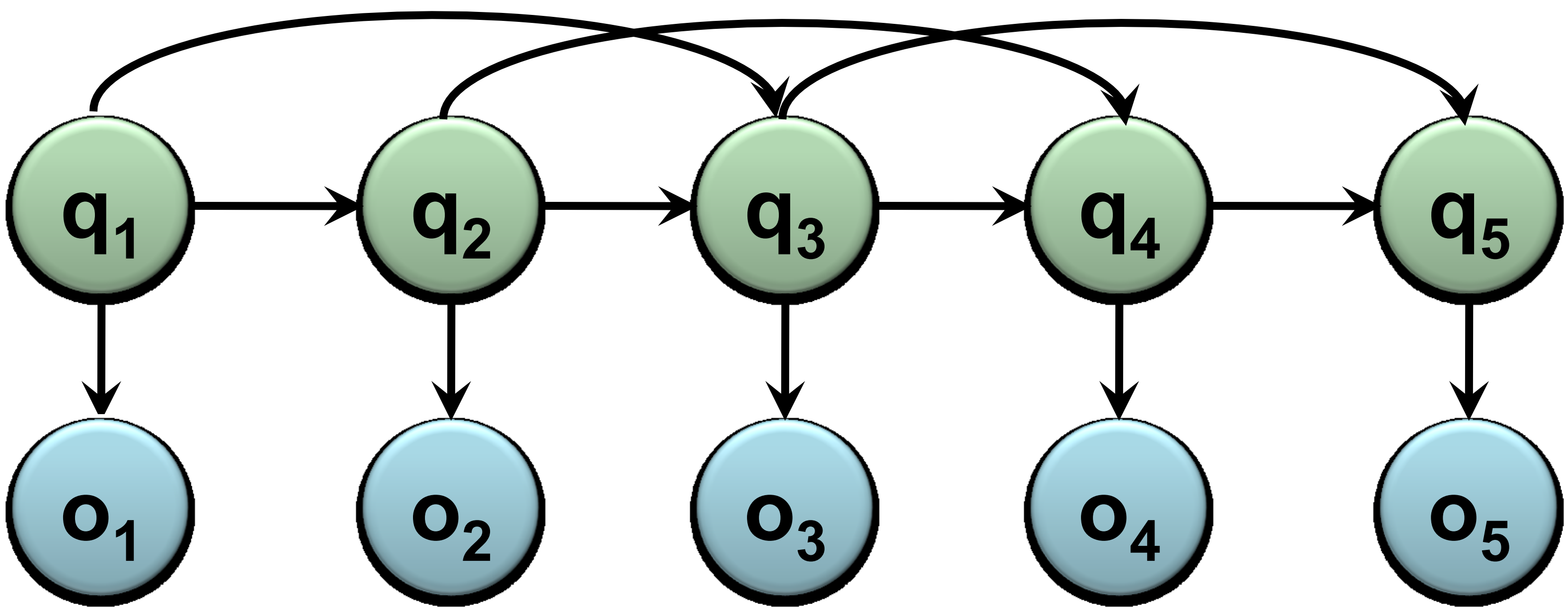
# HMM limitations

- In their simplest form HMMs make strong assumptions
  - State only depends on previous state
  - Discrete state variables
  - Output only depends on hidden state
- These assumptions can be helpful
  - Inference is relatively easy
  - Few parameters needed

- Sometimes these assumptions are too restrictive
- High level overviews of how assumptions are relaxed
- Inference and learning can be much more difficult for some of the following extensions

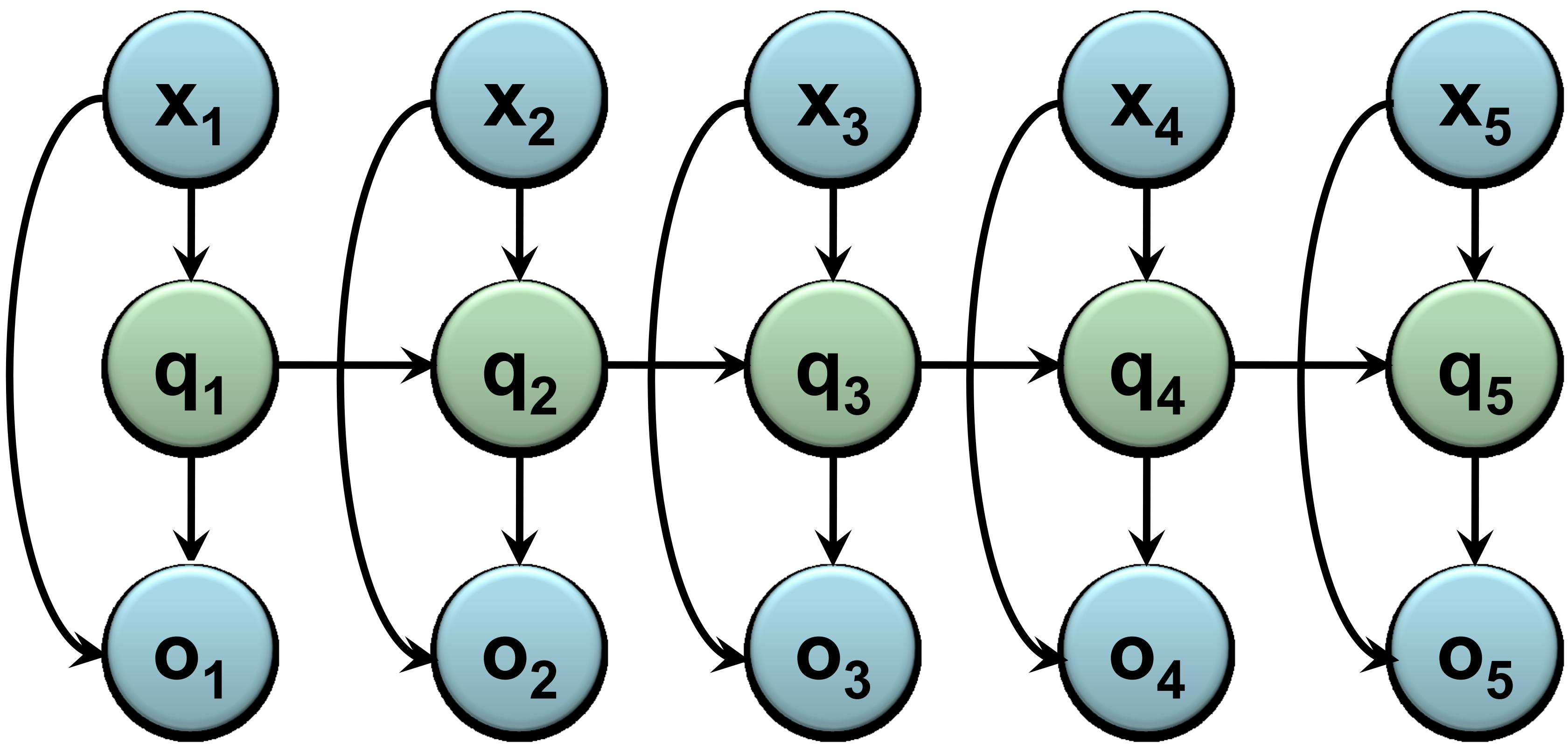
# Second order HMMs

- Hidden state depends on the two previous states
- Useful for natural language processing
- Can be extended to nth order HMM
  - State depends on n previous states



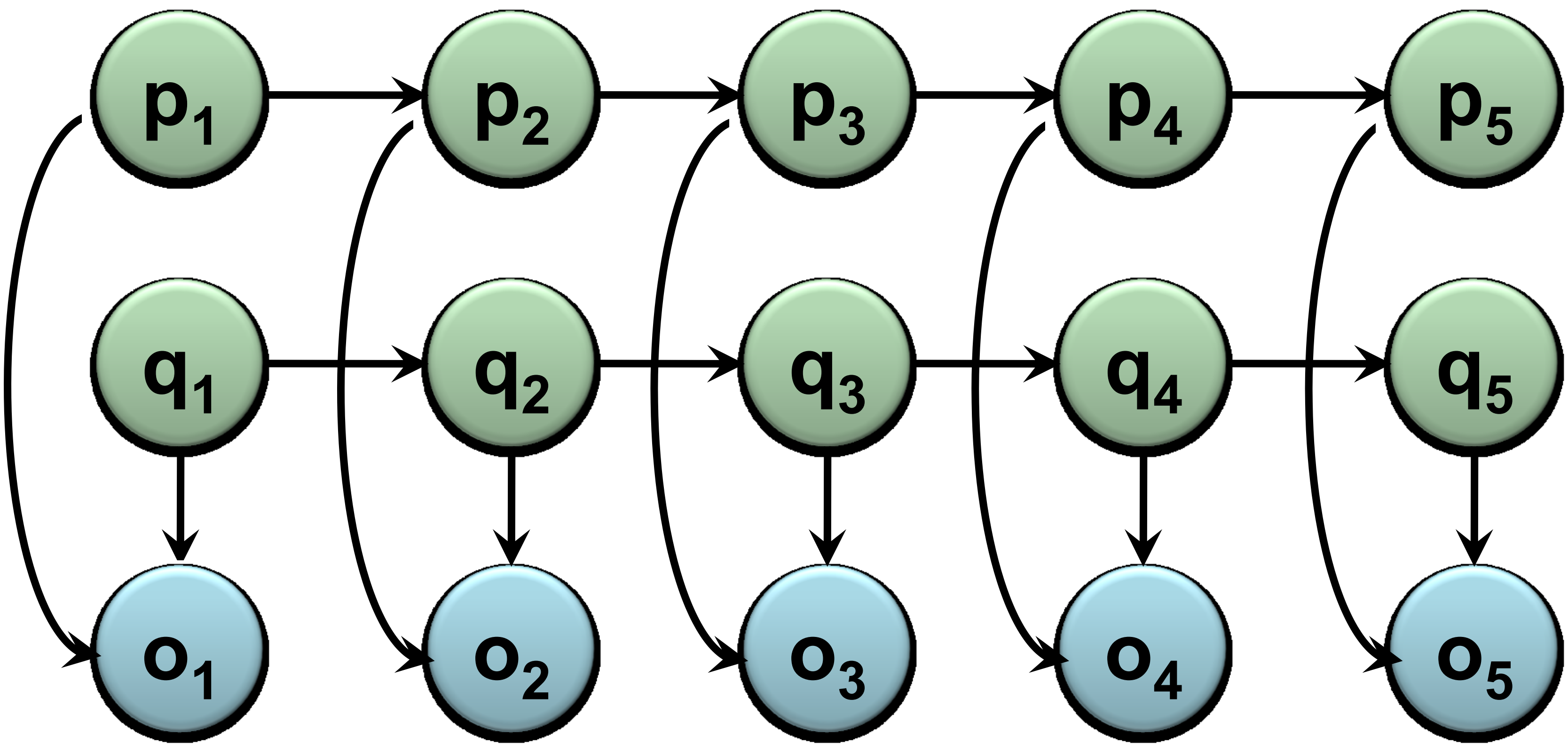
# Input-Output HMMs

- Hidden states and output depend on another observed sequence
- Still have $q_3 \perp q_1 \mid q_2$



# Factorial HMMs

- Using a single hidden variable for all hidden states would often lead to huge state space
- Instead use more than one chain of hidden variables
- Output depends on both hidden states



# Linear dynamical systems

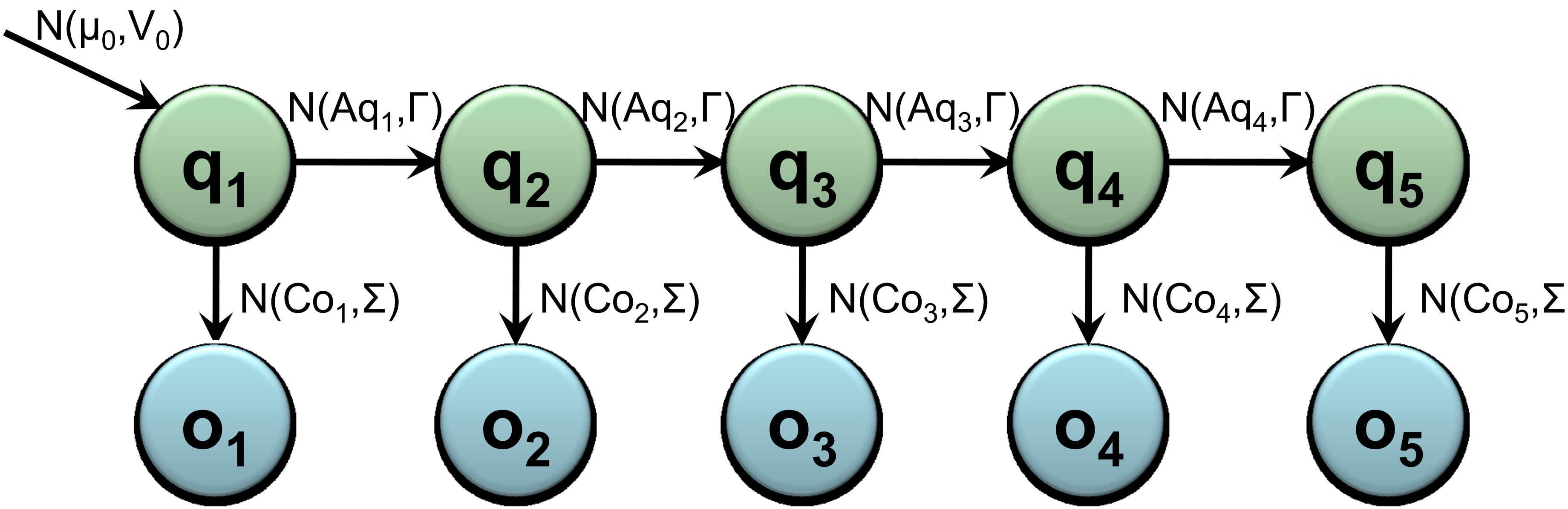
- Hidden states are multivariate Gaussian distributions
- State $q_t$ is linear function of state $q_{t-1}$ plus noise

$$q_1 = \mu_0 + u \qquad u \sim N(u \mid 0, V_0)$$

$$q_t = Aq_{t-1} + w_t \qquad w \sim N(w \mid 0, \Gamma)$$

$$o_t = Cq_t + v_t \qquad v \sim N(v \mid 0, \Sigma)$$

- A.k.a. Kalman filters



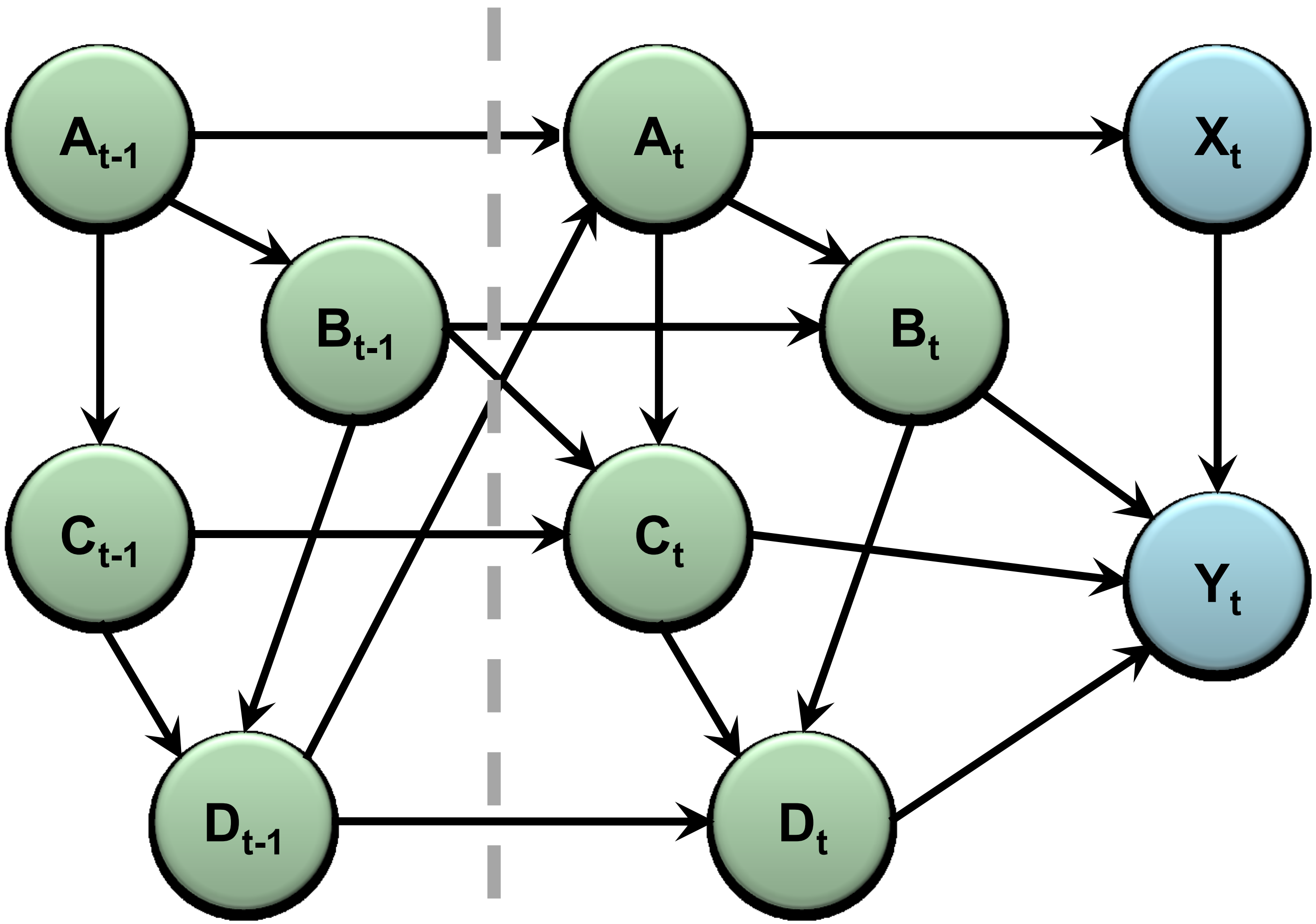
# Dynamic Bayesian networks

- All of the previous models are special cases of dynamic Bayesian networks (DBNs)
- At each time point
  - Set of hidden variables
  - Set of observed variables
- Variables can be discrete or continuous
- Two-slice temporal Bayesian network defines the structure and distributions

- Kevin Murphy’s DBN tutorial for much more detail

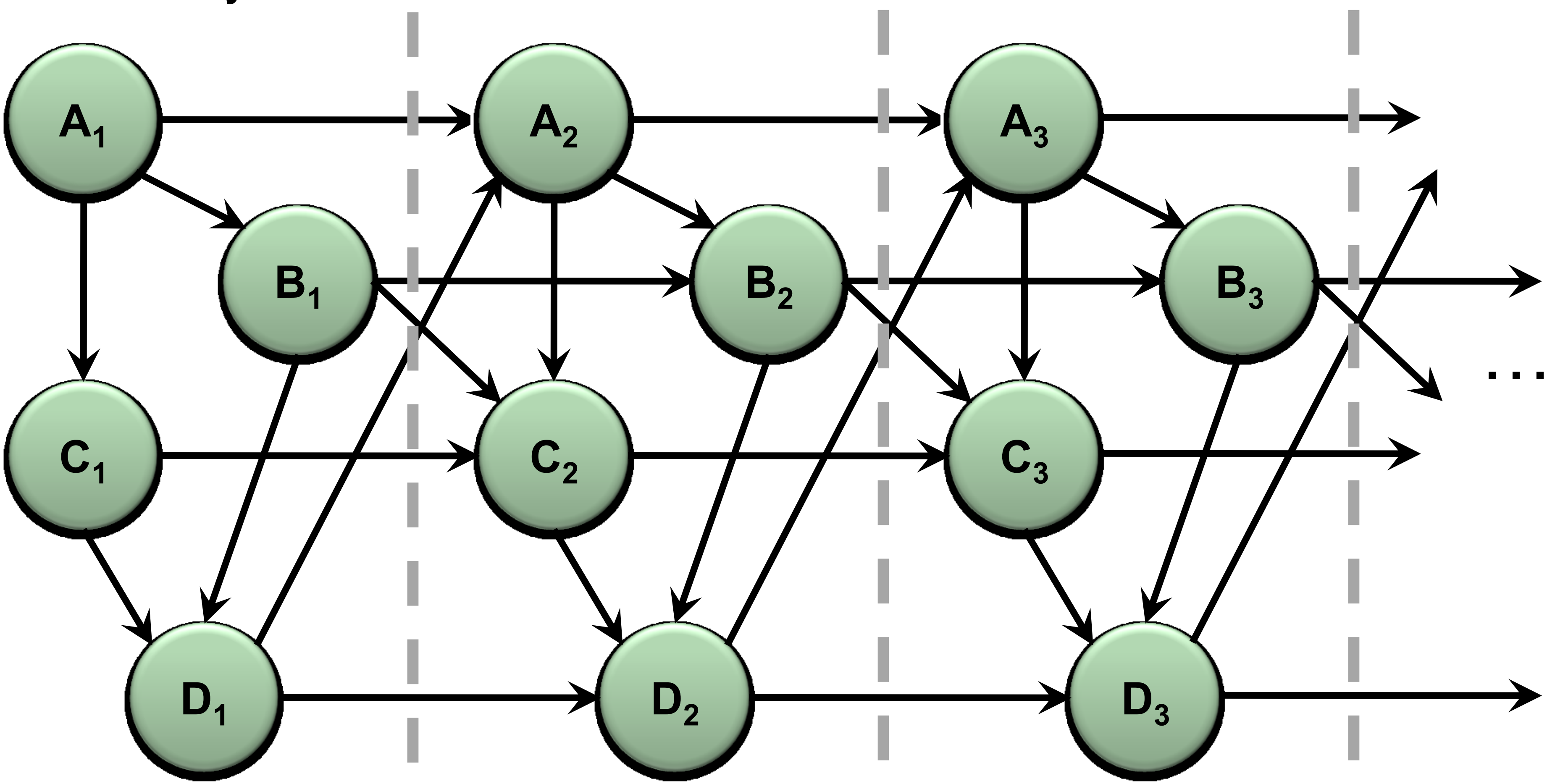
# Two-slice temporal Bayesian network

- Hidden variables at time t-1 and t
- Observed variables at time t



# Two-slice temporal Bayesian network

- Structure and distributions hold between all consecutive time points
- Only hidden nodes shown here



# PS5

- In 5.1 show the state transition diagram not the unrolled probabilistic graphical model

- Any questions?