Artificial Intelligence

Markov processes and Hidden Markov Models (HMMs)

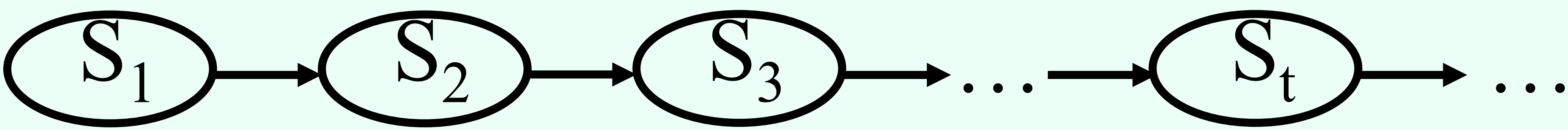
Instructor: Vincent Conitzer

Motivation

- The Bayes nets we considered so far were static: they referred to a single point in time
  - E.g., medical diagnosis
- Agent needs to model how the world evolves
  - Speech recognition software needs to process speech over time
  - Artificially intelligent software assistant needs to keep track of user's intentions over time
  - … … …

Markov processes

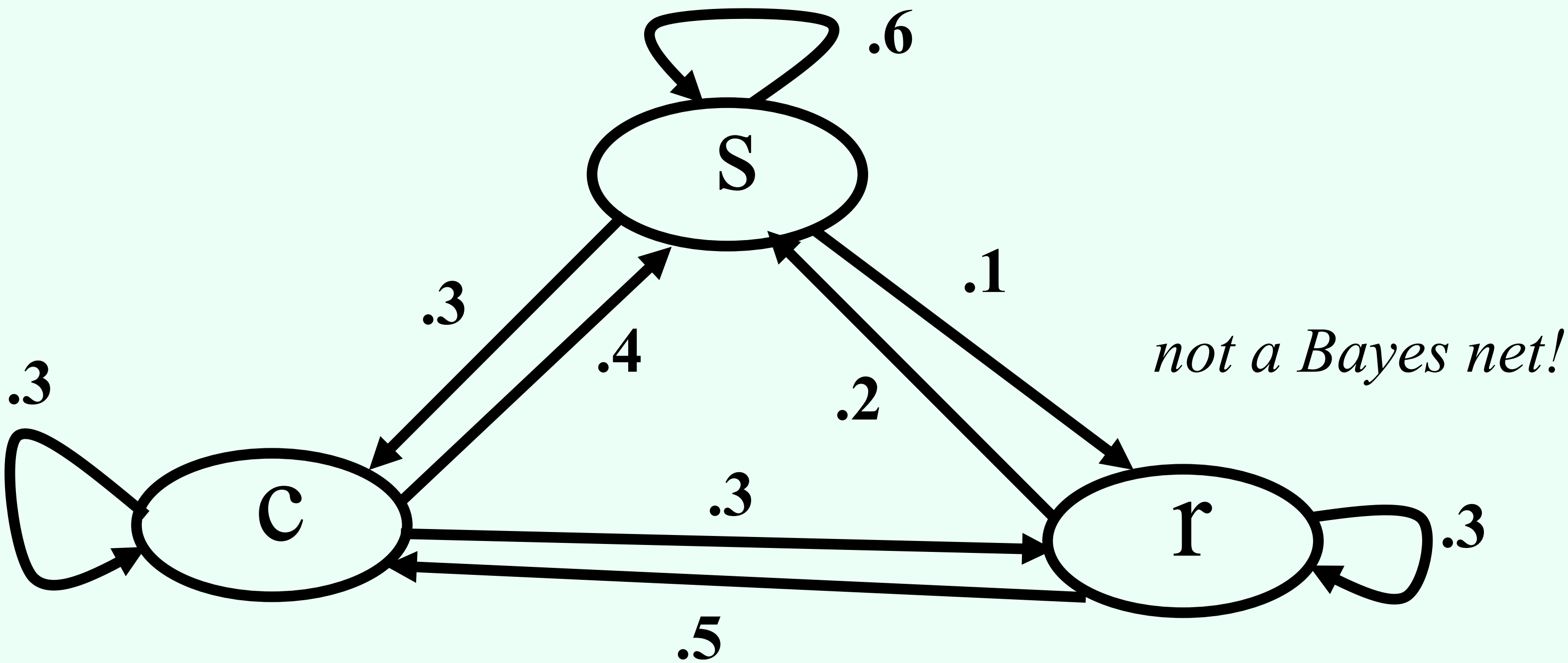
- We have time periods t = 0, 1, 2, …
- In each period t, the world is in a certain state S_t
- The Markov assumption: given S_t, S_{t+1} is independent of all S_i with $i < t$
  - $P(S_{t+1} \mid S_1, S_2, \ldots, S_t) = P(S_{t+1} \mid S_t)$
  - Given the current state, history tells us nothing more about the future



- Typically, all the CPTs are the same:
- For all t, $P(S_{t+1} = j \mid S_t = i) = a_{ij}$ (stationarity assumption)

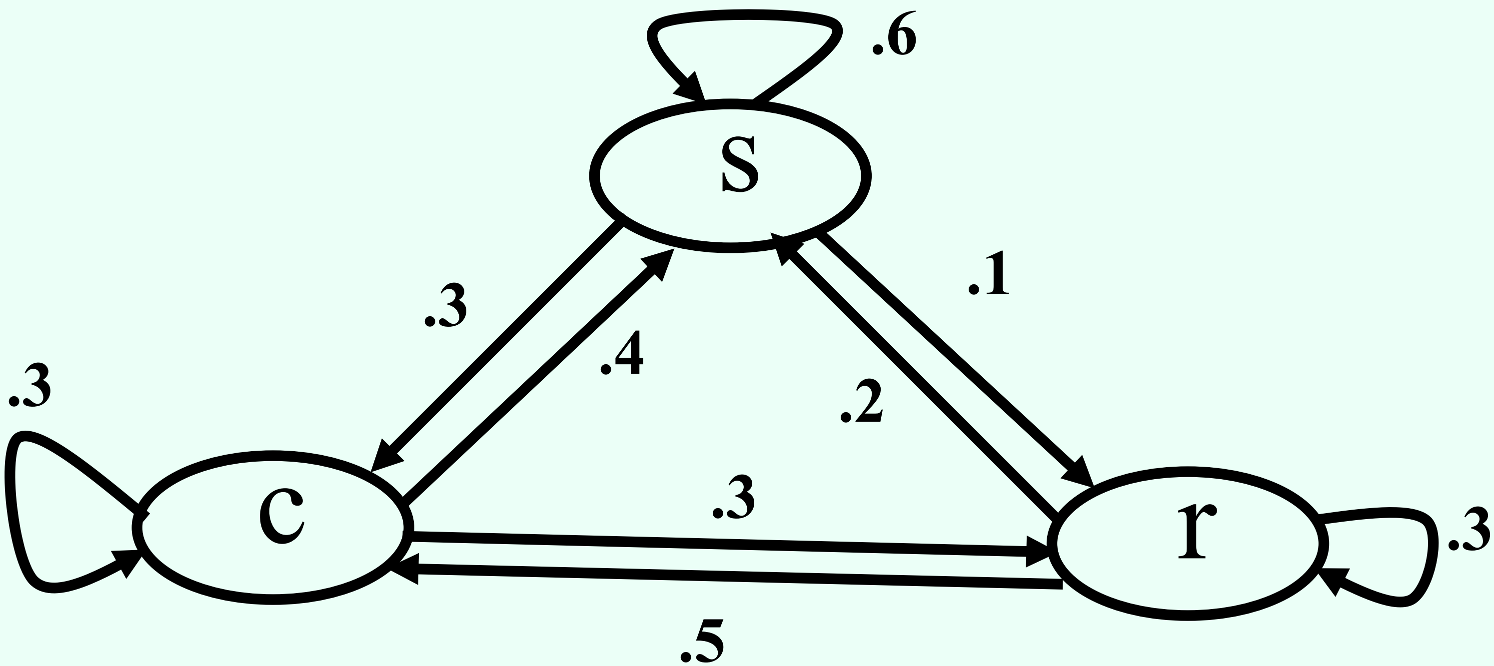
Weather example

- S_t is one of {s, c, r} (sun, cloudy, rain)
- Transition probabilities:



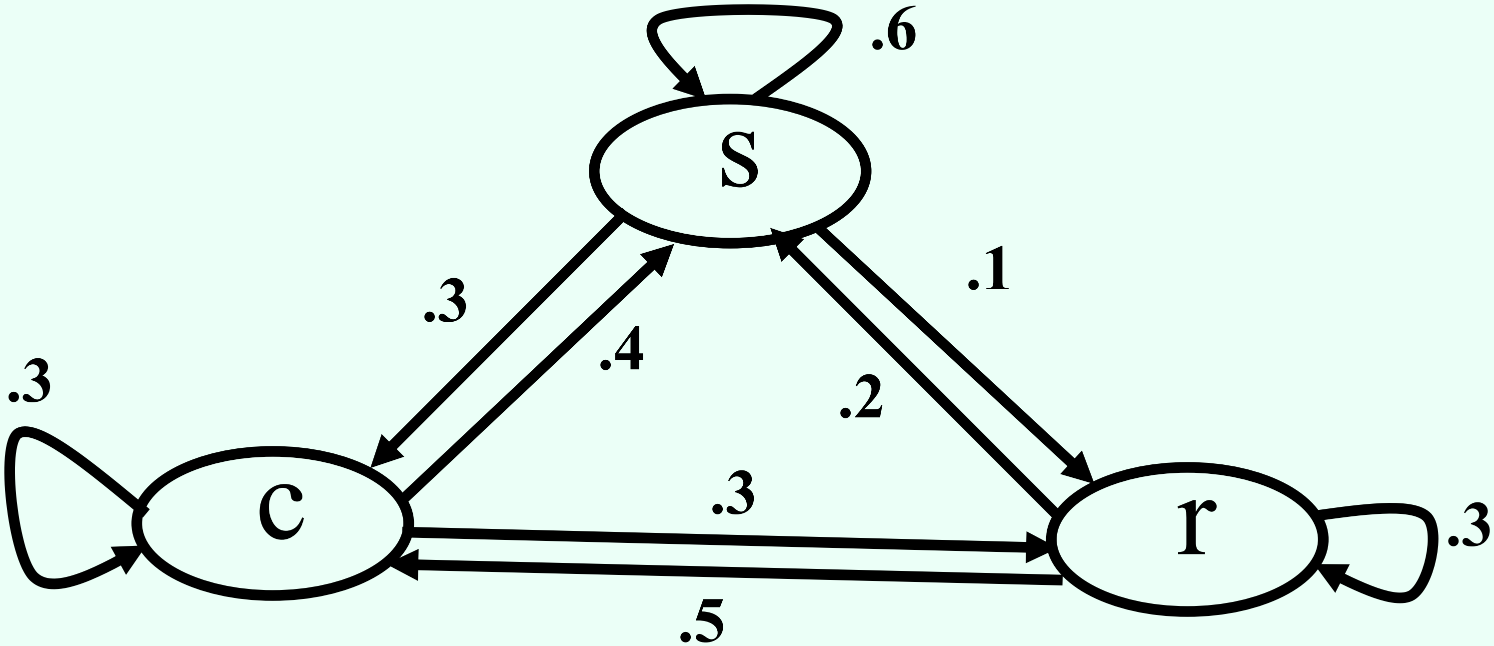
- Also need to specify an initial distribution $P(S_0)$
- Throughout, assume $P(S_0 = s) = 1$

Weather example…



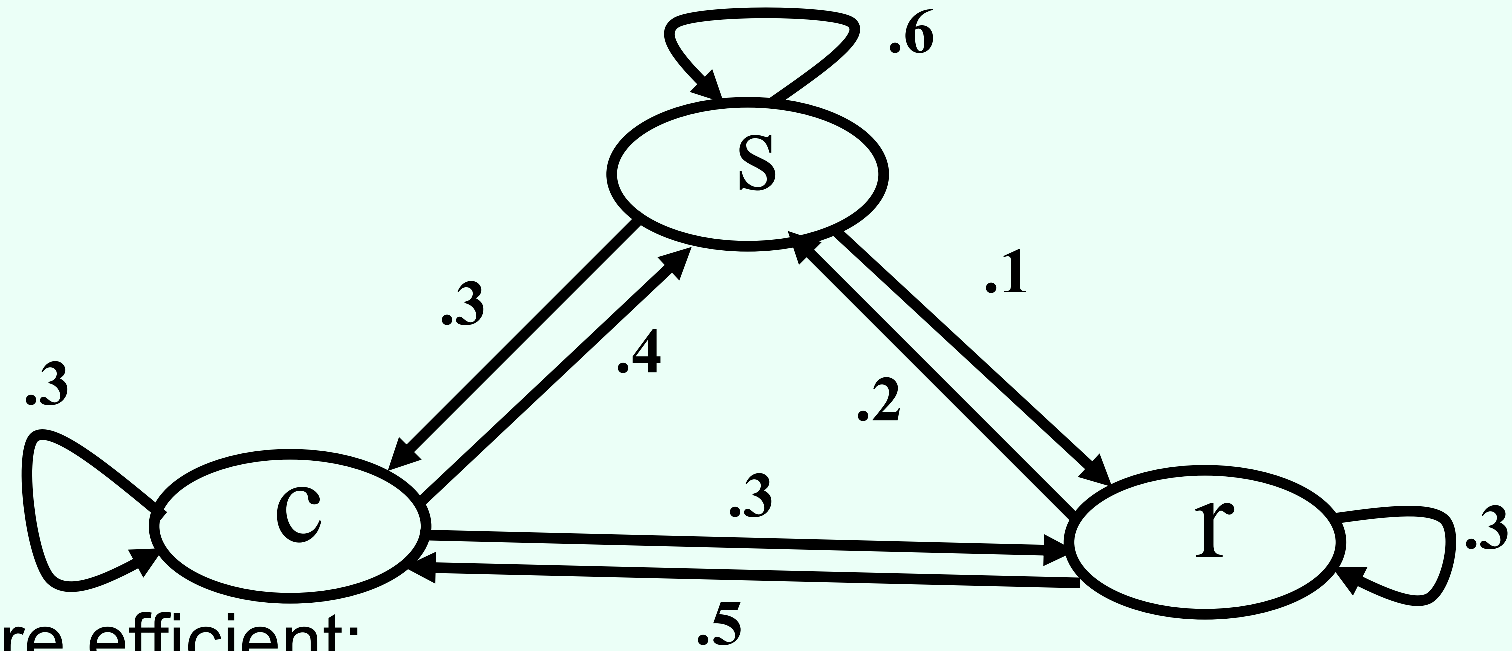
- What is the probability that it rains two days from now? $P(S_2 = r)$
- $P(S_2 = r) = P(S_2 = r, S_1 = r) + P(S_2 = r, S_1 = s) + P(S_2 = r, S_1 = c) = .1*.3 + .6*.1 + .3*.3 = .18$

Weather example…



- What is the probability that it rains **three** days from now?
- Computationally inefficient way: $P(S_3 = r) = P(S_3 = r, S_2 = r, S_1 = r) + P(S_3 = r, S_2 = r, S_1 = s) + \ldots$
- For n periods into the future, need to sum over 3^{n-1} paths

Weather example…



- More efficient:
- $P(S_3 = r) = P(S_3 = r, S_2 = r) + P(S_3 = r, S_2 = s) + P(S_3 = r, S_2 = c) = P(S_3 = r \mid S_2 = r)P(S_2 = r) + P(S_3 = r \mid S_2 = s)P(S_2 = s) + P(S_3 = r \mid S_2 = c)P(S_2 = c)$
- Only hard part: figure out $P(S_2)$
- Main idea: compute distribution $P(S_1)$, then $P(S_2)$, then $P(S_3)$
- Linear in number of periods!

example on board

Stationary distributions

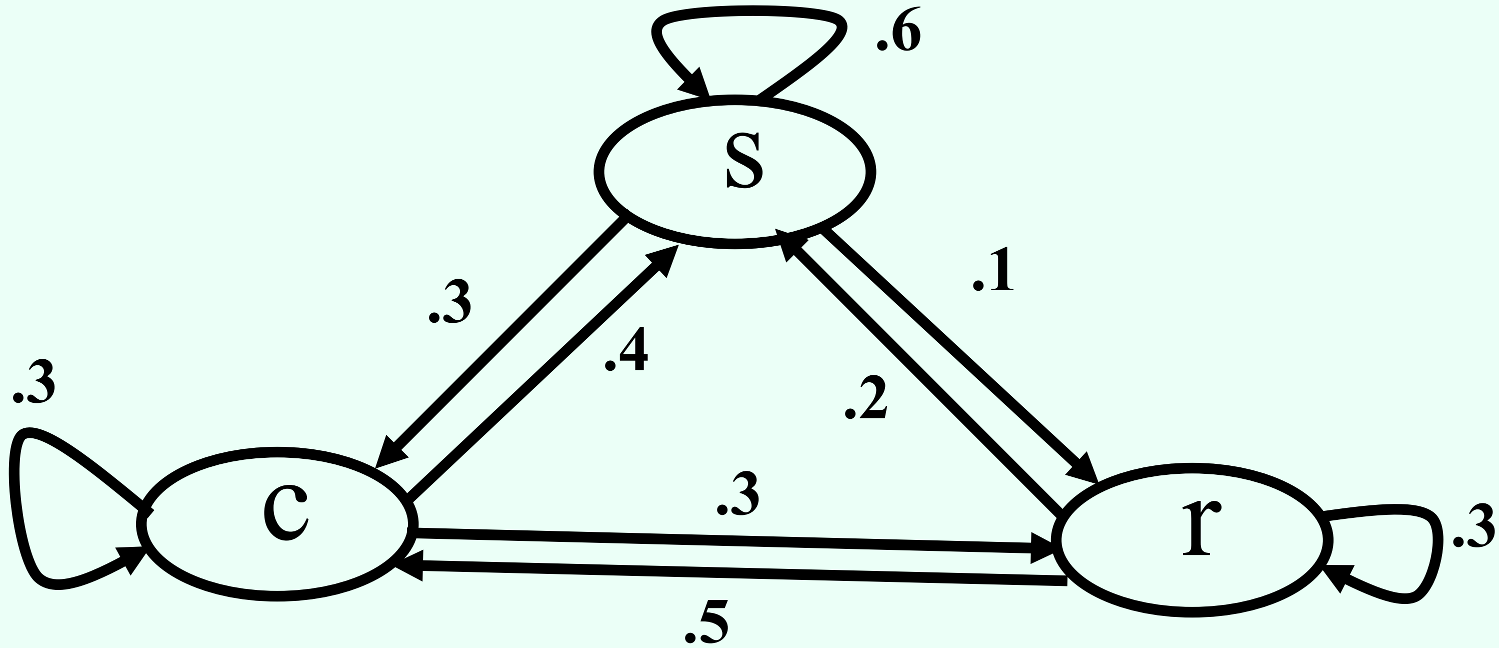
- As t goes to infinity, “generally,” the distribution $P(S_t)$ will converge to a stationary distribution
- A distribution given by probabilities π_i (where i is a state) is stationary if:

  $P(S_t = i) = \pi_i$ means that $P(S_{t+1} = i) = \pi_i$
- Of course,

  $P(S_{t+1} = i) = \Sigma_j P(S_{t+1} = i, S_t = j) = \Sigma_j P(S_t = j)\, a_{ji}$
- So, stationary distribution is defined by

  $\pi_i = \Sigma_j \pi_j a_{ji}$

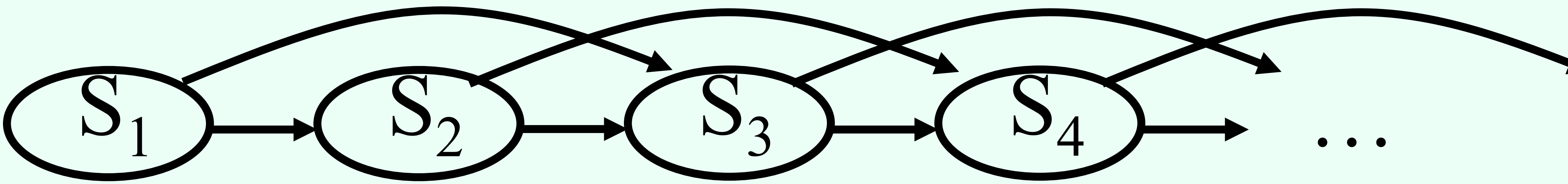
Computing the stationary distribution



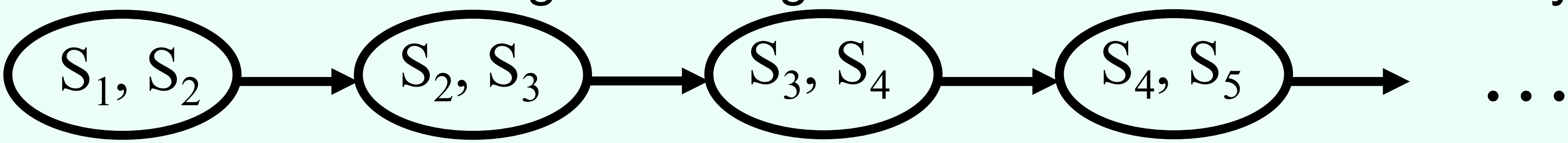
- $\pi_s = .6\pi_s + .4\pi_c + .2\pi_r$
- $\pi_c = .3\pi_s + .3\pi_c + .5\pi_r$
- $\pi_r = .1\pi_s + .3\pi_c + .3\pi_r$

Restrictiveness of Markov models

- Are past and future really independent given current state?
- E.g., suppose that when it rains, it rains for at most 2 days



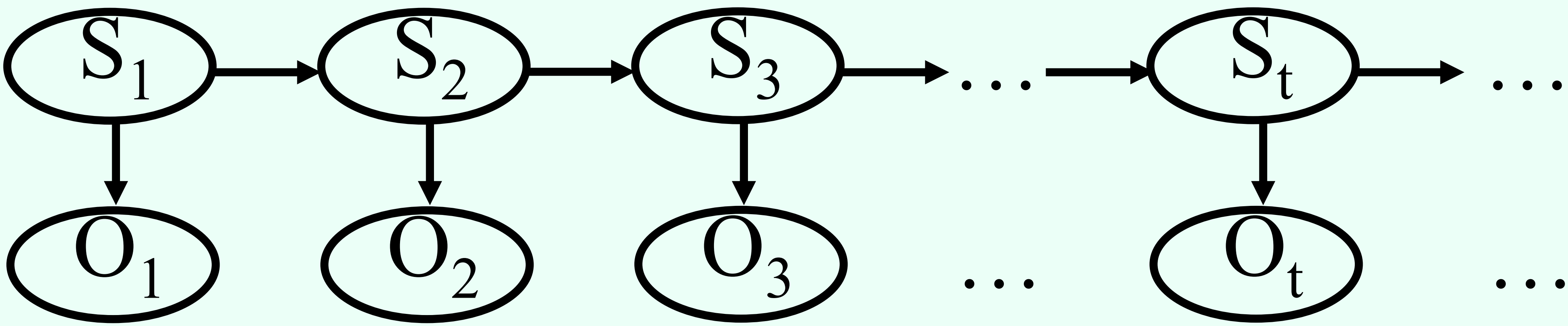
- Second-order Markov process
- Workaround: change meaning of “state” to events of last 2 days



- Another approach: add more information to the state
- E.g., the full state of the world would include whether the sky is full of water
  - Additional information may not be observable
  - Blowup of number of states…

Hidden Markov models (HMMs)

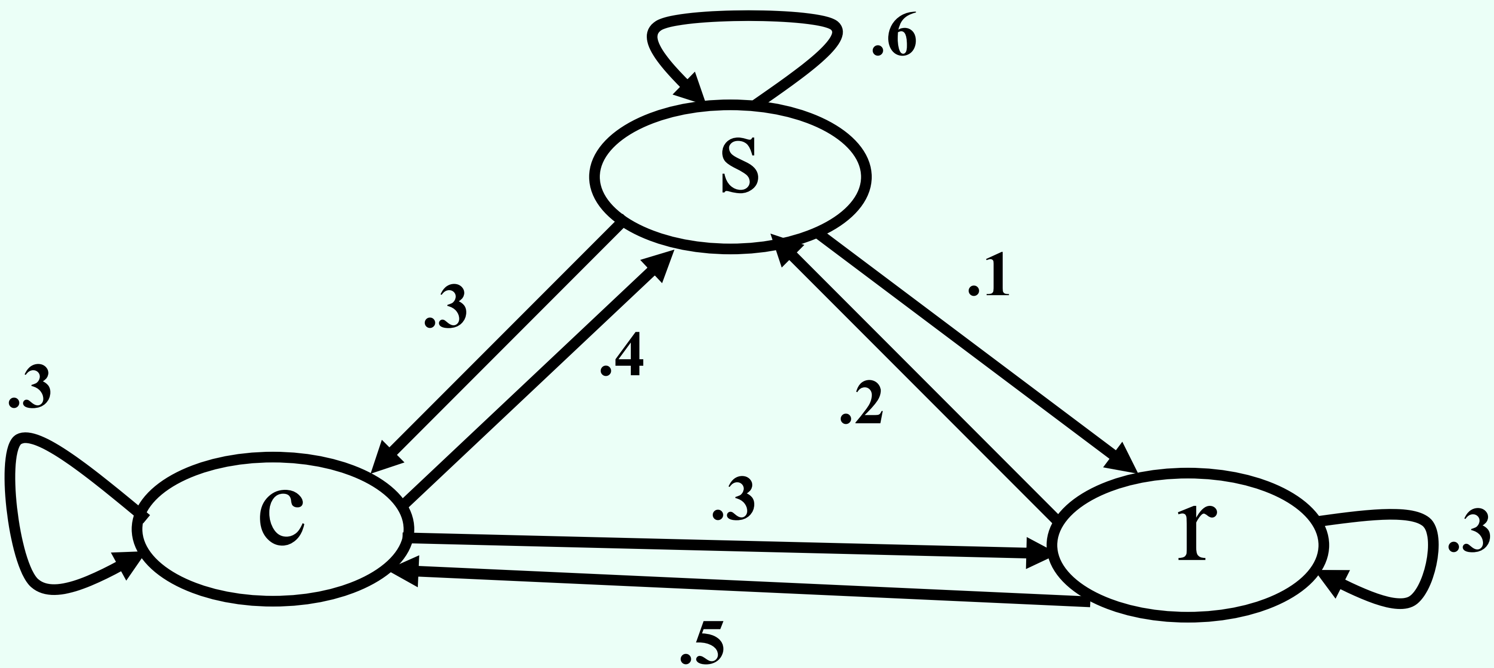
- Same as Markov model, except we cannot see the state
- Instead, we only see an observation each period, which depends on the current state



- Still need a transition model: $P(S_{t+1} = j \mid S_t = i) = a_{ij}$
- Also need an observation model: $P(O_t = k \mid S_t = i) = b_{ik}$

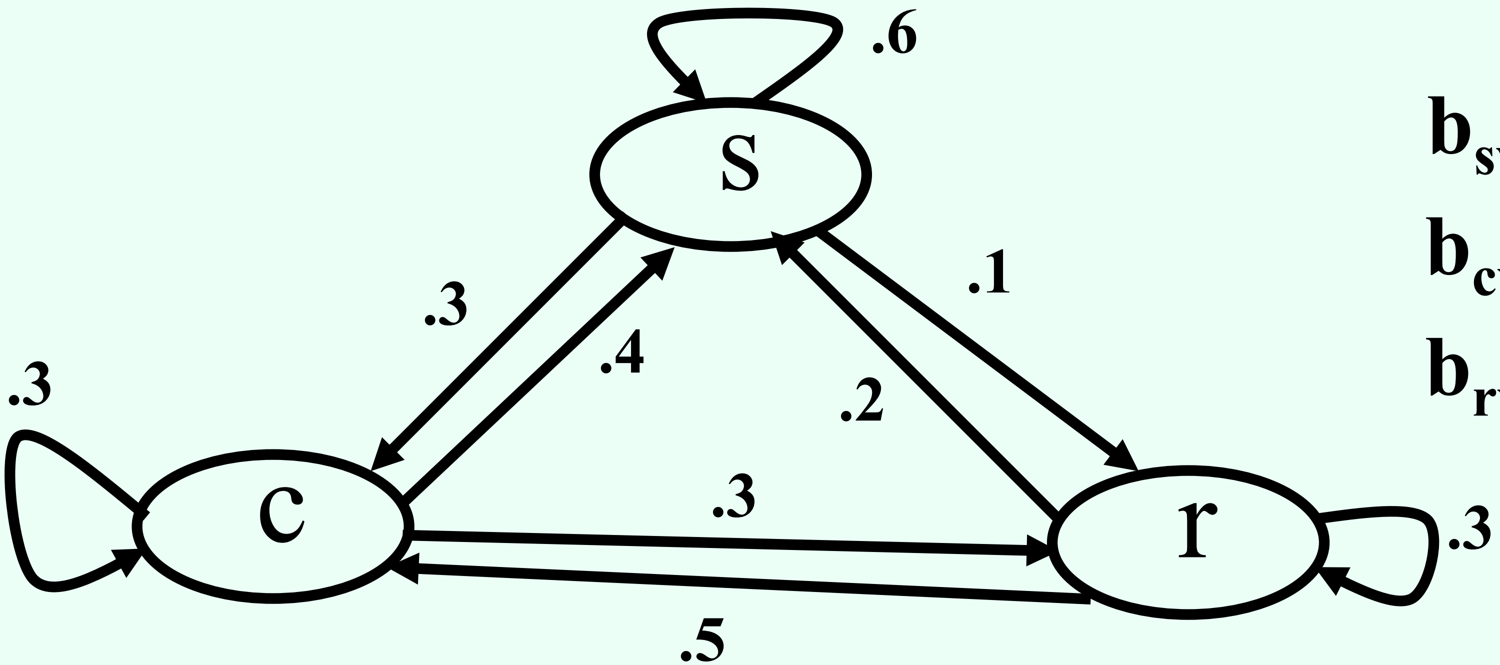
Weather example extended to HMM

- Transition probabilities:



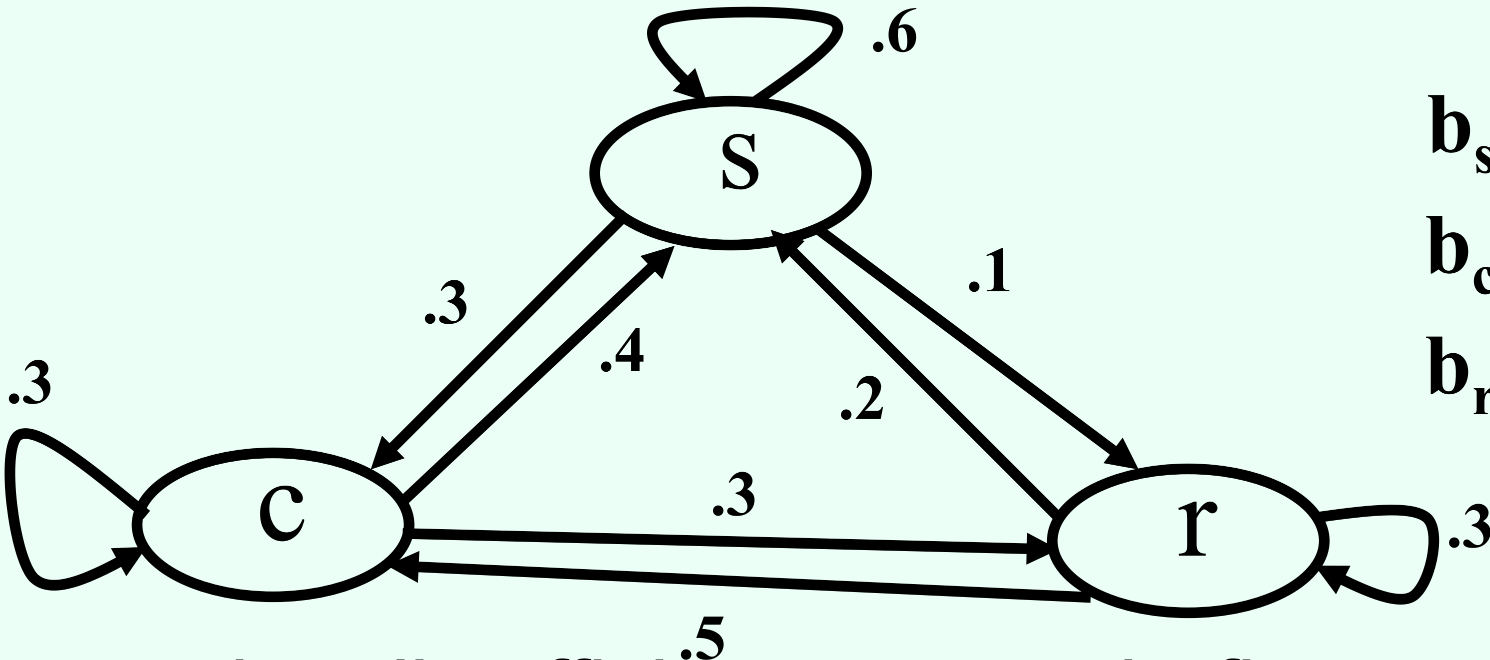
- Observation: labmate wet or dry
- $b_{sw} = .1$, $b_{cw} = .3$, $b_{rw} = .8$

HMM weather example: a question



- You have been stuck in the lab for three days (!)
- On those days, your labmate was dry, wet, wet, respectively
- What is the probability that it is now raining outside?
- $P(S_2 = r \mid O_0 = d, O_1 = w, O_2 = w)$
- By Bayes' rule, really want to know $P(S_2, O_0 = d, O_1 = w, O_2 = w)$

Solving the question



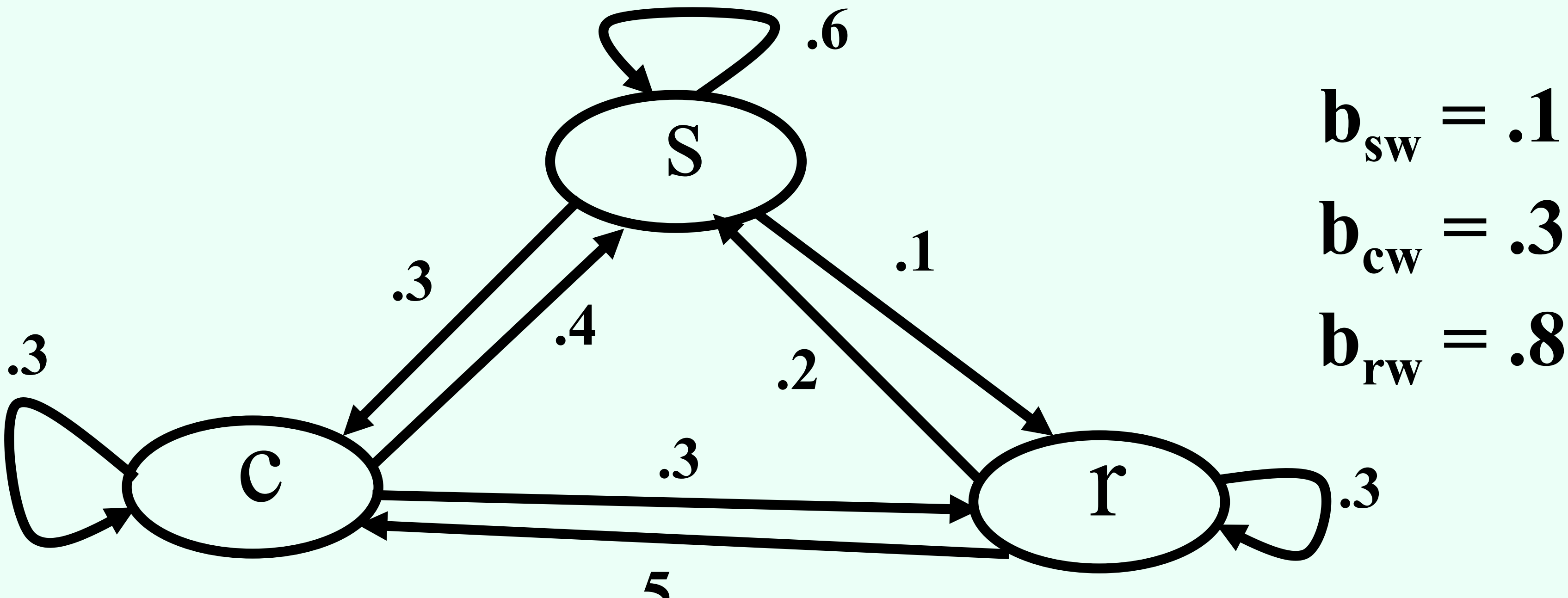
$b_{sw} = .1$

$b_{cw} = .3$

$b_{rw} = .8$

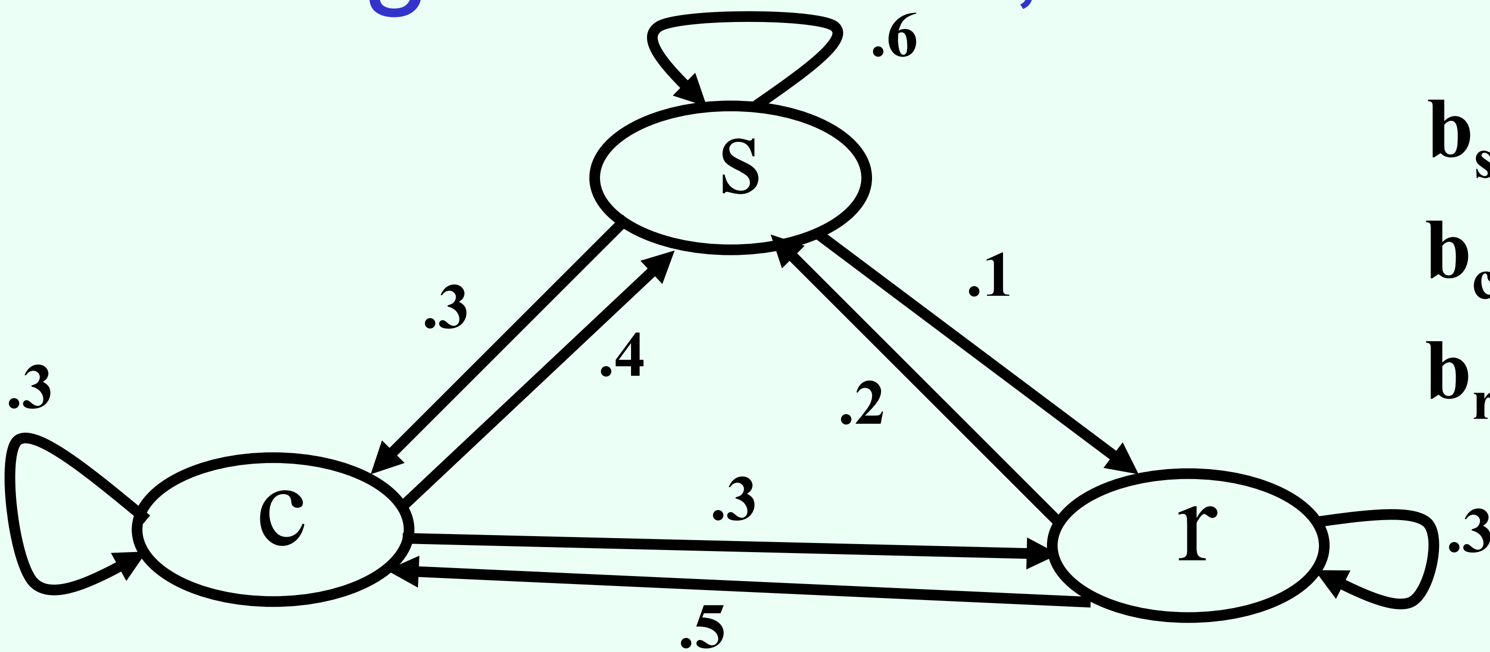
- Computationally efficient approach: first compute $P(S_1 = i, O_0 = d, O_1 = w)$ for all states i
- General case: solve for $P(S_t, O_0 = o_0, O_1 = o_1, \ldots, O_t = o_t)$ for t=1, then t=2, … This is called monitoring
- $P(S_t, O_0 = o_0, O_1 = o_1, \ldots, O_t = o_t) = \Sigma_{s_{t-1}} P(S_{t-1} = s_{t-1}, O_0 = o_0, O_1 = o_1, \ldots, O_{t-1} = o_{t-1})\, P(S_t \mid S_{t-1} = s_{t-1})\, P(O_t = o_t \mid S_t)$

Predicting further out



- You have been stuck in the lab for three days
- On those days, your labmate was dry, wet, wet, respectively
- What is the probability that **two days from now** it will be raining outside?
- $P(S_4 = r \mid O_0 = d, O_1 = w, O_2 = w)$

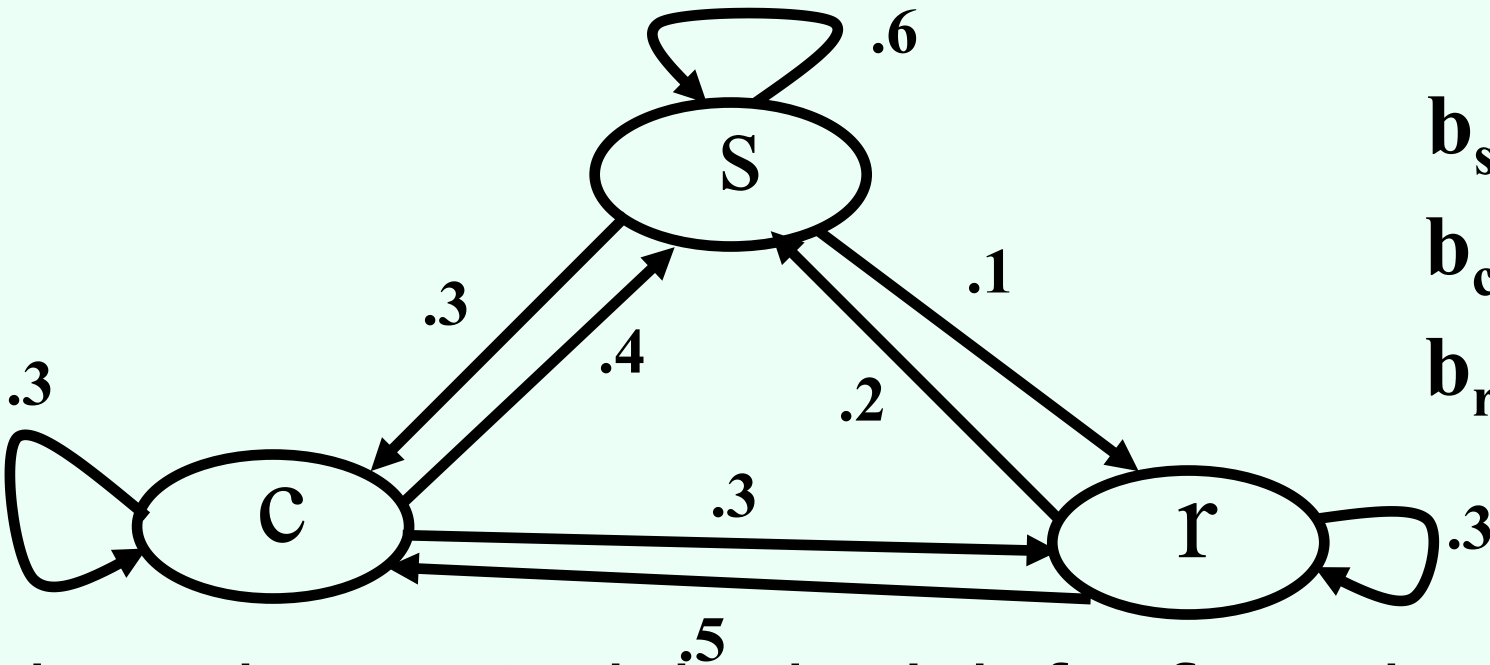
Predicting further out, continued…



$b_{sw} = .1$
$b_{cw} = .3$
$b_{rw} = .8$

- Want to know: $P(S_4 = r \mid O_0 = d, O_1 = w, O_2 = w)$
- Already know how to get: $P(S_2 \mid O_0 = d, O_1 = w, O_2 = w)$
- $P(S_3 = r \mid O_0 = d, O_1 = w, O_2 = w) =$
  $\Sigma_{s_2}\ P(S_3 = r, S_2 = s_2 \mid O_0 = d, O_1 = w, O_2 = w)$
  $\Sigma_{s_2}\ P(S_3 = r \mid S_2 = s_2)P(S_2 = s_2 \mid O_0 = d, O_1 = w, O_2 = w)$
- Etc. for S_4
- So: monitoring first, then straightforward Markov process updates

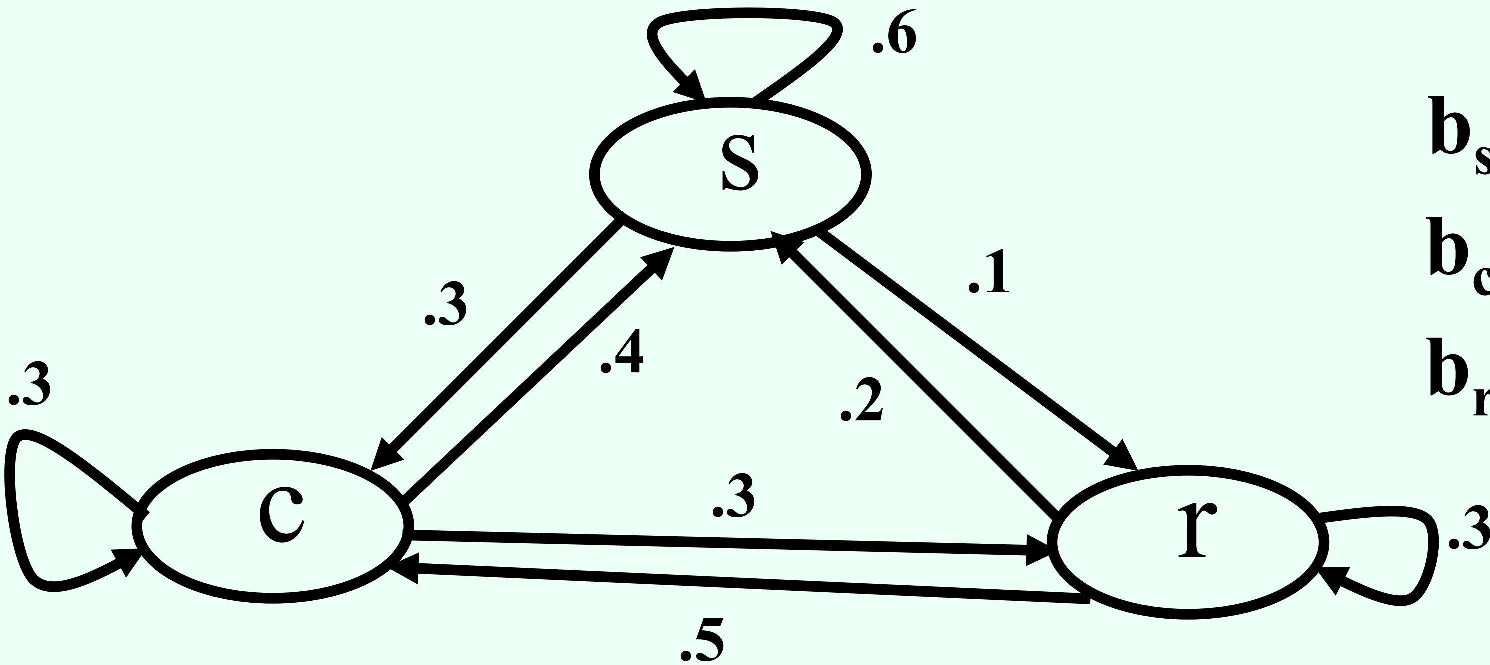
Integrating newer information



$b_{sw} = .1$
$b_{cw} = .3$
$b_{rw} = .8$

- You have been stuck in the lab for **four** days (!)
- On those days, your labmate was dry, wet, wet, dry respectively
- What is the probability that **two days ago** it was raining outside? $P(S_1 = r \mid O_0 = d, O_1 = w, O_2 = w, O_3 = d)$
  - Smoothing or hindsight problem

Hindsight problem continued…



$b_{sw} = .1$

$b_{cw} = .3$

$b_{rw} = .8$

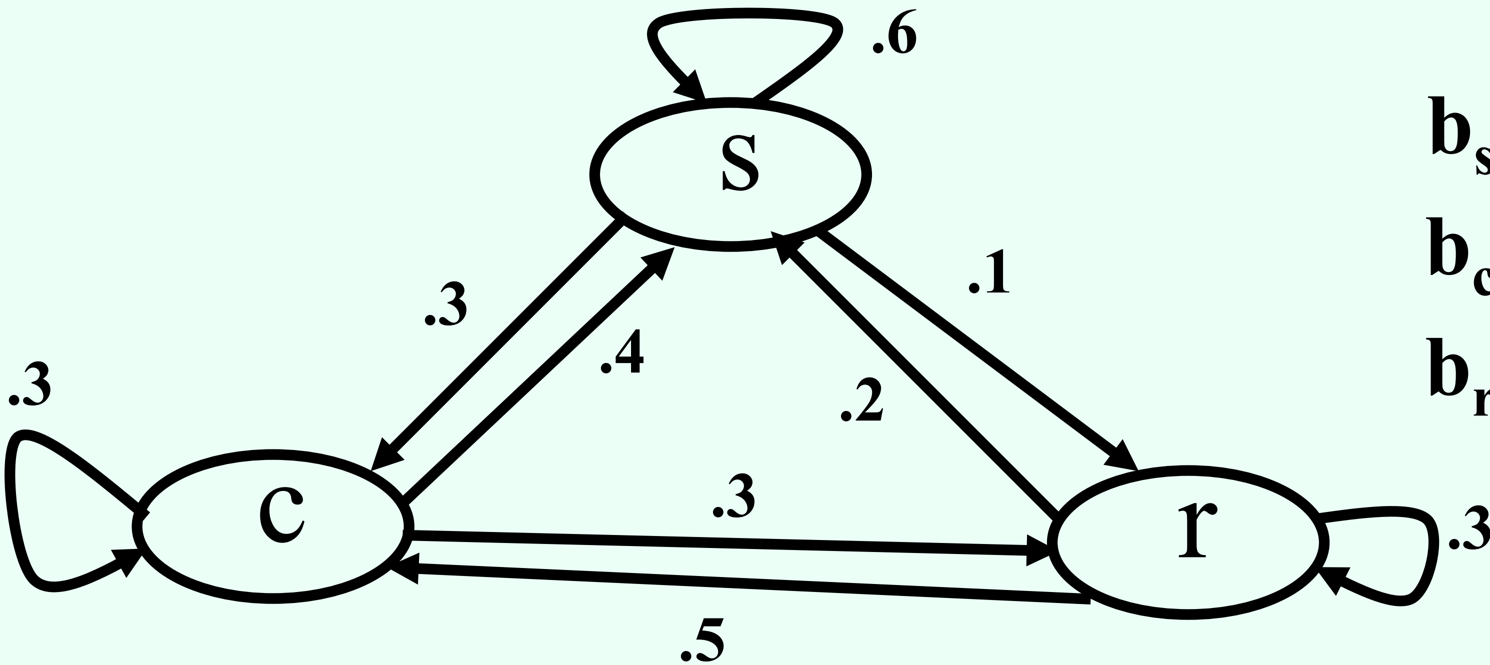
- Want: $P(S_1 = r \mid O_0 = d, O_1 = w, O_2 = w, O_3 = d)$
- "Partial" application of Bayes' rule:

  $P(S_1 = r \mid O_0 = d, O_1 = w, O_2 = w, O_3 = d) =$

  $P(S_1 = r, O_2 = w, O_3 = d \mid O_0 = d, O_1 = w) /$

  $P(O_2 = w, O_3 = d \mid O_0 = d, O_1 = w)$
- So really want to know $P(S_1, O_2 = w, O_3 = d \mid O_0 = d, O_1 = w)$

Hindsight problem continued…



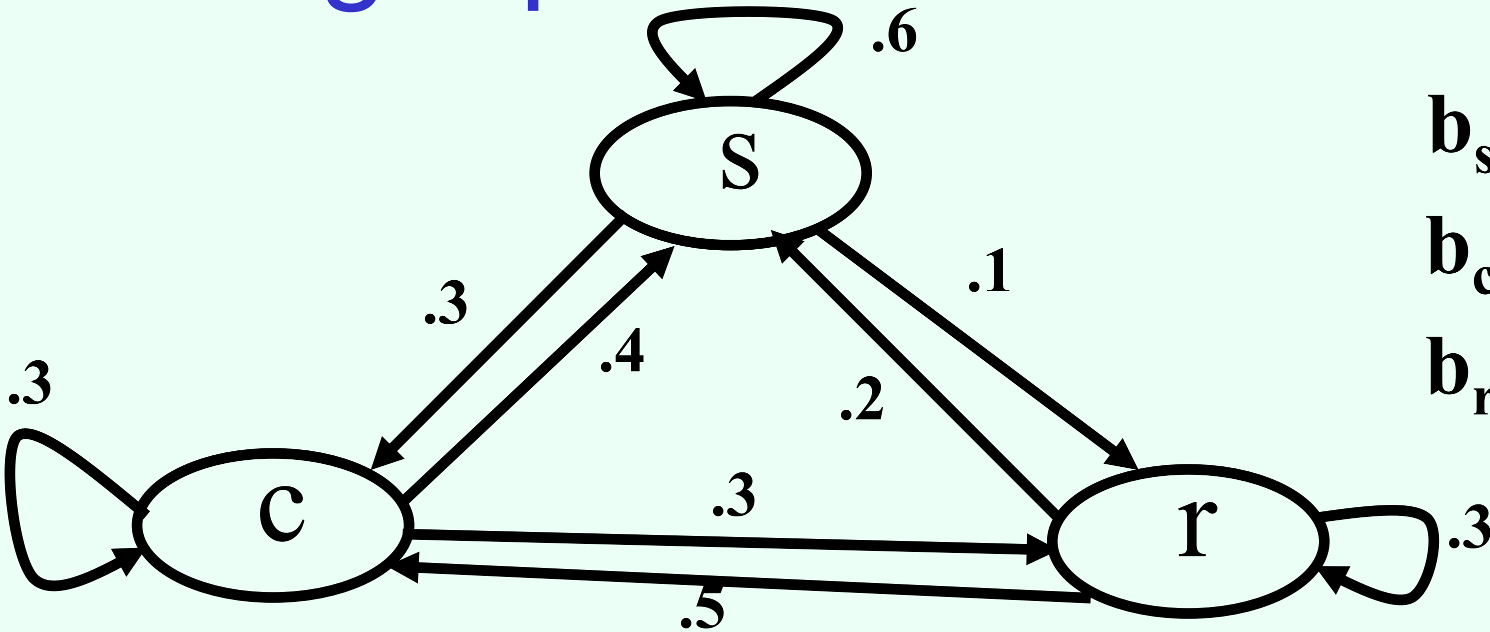
$b_{sw} = .1$

$b_{cw} = .3$

$b_{rw} = .8$

- Want to know $P(S_1 = r, O_2 = w, O_3 = d \mid O_0 = d, O_1 = w)$
- $P(S_1 = r, O_2 = w, O_3 = d \mid O_0 = d, O_1 = w) =$
  $P(S_1 = r \mid O_0 = d, O_1 = w)\, P(O_2 = w, O_3 = d \mid S_1 = r)$
- Already know how to compute $P(S_1 = r \mid O_0 = d, O_1 = w)$
- Just need to compute $P(O_2 = w, O_3 = d \mid S_1 = r)$

Hindsight problem continued…



$b_{sw} = .1$

$b_{cw} = .3$

$b_{rw} = .8$

- Just need to compute $P(O_2 = w, O_3 = d \mid S_1 = r)$
- $P(O_2 = w, O_3 = d \mid S_1 = r) =$

  $\Sigma_{s_2}\ P(S_2 = s_2, O_2 = w, O_3 = d \mid S_1 = r) =$

  $\Sigma_{s_2} P(S_2 = s_2 \mid S_1 = r)\, P(O_2 = w \mid S_2 = s_2)\, P(O_3 = d \mid S_2 = s_2)$
- First two factors directly in the model; last factor is a "smaller" problem of the same kind
- Use dynamic programming, backwards from the future
  - Similar to forwards approach from the past

Backwards reasoning in general

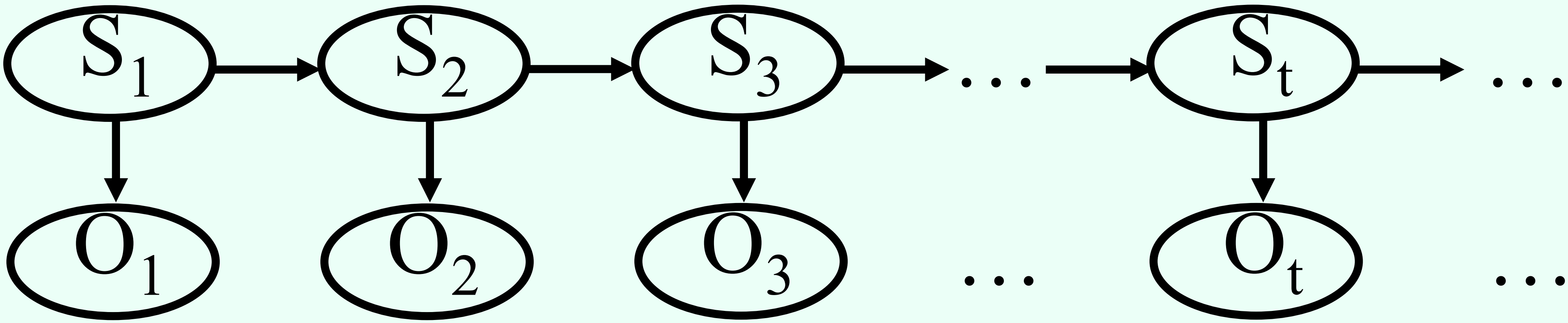
- Want to know $P(O_{k+1} = o_{k+1}, \ldots, O_t = o_t \mid S_k)$
- First compute

  $P(O_t = o_t \mid S_{t-1}) = \Sigma_{s_t} P(S_t = s_t \mid S_{t-1}) P(O_t = o_t \mid S_t = s_t)$
- Then compute

  $P(O_t = o_t, O_{t-1} = o_{t-1} \mid S_{t-2}) = \Sigma_{s_{t-1}} P(S_{t-1} = s_{t-1} \mid S_{t-2}) P(O_{t-1} = o_{t-1} \mid S_{t-1} = s_{t-1})\ P(O_t = o_t \mid S_{t-1} = s_{t-1})$
- Etc.

Variable elimination

- Because all of this is inference in a Bayes net, we can also just do variable elimination



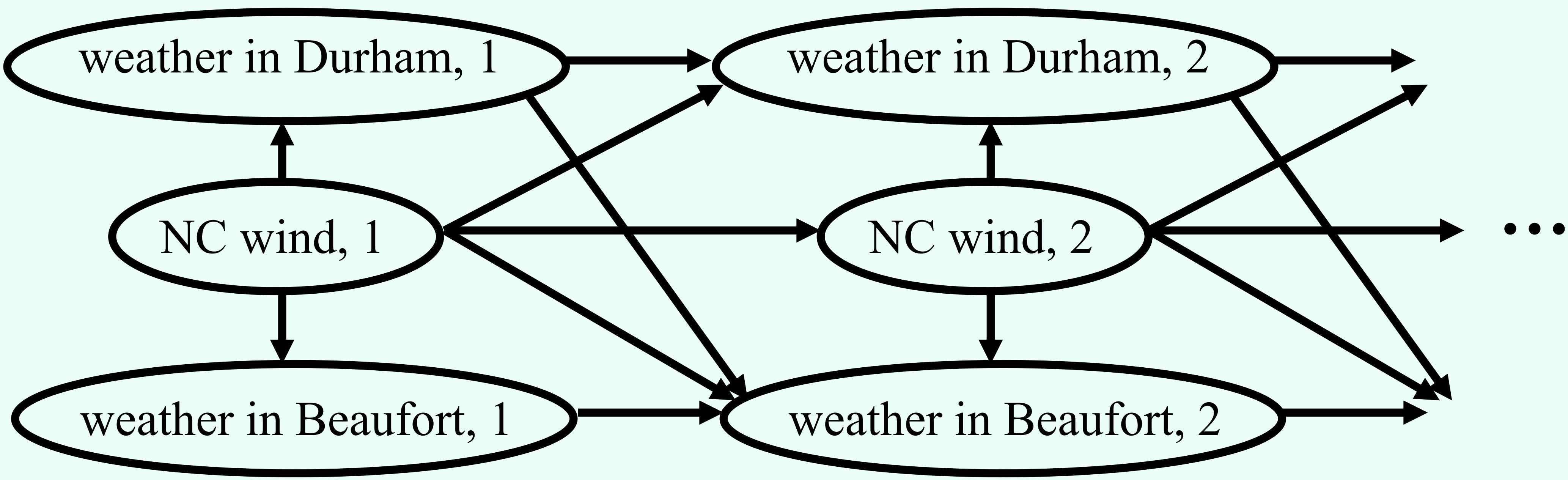
- E.g., $P(S_3 = r, O_1 = d, O_2 = w, O_3 = w) =$

  $\Sigma_{s_2}\Sigma_{s_1} P(S_1{=}s_1)P(O_1{=}d|S_1{=}s_1)P(S_2{=}s_2|S_1{=}s_1)$
  $P(O_2{=}w|S_2{=}s_2)P(S_3{=}r|S_2{=}s_2)P(O_3{=}w|S_3{=}r)$

- It's a tree, so variable elimination works well

Dynamic Bayes Nets

- So far assumed that each period has one variable for state, one variable for observation
- Often better to divide state and observation up into multiple variables



edges both within a period, and from one period to the next…

Some interesting things we skipped

- Finding the most likely sequence of states, given observations
  - Not necessary equal to the sequence of most likely states! (example?)
  - Viterbi algorithm
    - Key idea: for each period t, for every state, keep track of most likely sequence to that state at that period, given evidence up to that period
- Continuous variables
- Approximate inference methods
  - Particle filtering