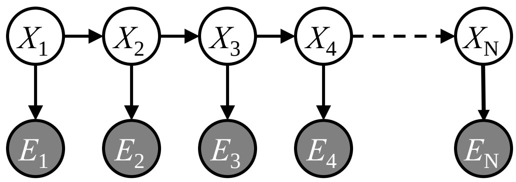
CSE 473

Lecture 22
(Chapters 14 & 15)

Probabilistic Inference and Hidden Markov Models



Recall: Probabilistic Inference

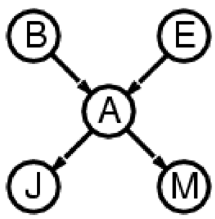
- Full joint distribution allows inference of all types of probabilities
  - E.g. Given random variables A, B, E, J, M, if you want P(B|J,M):

    $P(B|J,M) = \alpha\, P(B,J,M) = \alpha\, \Sigma_{E,A}\, P(B,J,M,E,A)$

- Problem: Full joint requires you to specify 2*2*2*2*2 = 32 values

Solution: Bayesian networks

- Simple graphical notation for conditional independence assertions
  - In many cases, allows *compact specification* of full joint distributions
  - Example BN for A, B, E, J, M



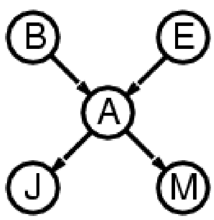
P(J,M,A,B,E) =

$\Pi_i \; P(X_i \mid Parents(X_j))$ =

P(J|A) P(M|A) P(A|B,E) P(B) P(E)

Only requires 2+2+4+1+1=10 values

Why is joint = $\Pi_i \; P(X_i \mid Parents(X_j))$?



Keep applying definition of conditional probability:

P(J,M,A,B,E) =

= P(J|M,A,B,E) P(M,A,B,E)

= P(J|A) P(M,A,B,E)

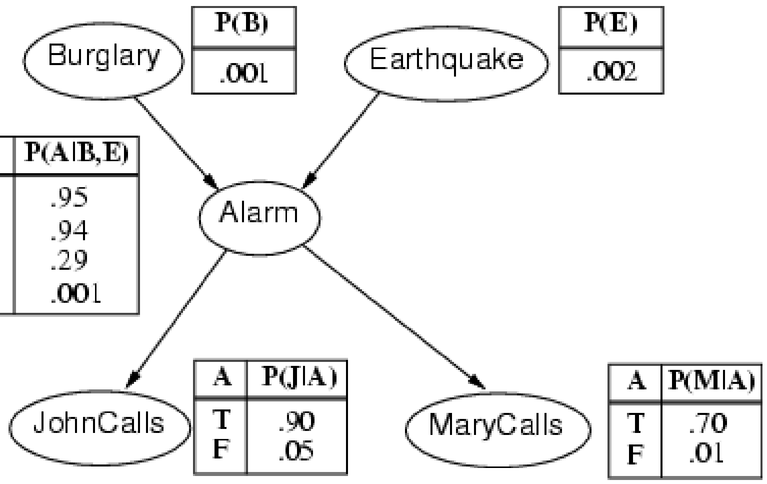
= P(J|A) P(M|A,B,E) P(A,B,E)

= P(J|A) P(M|A) P(A,B,E)

= P(J|A) P(M|A) P(A|B,E) P(B,E)

= P(J|A) P(M|A) P(A|B,E) P(B) P(E)

Bayesian Network for Burglars and Earthquakes



P(B)
.001

P(E)
.002

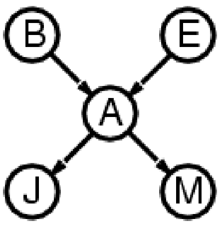
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

What is the probability of Burglary given that John and Mary called?

Compute P(B=true | J=true, M=true)



$P(b|j,m) = \alpha\, P(b,j,m)$

$= \alpha\, \Sigma_{e,a}\, P(b,j,m,e,a)$

$= \alpha\, \Sigma_{e,a}\, P(b)\, P(e)\, P(a|b,e)\, P(j|a)\, P(m|a)$

$= \alpha\, P(b)\, \Sigma_e\, P(e)\, \Sigma_a\, P(a|b,e) P(j|a) P(m|a)$

- Join all factors containing *a*
- Sum out *a* to get new function of b,e,j,m only

6

Variable Elimination (VE) Algorithm

- Eliminate variables one-by-one until there is a factor with only the query variables:
  1. *join* all factors containing that variable, multiplying probabilities
  2. *sum out* the influence of the variable

Remaining factor is a function of b, j, m

Eliminate e

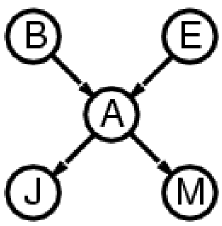
Eliminate a

$$P(b|j,m) = \alpha\, P(b)\, \Sigma_e P(e)\, \Sigma_a P(a|b,e) P(j|a) P(m|a)$$

↓

Function of b,j,m

Example of VE: P(J)



P(J)

$= \Sigma_{M,A,B,E}\, P(J,M,A,B,E)$

$= \Sigma_{M,A,B,E}\, P(J|A)P(M|A)\, P(A|B,E)\, P(B)\, P(E)$

$= \Sigma_A P(J|A)\, \Sigma_M P(M|A)\, \Sigma_B P(B)\, \Sigma_E P(A|B,E)P(E)$

$= \Sigma_A P(J|A)\, \Sigma_M P(M|A)\, \Sigma_B P(B)\, f1(A,B)$

$= \Sigma_A P(J|A)\, \Sigma_M P(M|A)\, f2(A)$

$= \Sigma_A P(J|A)\, f3(A)$

$= f4(J)$

8

Other Inference Algorithms

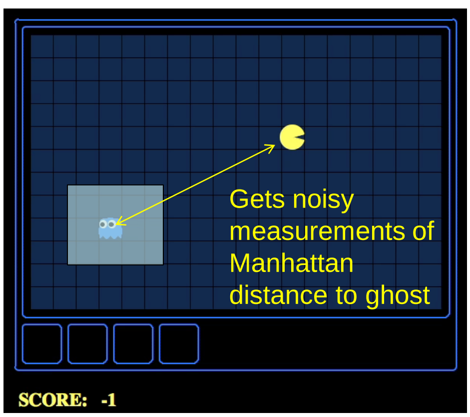
- Direct Sampling:
  - Repeat N times:
    - Use random number generator to generate sample values for each node
    - Start with nodes with no parents
    - Condition on sampled parent values for other nodes
  - Count frequencies of samples to get an approximation to desired distribution
- Other variants: Rejection sampling, likelihood weighting, Gibbs sampling and other MCMC methods (see text)
- Belief Propagation: A “message passing” algorithm for approximating P(X|evidence) for each node variable X
- Variational Methods: Approximate inference using distributions that are more tractable than original ones

(see text for details)

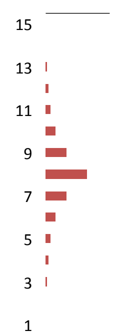
9

Pac-Man goes Ghost Hunting

Pac-Man does not know true position of the ghost



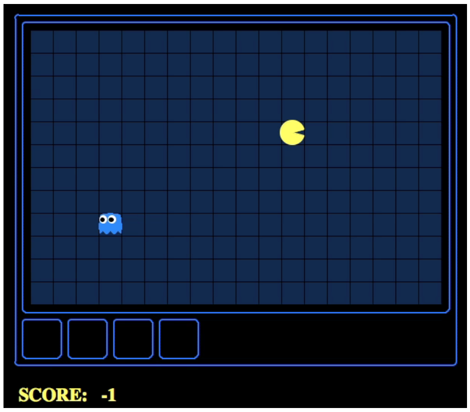
Noisy distance prob
(if true distance = 8)



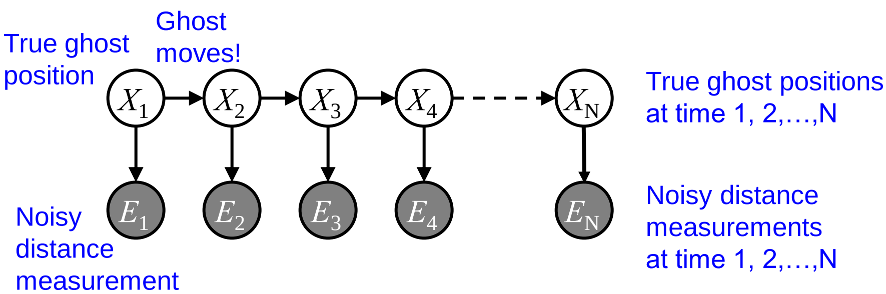
Must infer probability distribution over true ghost position

10

Example of Ghost Tracking (movie)



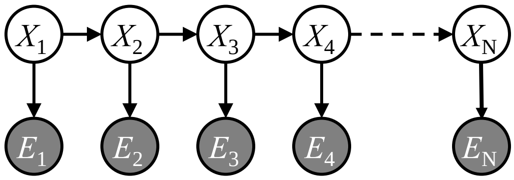
Bayesian Network for Tracking



This “Dynamic” Bayesian network is also called a Hidden Markov Model (HMM)

- Dynamic = time-dependent
- Hidden = state (ghost position) is hidden
- Markov = current state only depends on previous state
  Similar to MDP (Markov decision process) but no actions

Hidden Markov Model (HMM)



Hidden State at time t = 1, 2,…,N

Emissions (measurements) at time t = 1, 2,…,N

HMM is defined by 2 conditional probabilities:

$P(X_t \mid X_{t-1})$ Transition model $= P(X' \mid X)$

$P(E_t \mid X_t)$ Emission model $= P(E \mid X)$
(aka measurement/observation model)

plus initial state distribution $P(X_1)$

13

Project 4: Ghostbusters

- **Plot:** Pacman's grandfather, Grandpac, learned to hunt ghosts for sport.
  - Blinded by his power, but can hear the ghosts' banging and clanging sounds.
- **Transition Model:** Ghosts move randomly, but are sometimes biased.
- **Emission Model:** Pacman gets a “noisy” distance to each ghost.

Ghostbusters HMM

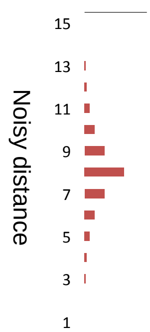
- $P(X_1)$ = uniform
- $P(X'|X)$ = ghost usually moves *clockwise*, but sometimes moves in a random direction or stays in place
- $P(E|X)$ = compute Manhattan distance to ghost from Pac-Man and emit a *noisy distance* given this true distance (see example for true distance = 8)

$P(X_1)$

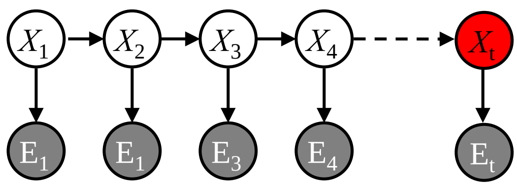
1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

P(X'|X=<1,2>)

1/6	1/6	1/2
0	1/6	0
0	0	0



HMM Inference Problem



Where is the ghost now?
Compute posterior probability over X_t

- Given evidence (all measurements made so far) $E_{1:t} = e_{1:t}$
- Main inference problem:
  - Filtering: Find posterior $P(X_t|e_{1:t})$ for current t

The "Forward" Algorithm for Filtering

- Want to compute the "belief" $B_t(X) = P(X_t | e_{1:t})$
- Derive belief update rule from probability definitions, Bayes' rule and Markov assumption:

$$P(X_t | e_1,\dots,e_t) = \alpha\, P(e_t | X_t, e_1,\dots,e_{t-1})\, P(X_t | e_1,\dots,e_{t-1}) \quad \text{Bayes}$$

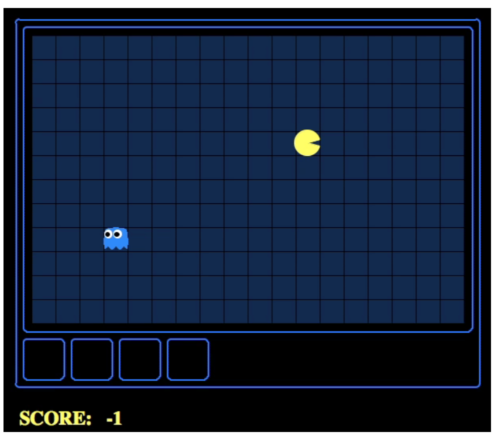
$$= \alpha\, P(e_t | X_t) \sum_{X_{t-1}} P(X_t, X_{t-1} | e_1,\dots,e_{t-1}) \quad \text{Markov + Marginalize}$$

$$= \alpha\, P(e_t | X_t) \sum_{X_{t-1}} P(X_t | X_{t-1}, e_1,\dots,e_{t-1}) P(X_{t-1} | e_1,\dots,e_{t-1})$$

$$= \alpha\, P(e_t | X_t) \sum_{X_{t-1}} P(X_t | X_{t-1}) P(X_{t-1} | e_1,\dots,e_{t-1})$$

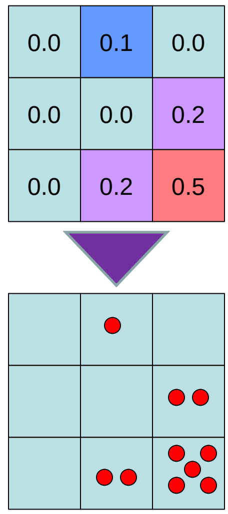
New estimate | Normalization constant | Emission model | Transition model | Previous estimate

Example of Filtering (Tracking) using the Forward Algorithm (movie)



Particle Filtering Motivation

- Sometimes |X| is too big for exact inference
  - |X| may be too big to even store $B_t(X)$
    E.g. when X is continuous
  - $|X|^2$ may be too big to do updates
- Solution: Approximate inference
  - Track a set of *samples* of X
  - Samples are called ***particles***
  - Number of samples for X=x is proportional to probability of x



Next Time

- Particle Filtering and its Applications
  - Guest lecture by Prof. Dieter Fox
- To Do:
  - Project 4 (last project! Assigned today)