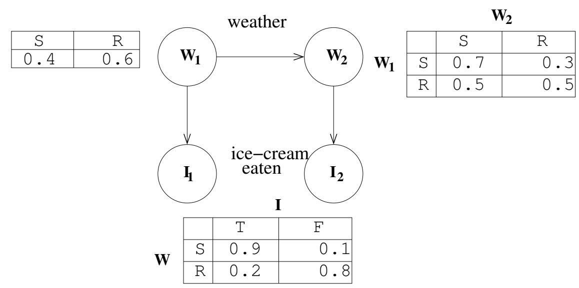
CS 188 Spring 2010 | Introduction to Artificial Intelligence | Section 10 Solutions Sampling and Bayes Nets

Q1. Sampling

The Bayes Net in the diagram to the right describes a person's ice-cream eating habits. The nodes W_1 and W_2 stand for the weather on days 1 and 2, which can either be rainy R or sunny S, The nodes I_1 and I_2 stand for whether or not the person ate ice-cream that day. They can take the values T or F. The conditional probability distributions relevant to the graphical model are also given to you, note that there is a single conditional probability distribution $P(I|W)$, which I_1 and I_2 follow.



Suppose we use prior sampling to produce the following samples from the weather/ice-cream model:

(W_1, I_1, W_2, I_2)	(W_1, I_1, W_2, I_2)
R, F, R, F excluded by rejection sampling	R, F, R, F excluded by rejection sampling
S, F, S, T excluded by rejection sampling	S, T, S, T excluded by rejection sampling
S, T, S, T excluded by rejection sampling	R, F, R, T excluded by rejection sampling
S, T, S, T excluded by rejection sampling	S, T, R, F
S, T, R, F	R, F, S, T excluded by rejection sampling

(a) What is $\hat{P}(W_2 = R)$?

Number of samples in which $W_2 = R$: 5. Total number of samples: 10. Answer $5/10 = 0.5$.

(b) Cross off samples rejected by rejection sampling if we're computing $\hat{P}(W_2|I_1 = T, I_2 = F)$

(c) Now use likelihood weighting, and assume we've generated the following samples, given the evidence $I_1 = T$ and $I_2 = F$.

(W_1, I_1, W_2, I_2)	(W_1, I_1, W_2, I_2)
S, T, R, F	S, T, S, F
R, T, R, F	S, T, S, F
S, T, R, F	R, T, S, F

(i) What is the weight of the first sample (S, T, R, F) above?

The weight given to a sample in likelihood weighting is $\prod_{\text{evidence variables } e} \Pr(e|\text{Parents}(e))$. In this case, the evidence is $I_1 = T, I_2 = F$. The weight of the first sample is therefore $w = \Pr(I_1 = T|W_1 = S) \cdot \Pr(I_2 = F|W_2 = R) = 0.9 \cdot 0.8 = 0.72$.

(ii) Use likelihood weighting to estimate $\hat{P}(W_2|I_1 = T, I_2 = F)$

The sample weights are given by

(W_1, I_1, W_2, I_2)	w	(W_1, I_1, W_2, I_2)	w
S, T, R, F	0.72	S, T, S, F	0.09
R, T, R, F	0.16	S, T, S, F	0.09
S, T, R, F	0.72	R, T, S, F	0.02

$$\hat{\Pr}(W_2 = R|I_1 = T, I_2 = F) = \frac{0.72 + 0.16 + 0.72}{0.72 + 0.16 + 0.72 + 0.09 + 0.09 + 0.02} = 0.889$$

$$\hat{\Pr}(W_2 = S|I_1 = T, I_2 = F) = 1 - 0.889 = 0.111.$$

Q2. HMMs: Tracking a Jabberwock

You have been put in charge of a Jabberwock for your friend Lewis. The Jabberwock is kept in a large tugley wood which is conveniently divided into an $N \times N$ grid. It wanders freely around the N^2 possible cells. At each time step $t = 1, 2, 3, \ldots$, the Jabberwock is in some cell $X_t \in \{1, \ldots, N\}^2$, and it moves to cell X_{t+1} randomly as follows: with probability $1 - \epsilon$, it chooses one of the (up to 4) valid neighboring cells uniformly at random; with probability ϵ, it uses its magical powers to teleport to a random cell uniformly at random among the N^2 possibilities (it might teleport to the same cell). Suppose $\epsilon = \frac{1}{2}$, $N = 10$ and that the Jabberwock always starts in $X_1 = (1, 1)$.

(a) Compute the probability that the Jabberwock will be in $X_2 = (2, 1)$ at time step 2. What about $\Pr(X_2 = (4, 4))$?

$\Pr(X_2 = (2, 1)) = 1/2 \cdot 1/2 + 1/2 \cdot 1/100 = 0.255$

$\Pr(X_2 = (4, 4)) = 1/2 \cdot 1/100 = 0.005$

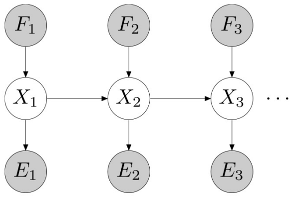
At each time step t, you dont see X_t but see E_t, which is the row that the Jabberwock is in; that is, if $X_t = (r, c)$, then $E_t = r$. You still know that $X_1 = (1, 1)$.

(b) Suppose we see that $E_1 = 1$, $E_2 = 2$, $E_3 = 10$. Fill in the following table with the distribution over X_t after each time step, taking into consideration the evidence. Your answer should be concise. *Hint*: you should not need to do any heavy calculations.

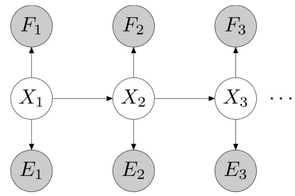
t	$\Pr(X_t, e_{1:t-1})$	$\Pr(X_t, e_{1:t})$
1	(1,1) : 1.0, (others) : 0.0	(1,1): 1.0, (others): 0.0
2	(1,2), (2,1): 51/200, (others): 1/200	(2,1): 51/200, (2,2+): 1/200

You are a bit unsatisfied that you cant pinpoint the Jabberwock exactly. But then you remembered Lewis told you that the Jabberwock teleports only because it is frumious on that time step, and it becomes frumious independently of anything else. Let us introduce a variable $F_t \in \{0, 1\}$ to denote whether it will teleport at time t. We want to to add these frumious variables to the HMM.

Consider the two candidates:



(A)



(B)

(A)	(B)
$X_1 \perp X_3 \mid X_2$ +	$X_1 \perp X_3 \mid X_2$ +
$X_1 \perp E_2 \mid X_2$ +	$X_1 \perp E_2 \mid X_2$ +
$X_1 \perp F_2 \mid X_2$	$X_1 \perp F_2 \mid X_2$ +
$X_1 \perp E_4 \mid X_2$ +	$X_1 \perp E_4 \mid X_2$ +
$X_1 \perp F_4 \mid X_2$ +	$X_1 \perp F_4 \mid X_2$ +
$E_3 \perp F_3 \mid X_3$ +	$E_3 \perp F_3 \mid X_3$ +
$E_1 \perp F_2 \mid X_2$	$E_1 \perp F_2 \mid X_2$ +
$E_1 \perp F_2 \mid E_2$	$E_1 \perp F_2 \mid E_2$

(c) For each model, circle the conditional independence assumptions above which are true in that model.

(d) Which Bayes net is more appropriate for the problem domain here, (A) or (B)? Justify your answer.

(A) because the choice of X depends on F in the problem description.

For the following questions, your answers should be fully general for models of the structure shown above, not specific to the teleporting Jabberwock. For full credit, you should also simplify as much as possible (including pulling constants outside of sums, etc.).

(e) For (A), express $\Pr(X_{t+1}, e_{1:t+1}, f_{1:t+1})$ in terms of $\Pr(X_t, e_{1:t}, f_{1:t})$ and the CPTs used to define the network. Assume the E and F nodes are all observed.

$$\Pr(x_{t+1}, e_{1:t+1}, f_{1:t+1}) = \Pr(e_{t+1}|x_{t+1}) \sum_{x_t} \Pr(x_{t+1}|x_t, f_{t+1}) \Pr(x_t, e_{1:t}, f_{1:t}).$$

(f) For (B), express $\Pr(X_{t+1}, e_{1:t+1}, f_{1:t+1})$ in terms of $\Pr(X_t, e_{1:t}, f_{1:t})$ and the CPTs used to define the network. Assume the E and F nodes are all observed.

$$\Pr(x_{t+1}, e_{1:t+1}, f_{1:t+1}) = \Pr(e_{t+1}|x_{t+1}) \Pr(f_{t+1}|x_{t+1}) \sum_{x_t} \Pr(x_{t+1}|x_t) \Pr(x_t, e_{1:t}, f_{1:t}).$$

Suppose that we don't actually observe the F_ts.

(g) For (A), express $\Pr(X_{t+1}, e_{1:t+1})$ in terms of $\Pr(X_t, e_{1:t})$ and the CPTs used to define the network.

$$\Pr(x_{t+1}, e_{1:t+1}) = \Pr(e_{t+1}|x_{t+1}) \sum_{f_{t+1}} \Pr(f_{t+1}) \sum_{x_t} \Pr(x_{t+1}|x_t, f_{t+1}) \Pr(x_t, e_{1:t}).$$

(h) For (B), express $\Pr(X_{t+1}, e_{1:t+1})$ in terms of $\Pr(X_t, e_{1:t})$ and the CPTs used to define the network.

$$\Pr(x_{t+1}, e_{1:t+1}) = \Pr(e_{t+1}|x_{t+1}) \sum_{x_t} \Pr(x_{t+1}|x_t) \Pr(x_t, e_{1:t}).$$