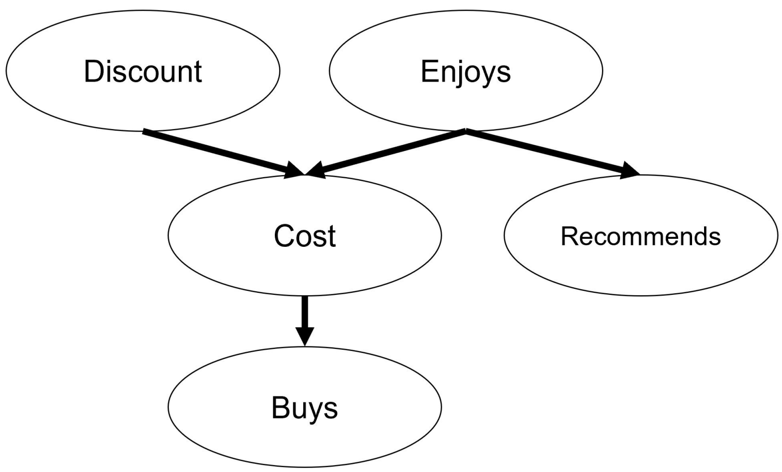
1 Inference

Realizing that students aren't particularly fond of reading the textbook, the 281 course staff have developed a software that automatically scans the textbook and outputs key points for each individual chapter. However, since the development of the software requires time and computational resources, the 281 staff decides to offer a free one month trial to students, after which a paid subscription is necessary to keep using the software. The following network and variables are used to represent the problem:

- $Discount(D)$: $+d$ if a discount is offered, $-d$ otherwise
- $Enjoys(E)$: $+e$ if a student enjoys the software, $-e$ otherwise
- $Cost(C)$: $+c$ if the software cost is < 20, $-c$ otherwise
- $Recommends(R)$: $+s$ if the student recommends the software to a friend, $-s$ otherwise
- $Buys(B)$: $+b$ if the student buys a software subscription, $-b$ otherwise



(a) How can we represent the probability that a student buys and recommends the software using the conditional probabilities at each node?

(b) The staff has surveyed students and collected data on whether the students enjoyed the software or not. With this information, we want to perform a inference on a joint distribution where the query variable is $Buys$ (B).

(i) How can we represent the probability expression in terms of conditional probabilities from the network?

(ii) What are the hidden and evidence variable(s)?

(c) Using the probability expression from the previous part, we want to compute the query B given evidence that the student enjoys the software. Assume the variable ordering is in alphabetical order.

(i) How many factors are there, and what are the dimensions of each factor?

(ii) Run the variable elimination algorithm to eliminate repeated computations for the expression $P(B|+e)$.

(iii) How does the resulting expression change if the variable ordering is instead in reverse alphabetical order?

(iv) How do the two orderings compare with respect to time and space complexity?

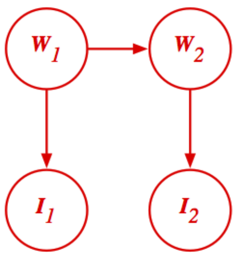
(v) Describe a heuristic that could be useful in determining a variable ordering to minimize the size of the largest factor.

2 Sampling

(a) Compared to other sampling methods (rejection, likelihood weighting, Gibbs), what kind of information can prior sampling not use (that other methods can)?

(b) How does reject sampling work on a high level, and what is its biggest/immediate weakness?

The diagram below describes a person's ice-cream eating habits based on the weather. The nodes W_i stand for the weather on a day i, which can either be **s** (sunny) or **r** (rainy). The nodes I_i represent whether the person ate ice-cream on day i, which can either be **t** (true) or **f** (false).



W_1	$P(W_1)$
s	0.6
r	0.4

I	W	$P(I \vert W)$
t	s	0.9
f	s	0.1
t	r	0.2
f	r	0.8

W_2	W_1	$P(W_2 \vert W_1)$
s	s	0.7
r	s	0.3
s	r	0.5
r	r	0.5

Assume we generate the following six samples given the evidence $I_1 = t$ and $I_2 = f$:

$$(W_1, I_1, W_2, I_2) = < s, t, r, f >, < r, t, r, f >, < s, t, r, f >, < s, t, s, f >, < s, t, s, f >, < r, t, s, f >$$

Using these samples, we will complete the following table:

(W_1, I_1, W_2, I_2)	Count/N	w	Joint
s, t, s, f	2/6	0.09	0.03
s, t, r, f	/6		
r, t, s, f	/6		
r, t, r, f	/6		

(c) What is the weight of the sample (s, t, r, f) above? Recall that the weight given to a sample in likelihood weighting is:

$$w = \prod_{\text{Evidence variables } e} P(e|\text{Parents}(e)).$$

(d) What is the estimate of P(s, t, r, f) given the samples?

(e) Compute the rest of the entries in the table. Use the estimated joint probabilities to estimate $P(W_2|I_1 = t, I_2 = f)$.

(f) What is a weakness of likelihood weighing sampling? How does Gibbs sampling work, and how does it address this limitation?

3 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 $X_1 = (1, 1)$.

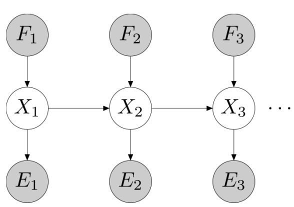
(a) Compute the probability that the Jabberwock will be in $X_2 = (2, 1)$ at time step 2. What about $P(X_2 = (4, 4))$?

At each time step t, you don't 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 $X_1 = (1, 1)$.

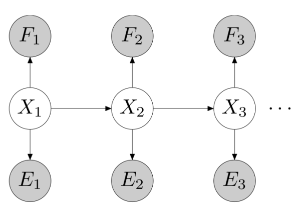
(b) Suppose we see that $E_1 = 1$, $E_2 = 2$, $E_3 = 10$. Fill in the following table with the distribution over X_t after each time step, taking into consideration the evidence. Your answer should be concise. Hint: you should not need to do any heavy calculations.

t	$P(X_t \mid e_{1:t-1}, X_1 = (1,1))$	$P(X_t \mid e_{1:t}, X_1 = (1,1))$
1		
2		

You are a bit unsatisfied that you can't 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.

For the following questions, your answers should be fully general for models of the structure shown above, not specific to the teleporting Jabberwock.

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

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

Suppose that we don't actually observe the F_ts.

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

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