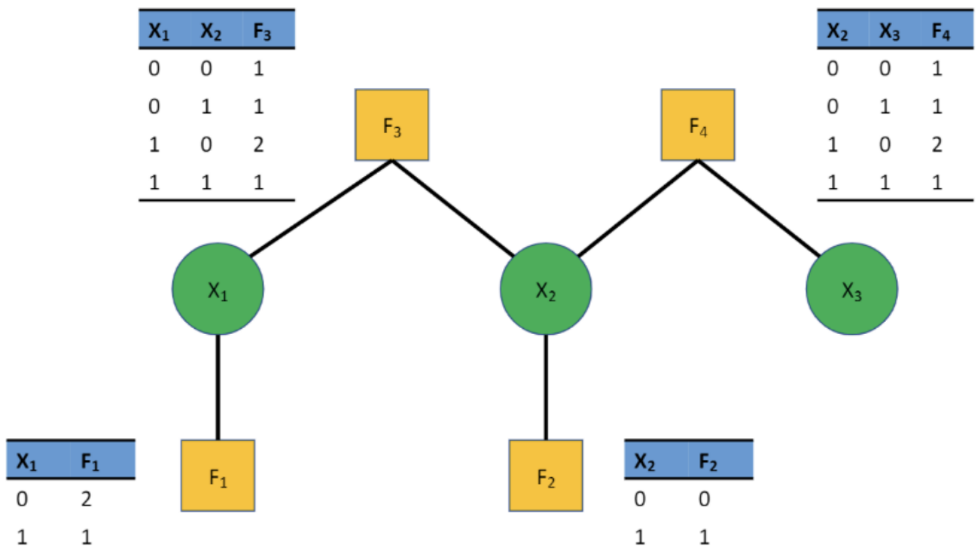
CS221 Problem Workout Solutions

Week 7

1) **Problem 1: Markov Networks** This problem will give you some practice on computing probabilities given a Markov network. Specifically, given the Markov network below, we will ask you questions about the probability distribution $p(X_1, X_2, X_3)$ over the binary random variables X_1, X_2, and X_3.



(a) What is $p(X_1 = 0, X_2 = 0, X_3 = 0)$?

Solution We have
$p(X_1 = 0, X_2 = 0, X_3 = 0) = \frac{F_1(X_1=0)F_2(X_2=0)F_3(X_1=0,X_2=0)F_4(X_2=0,X_3=0)}{Z} = \frac{2\times 0\times 1\times 1}{Z} = 0,$
where
$Z = \sum_{X_1\in\{0,1\}} \sum_{X_2\in\{0,1\}} \sum_{X_3\in\{0,1\}} F_1(X_1)F2(X_2)F_3(X_1, X_2), F_4(X_2, X_3) = 9.$

(b) What is $p(X_1 = 0, X_2 = 1, X_3 = 0)$?

Solution We have
$p(X_1 = 0, X_2 = 1, X_3 = 0) = \frac{F_1(X_1=0)F_2(X_2=1)F_3(X_1=0,X_2=1)F_4(X_2=1,X_3=0)}{Z} = \frac{2\times 1\times 1\times 2}{Z} = \frac{4}{9},$
where $Z = 9$ as before.

(c) What is $p(X_2 = 0)$?

Solution $p(X_2 = 0) = 0$.

(d) What is $p(X3 = 0)$?

Solution $p(X_3 = 0) = \frac{\sum_{X_1 \in \{0,1\}} \sum_{X_2 \in \{0,1\}} p(X_3=0, X_1, X_2)}{Z} = \frac{6}{9} = \frac{2}{3}$.

2) Problem 2: Bayesian Networks Basics

$P(A\|D,X)$			
$+d$	$+x$	$+a$	0.9
$+d$	$+x$	$-a$	0.1
$+d$	$-x$	$+a$	0.8
$+d$	$-x$	$-a$	0.2
$-d$	$+x$	$+a$	0.6
$-d$	$+x$	$-a$	0.4
$-d$	$-x$	$+a$	0.1
$-d$	$-x$	$-a$	0.9

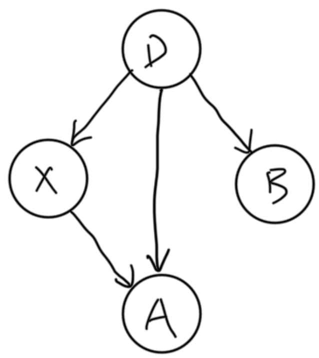
$P(D)$	
$+d$	0.1
$-d$	0.9

$P(X\|D)$		
$+d$	$+x$	0.7
$+d$	$-x$	0.3
$-d$	$+x$	0.8
$-d$	$-x$	0.2

$P(B\|D)$		
$+d$	$+b$	0.7
$+d$	$-b$	0.3
$-d$	$+b$	0.5
$-d$	$-b$	0.5

(a) Given the tables above, draw a minimal representative Bayesian network of this model. Be sure to label all nodes and the directionality of the edges.

Solution



(b) Compute the following probabilities: $P(+d \mid +b)$, $P(+d, +a)$, $P(+d \mid +a)$.

Solution

$$\begin{aligned} P(+d|+b) &= \frac{P(+b|+d)P(+d)}{P(+b)} \\ &= \frac{P(+b|+d)P(+d)}{\sum_d P(+b|d)P(d)} \\ &= \frac{0.7 * 0.1}{0.7 * 0.1 + 0.5 * 0.9} \\ &= 0.135 \end{aligned}$$

$$\begin{aligned} P(+d,+a) &= \sum_x P(+a|+d,x)P(x|+d)P(+d) \\ &= 0.1(0.9 * 0.7 + 0.8 * 0.3) \\ &= 0.087 \end{aligned}$$

In order to compute $P(+d|+a)$, we need to compute P(-d, +a).

$$\begin{aligned} P(-d, +a) &= \sum_x P(-d, x, +a) \\ &= \sum_x P(+a|-d, x)P(x|-d)P(-d) \\ &= 0.90(0.60 * 0.80 + 0.10 * 0.20) \\ &= 0.45 \end{aligned}$$

$$P(+d|+a) = \frac{P(+a, +d)}{\sum_d P(+a, d)} = \frac{0.087}{0.087 + 0.45} = 0.162$$

(c) Which of the following conditional independences are guaranteed by the above network?

- ☐ $X \perp\!\!\!\perp B \,|\, D$
  (X and B are cond. ind. given D)
- ☐ $D \perp\!\!\!\perp A \,|\, X$
- ☐ $D \perp\!\!\!\perp A \,|\, B$
- ☐ $D \perp\!\!\!\perp X \,|\, A$

Solution Only $X \perp\!\!\!\perp B \,|\, D$ is guaranteed.

3) **Problem 3: Bayesian Networks Trivia**

As the president of the National Trivia Association, you must choose between the Bayesians and the Markovians, the nation's top two rival trivia teams, to represent the US at the World Trivia Olympics. To determine the more popular team, you decide to model the change in monthly TV viewership using a Bayesian network.

Let B_t and M_t denote the number of TV viewers that the Bayesians and Markovians have in month t respectively. You have no way of observing these quantities directly, but you can observe two other quantities which they influence: let S_t denote the number of times internet users searched for the Bayesians in month t, and let A_t denote the attendance of the friendly match at a neighborhood pub between the Bayesians and the Markovians in month t.

The viewerships of the two teams evolve according to the following model, where each month a fan is either gained or lost with equal probability:

$$\Pr(M_{t+1}|M_t) = \begin{cases} \frac{1}{2} & \text{if } M_{t+1} = M_t - 1 \\ \frac{1}{2} & \text{if } M_{t+1} = M_t + 1 \\ 0 & \text{otherwise} \end{cases} \qquad \Pr(B_{t+1}|B_t) = \begin{cases} \frac{1}{2} & \text{if } B_{t+1} = B_t - 1 \\ \frac{1}{2} & \text{if } B_{t+1} = B_t + 1 \\ 0 & \text{otherwise} \end{cases}$$

The Bayesian fans like to rewatch their trivia shows by searching the recaps online! We model the fan's size's influence on the number of internet searches by:

$$\Pr(S_t|B_t) = \begin{cases} 0.3 & \text{if } S_t = B_t \\ 0.25 & \text{if } S_t = B_t - 1 \\ 0.2 & \text{if } S_t = B_t - 2 \\ 0.15 & \text{if } S_t = B_t - 3 \\ 0.1 & \text{if } S_t = B_t - 4 \\ 0 & \text{otherwise} \end{cases}$$

Lastly, because most TV viewers attend each monthly friendly matches (although sometimes more, and sometimes fewer), we model the influence of the TV viewership number on the friendly match attendance by:

$$\Pr(A_t|B_t, M_t) = \begin{cases} 0.14 & \text{if } A_t = B_t + M_t \\ 0.13 & \text{if } |A_t - (B_t + M_t)| = 1 \\ 0.11 & \text{if } |A_t - (B_t + M_t)| = 2 \\ 0.09 & \text{if } |A_t - (B_t + M_t)| = 3 \\ 0.06 & \text{if } |A_t - (B_t + M_t)| = 4 \\ 0.04 & \text{if } |A_t - (B_t + M_t)| = 5 \\ 0 & \text{otherwise} \end{cases}$$

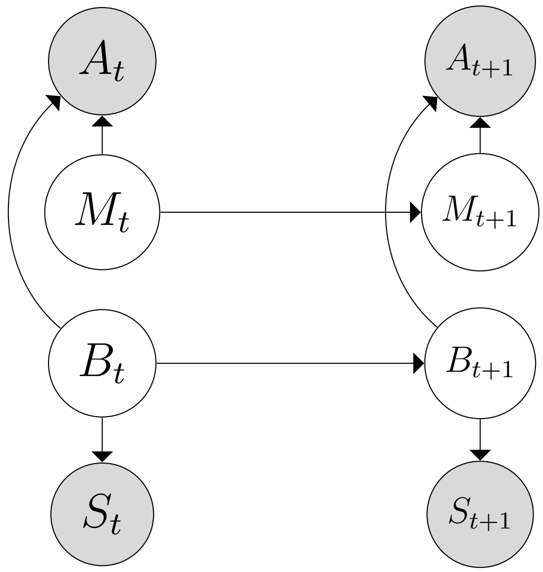


Figure 1: The changing TV viewership count modeled as a dynamic Bayesian network. The unshaded nodes correspond to the latent/hidden TV viewership counts, and the shaded nodes correspond to the observable emissions.

a. (*10 points*) **Inference**

Suppose the Bayesian's trivia team captain took a nationwide poll in month t that concluded they had exactly 75 TV viewers. Suppose additionally that in month $t+2$, the search engine reported 73 people search for the Bayesians online. What is the probability that in month $t+2$ the Bayesians have 77 TV viewers?

$\Pr(B_{t+2} = 77 | B_t = 75, S_{t+2} = 73) =$

Solution By Bayes rule, we have:

$$\Pr(B_{t+2} = 77 | B_t = 75, S_{t+2} = 73) = \frac{\Pr(S_{t+2} = 73 | B_t = 75, B_{t+2} = 77)\Pr(B_{t+2} = 77 | B_t = 75)}{\Pr(S_{t+2} = 73 | B_t = 75)}$$

We'll begin with the first term in the numerator; because S_{t+2} is conditionally independent of B_t given B_{t+2}, we have $\Pr(S_{t+2} = 73 | B_t = 75, B_{t+2} = 77) = \Pr(S_{t+2} = 73 | B_{t+2} = 77)$. This is simply given by our jersey sales model; the probability that the Bayesians sell four fewer jerseys than they have fans is 0.1.

We turn next to the second term in the numerator; if there are 75 fans in month t, then with equal probability there are either 74 or 76 fans in month $t+1$. If there were 74 in month $t+1$, then there would be either 73 or 75 in month $t+2$ with equal probability, and if there were 76 in month $t+1$, then there would be either 75 or 77 in month $t+2$ with equal probability. Thus, we have that $\Pr(B_{t+2} = 73 | B_t = 75) = \Pr(B_{t+2} = 77 | B_t = 75) = 0.25$, and $\Pr(B_{t+2} = 75 | B_t = 75) = 0.5$.

Now, to compute the denominator, we simply sum the expression in the numerator across all possible values for B_{t+2}:

$$\Pr(S_{t+2} = 73 | B_t = 75) = \sum_x \Pr(S_{t+2} = 73 | B_t = 75, B_{t+2} = x)\Pr(B_{t+2} = x | B_t = 75)$$

Following the same reasoning as we used for the numerator, this evaluates to:

$$= 0.25 \cdot \Pr(S_{t+2} = 73 | B_{t+2} = 73) + 0.5 \cdot \Pr(S_{t+2} = 73 | B_{t+2} = 75) + 0.25 \cdot \Pr(S_{t+2} = 73 | B_{t+2} = 77)$$

$$= 0.25 \cdot 0.3 + 0.5 \cdot 0.2 + 0.25 \cdot 0.1 = 0.075 + 0.1 + 0.025 = 0.2$$

So altogether, we have:

$$\Pr(B_{t+2} = 77 | B_t = 75, S_{t+2} = 73) = \frac{0.1 \cdot 0.25}{0.2} = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$$

b. (*4 points*) **Extra Practice - Gibbs Sampling**

Inference is exhausting; you decide that you'd be satisfied with simply being able to draw samples from distributions rather than specifying them exactly. In particular, you want to sample joint assignments to the variables $\{B_t, M_t, A_t, S_t\}_{t=1}^{T}$ for some time horizon T. You decide to implement Gibbs sampling for this purpose, but something's not right! What additional information, beyond what we've given you, would allow you to perform Gibbs sampling? Briefly explain.

Solution (The following argument applies identically to M_t as well as B_t): In order to sample B_t, we need to have first assigned a value to B_{t-1}; but in order to have sampled a value for B_{t-1}, we need to have first assigned a value to B_{t-2}, and so on. Continuing in this way, we realize that we must have a way of assigning a value to B_1 in order to perform Gibbs sampling. But to do this, we would either need to specify a fixed value for B_1, or specify a prior distribution $\Pr(B_1)$ from which to sample.

c. (*10 points*) **Particle Filtering**

Throughout this problem, you are free to leave quantities in terms of unevaluated expressions (i.e. you may write $0.75 \cdot 0.5$ instead of 0.375).

Computing all of those terms exactly seems tedious, so you instead decide to employ particle filtering to quickly and painlessly provide you with approximate solutions. You're fine with a (very) crude approximation, so you only use two particles.

(i) [2 points] Suppose you begin with the two particles $(B_1 = 80, M_1 = 75)$ and $(B_1 = 82, M_1 = 74)$. You then observe that $S_1 = 79$ and $A_1 = 154$. Compute the weights that you should assign to the two particles based on this evidence.

Solution For the first particle, we have $\Pr(A_1 = 154|B_1 = 80, M_1 = 75) = 0.13$ and $\Pr(S_1 = 79|B_1 = 80) = 0.25$. Thus, the first particle should get a weight of $0.13 * 0.25 = 0.0325$.

Similarly, for the second particle, we have $\Pr(A_1 = 154|B_1 = 82, M_1 = 74) = 0.11$ and $\Pr(S_1 = 79|B_1 = 82) = 0.15$. Thus, the second particle should get a weight of $0.11 * 0.15 = 0.0165$.

(ii) [2 points] Using these weights, we now resample two new particles. Provide this sampling distribution.

Probability of sampling a new particle to be $(B_1 = 80, M_1 = 75) =$

Solution $\frac{0.0325}{0.0325+0.0165}$

Probability of sampling a new particle to be $(B_1 = 82, M_1 = 74) =$

Solution $\frac{0.0165}{0.0325+0.0165}$

(iii) [3 points] Suppose both of our new particles are sampled to be $(B_1 = 80, M_1 = 75)$. We now extend these particles using our dynamics models. What is the probability that a particular one of these two particles is extended to:

$(B_1 = 80, M_1 = 75, B_2 = 78, M_2 = 76)$?

Solution Zero. Under the given model for $\Pr(B_{t+1}|B_t)$, the only possible values for B_2 are 79 and 81.

$(B_1 = 80, M_1 = 76, B_2 = 79, M_2 = 75)$?

Solution Zero. The value assigned to M_1 cannot change upon extending the particle.