14.7

These proofs are tricky for those not accustomed to manipulating probability expres

sions, and students may require some hints.

a. There are several ways to prove this. Probably the simplest is to work directly from the global semantics. First, we rewrite the required probability in terms of the full joint:

$$\begin{aligned}
P(x_i|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n) &= \frac{P(x_1,\ldots,x_n)}{P(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)} \\
&= \frac{P(x_1,\ldots,x_n)}{\sum_{x_i} P(x_1,\ldots,x_n)} \\
&= \frac{\prod_{j=1}^{n} P(x_j|parents X_j)}{\sum_{x_i}\prod_{j=1}^{n} P(x_j|parents X_j)}
\end{aligned}$$

Now, all terms in the product in the denominator that do not contain xi can be moved outside the summation, and then cancel with the corresponding terms in the numerator. This just leaves us with the terms that do mention x_i

those in which X_i is a child or a parent. Hence $P(x_i|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$ is equal to

$$\frac{P(x_i|parents X_i)\prod_{Y_j\in Children(X_i)} P(y_j|parents(Y_j))}{\sum_{x_i} P(x_i|parents X_i)\prod_{Y_j\in Children(X_i)} P(y_j|parents(Y_j))}$$

Now, by reversing the argument in part (b), we obtain the desired result.

b.

This is a relatively straightforward application of Bayes' rule. Let **Y** = Y1,... ,yℓ be the children of Xi and let **Z**j be the parents of Yj other than Xi . Then we have

$$\begin{aligned}
&\mathbf{P}(X_i|MB(X_i)) \\
&= \mathbf{P}(X_i|Parents(X_i),\mathbf{Y},\mathbf{Z}_1,\ldots,\mathbf{Z}_\ell) \\
&= \alpha\mathbf{P}(X_i|Parents(X_i),\mathbf{Z}_1,\ldots,\mathbf{Z}_\ell)\mathbf{P}(\mathbf{Y}|Parents(X_i),X_i,\mathbf{Z}_1,\ldots,\mathbf{Z}_\ell) \\
&= \alpha\mathbf{P}(X_i|Parents(X_i))\mathbf{P}(\mathbf{Y}|X_i,\mathbf{Z}_1,\ldots,\mathbf{Z}_\ell) \\
&= \alpha\mathbf{P}(X_i|Parents(X_i))\prod_{Y_j\in Children(X_i)} P(Y_j|Parents(Y_j))
\end{aligned}$$

where the derivation of the third line from the second relies on the fact that a node is independent of its nondescendants given its children.

14.14

a. (2),(3) can be asserted

b. $p(b, i, \not{m}, g, j) = p(b) * p(\not{m})p(i|b, \not{m}) * p(g|b, i, \not{m}) * p(j|g, b, i, \not{m}) = 0.2916$

c. $p(j) = p(g|b, i, m) * p(j|g) = 0.81$

e. A pardon is unnecessary if the person is not indicted or not found guilty; so I and G are parents of P. One could also add B and M as parents of P, since a pardon is more likely if the person is actually innocent and if the prosecutor is politically motivated. (There are other causes of Pardon, such as LargeDonationToPresidentsParty, but such variables are not currently in the model.) The pardon (presumably) is a get out-of-jail-free card, so P is a parent of J.

14.18

a. There are two uninstantiated Boolean variables (Cloudy and Rain) and therefore four possible states.

b. First, we compute the sampling distribution for each variable, conditioned on its Markov blanket

$$\begin{aligned}
\mathbf{P}(C|r, s) &= \alpha\mathbf{P}(C)\mathbf{P}(s|C)\mathbf{P}(r|C) \\
&= \alpha\langle 0.5, 0.5\rangle\langle 0.1, 0.5\rangle\langle 0.8, 0.2\rangle = \alpha\langle 0.04, 0.05\rangle = \langle 4/9, 5/9\rangle \\
\mathbf{P}(C|\neg r, s) &= \alpha\mathbf{P}(C)\mathbf{P}(s|C)\mathbf{P}(\neg r|C) \\
&= \alpha\langle 0.5, 0.5\rangle\langle 0.1, 0.5\rangle\langle 0.2, 0.8\rangle = \alpha\langle 0.01, 0.20\rangle = \langle 1/21, 20/21\rangle \\
\mathbf{P}(R|c, s, w) &= \alpha\mathbf{P}(R|c)\mathbf{P}(w|s, R) \\
&= \alpha\langle 0.8, 0.2\rangle\langle 0.99, 0.90\rangle = \alpha\langle 0.792, 0.180\rangle = \langle 22/27, 5/27\rangle \\
\mathbf{P}(R|\neg c, s, w) &= \alpha\mathbf{P}(R|\neg c)\mathbf{P}(w|s, R) \\
&= \alpha\langle 0.2, 0.8\rangle\langle 0.99, 0.90\rangle = \alpha\langle 0.198, 0.720\rangle = \langle 11/51, 40/51\rangle
\end{aligned}$$

Strictly speaking, the transition matrix is only well-defifined for the variant of MCMC in which the variable to be sampled is chosen randomly. (In the variant where the variables are chosen in a fifixed order, the transition probabilities depend on where we are in the ordering.) Now consider the transition matrix

Entries on the diagonal correspond to self-loops. Such transitions can occur by sampling *either* variable. For example,

$$q((c,r) \to (c,r)) = 0.5P(c|r,s) + 0.5P(r|c,s,w) = 17/27$$

Entries where one variable is changed must sample that variable. For example,

$$q((c,r) \to (c,\neg r)) = 0.5P(\neg r|c,s,w) = 5/54$$

Entries where both variables change cannot occur. For example,

$$q((c,r) \to (\neg c,\neg r)) = 0$$

This gives us the following transition matrix, where the transition is from the state given by the row label to the state given by the column label:

$$\begin{array}{l} \\ (c,r) \\ (c,\neg r) \\ (\neg c,r) \\ (\neg c,\neg r) \end{array} \begin{array}{c} \begin{array}{cccc} (c,r) & (c,\neg r) & (\neg c,r) & (\neg c,\neg r) \end{array} \\ \begin{pmatrix} 17/27 & 5/54 & 5/18 & 0 \\ 11/27 & 22/189 & 0 & 10/21 \\ 2/9 & 0 & 59/153 & 20/51 \\ 0 & 1/42 & 11/102 & 310/357 \end{pmatrix} \end{array}$$

c. Q^2 represents the probability of going from each state to each state in two steps.

d. Q^n (as n → ∞) represents the long-term probability of being in each state starting in each state; for ergodic **Q** these probabilities are independent of the starting state, so every row of **Q** is the same and represents the posterior distribution over states given the evidence.

e. We can produce very large powers of **Q** with very few matrix multiplications. For example, we can get Q^2 with one multiplication, Q^4 with two, and Q^2 with k. Unfortunately, in a network with n Boolean variables, the matrix is of size 2n × 2n , so each multiplication takes $O(2^{3n})$ operations.

14.21

a. we can create class **Team**, and create three instance A,B,C. Then we can create class **match**, with three instance AB,BC,CD. Each team has a quality Q, the probabilty of X win Y increases with Q(X)-Q(Y)

c.The exact result will depend on the probabilities used in the model. With any prior on quality that is the same across all teams, we expect that the posterior over BC.Outcome will show that C is more likely to win than B.

d. The inference cost in such a model will be $O(2^n)$ because all the team qualities become coupled.

e. MCMC appears to do well on this problem, provided the probabilities are not too skewed. Our results show scaling behavior that is roughly linear in the number of teams, although we did not investigate very large n.

20.9

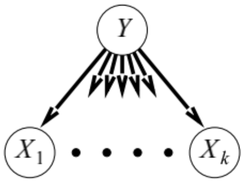
a. The probability of a positive example is π and of a negative example is $(1-\pi)$, and the data are independent, so the probability of the data is $\pi^p(1-\pi)^n$

b. We have $L = plog\pi + nlog(1-\pi)$; if the derivative is zero, we have

$$\frac{\partial L}{\partial \pi} = \frac{p}{\pi} - \frac{n}{1-\pi} = 0$$

so the ML value is $\pi = p/(p+n)$, i.e., the proportion of positive examples in the data.

c. This is the "naive Bayes" probability model.



d.

The likelihood of a single instance is a product of terms. For a positive example, π times α_i for each true attribute and $(1-\alpha_i)$ for each negative attribute; for a negative example, $(1-\pi)$ times β_i for each true attribute and $(1-\beta_i)$ for each negative attribute. Over the whole data set, the likelihood is $\pi^p(1-\pi)^n \prod_i \alpha_i^{p_i^+}(1-\alpha_i)^{n_i^+}\beta_i^{p_i^-}(1-\beta_i)^{n_i^-}$.

*f.** In the data set we have$p = 2, n = 2, p_i^+ = 1, n_i^+ = 1, p_i = 1, n_i = 1$. From our formulæ, we obtain π = α1 = α2 = β1 = β2 = 0.5.

g. Each example is predicted to be positive with probability 0.5.