8 Conditional Probability

(**8**) We would like to compute $\Pr(a, b|c, d)$ but we only have available to us the following quantities: $\Pr(a)$, $\Pr(b)$, $\Pr(c)$, $\Pr(a|d)$, $\Pr(b|d)$, $\Pr(c|d)$, $\Pr(d|a)$, $\Pr(a, b)$, $\Pr(c, d)$, $\Pr(a|c, d)$, $\Pr(b|c, d)$, $\Pr(c|a, b)$, $\Pr(d|a, b)$.

For each of the assumptions below, give a set of terms that is sufficient to compute the desired probability, or "none" if it can't be determined from the given quantities.

a. A and B are conditionally independent given C and D

b. C and D are conditionally independent given A and B

c. A and B are independent

d. A, B, and C are all conditionally independent given D

8 Conditional Probability

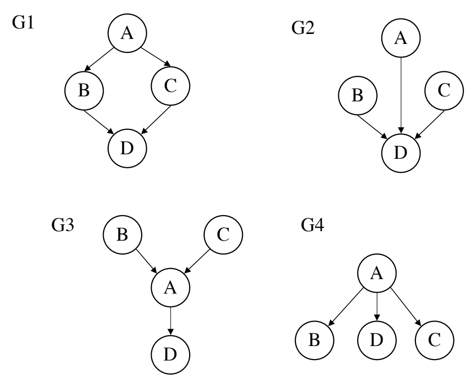
a. $\Pr(a|c, d)$, $\Pr(b|c, d)$

b. $\Pr(c|a, b)$, $\Pr(d|a, b)$, $\Pr(a, b)$, $\Pr(c, d)$

c. none

d. $\Pr(a|d)$, $\Pr(b|d)$

9 Network Structures



(**12**) Following is a list of conditional independence statements. For each statement, name all of the graph structures, G1 – G4, or "none" that imply it.

a. A is conditionally independent of B given C

b. A is conditionally independent of B given D

c. B is conditionally independent of D given A

d. B is conditionally independent of D given C

e. B is independent of C

f. B is conditionally independent of C given A

9 Network Structures

a. G2

b. none

c. G3, G4

d. none

e. G2, G3

f. G1, G2, G4

10 Counting Parameters

(**4**) How many independent parameters are required to specify a Bayesian network given each of the graph structures G1 – G4? Assume the nodes are binary.

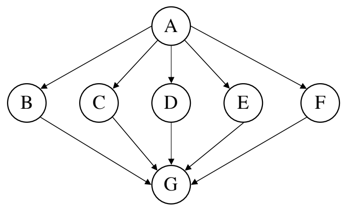
10 Counting Parameters

a. G1. 9

b. G2. 11

c. G3. 8

d. G4. 7



a. In this network, what is the size of the biggest factor that gets generated if we do variable elimination with elimination order A,B,C,D,E, F, G?

b. Give an elimination order that has a smaller largest factor.

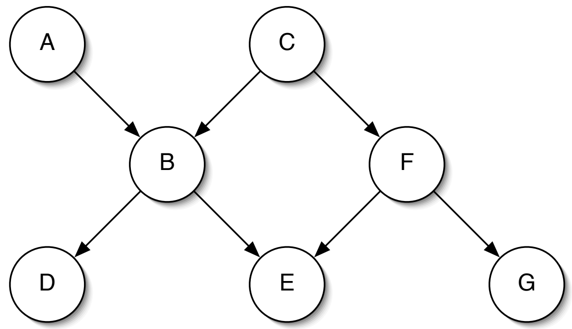
11 Variable Elimination

a. 5

b. B, C, D, E, F, A, G

1. (20 points)

Consider the following Bayesian network:



(a) (1 pt) Is it a polytree?

No

(b) (1 pt) Is A independent of C?

Yes

(c) (1 pt) Is C independent of E?

No

(d) (1 pt) Is D independent of C?

No

(e) (1 pt) Name a variable that, if it were an evidence variable, your answer to the question in part (b) would be different, or say that there is no such variable. (So, if your answer to (b) was that they are independent, then name a variable X for which A is not conditionally independent of C given X.)

B (makes A and C dependent)

(f) (1 pt) Name a variable that, if it were an evidence variable, your answer to part (c) would be different, or say that there is none.

No such (single) variable (2 variables B,F)

(g) (1 pt) Name a variable that, if it were an evidence variable, your answer to part (d) would be different, or say that there is none.

B

(h) (2 pts) If all the nodes are binary, how many parameters would be required to specify all the CPTs in this network? (Remember that if p is specified then it is not necessary to specify $1 - p$ as well.)

16

(i) (3 pts) Give an expression for $\Pr(D|C)$ given probabilities that are stored in the CPTs. Don't include any unnecessary terms.

$\Pr(D|C) = \sum_b \sum_a Pr(D|b) \times Pr(b|a, C) \times Pr(a)$

(j) (3 pts) What factor is created if we eliminate B first in the course of using variable elimination to compute $\Pr(A|G)$?

$f\{A, C, D, E, F\}$
There are many correct answers to this problem because A is independent of G.

(k) (2 pts) What is the Markov blanket of B?

A, C, D, E, F

(l) (3 pts) Imagine that you're doing likelihood weighting to compute $\Pr(E = e|A = a)$. What weight would you have to assign to sample $\langle a, b, c, d, f, g\rangle$?

$P(A = a)$

3. **Independence relations**

Draw a Bayesian network graph that encodes the following independence relations, or show that no such graph exists.

(a)
- A is independent of B
- A is independent of C given B
- A is not independent of C

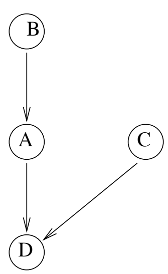
Answer: The last two statements taken together imply that the graph is fully connected, there is no direct link between A and C, and the structure where B is a common effect of A and C is not possible. This leaves three possible structures.

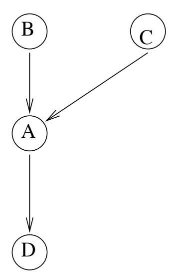
i. B is a common cause of A and C. But this is not possible because then A is not independent of B.

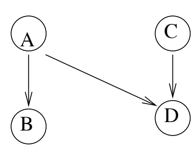
ii. The structures $A \rightarrow B \rightarrow C$ and $C \rightarrow B \rightarrow A$, neither of which are possible because A is not independent of B in either.

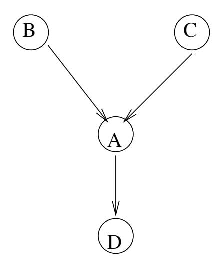
(b)
- D is independent of B given A
- B is independent of C
- B is not independent of D
- B is not independent of C given D

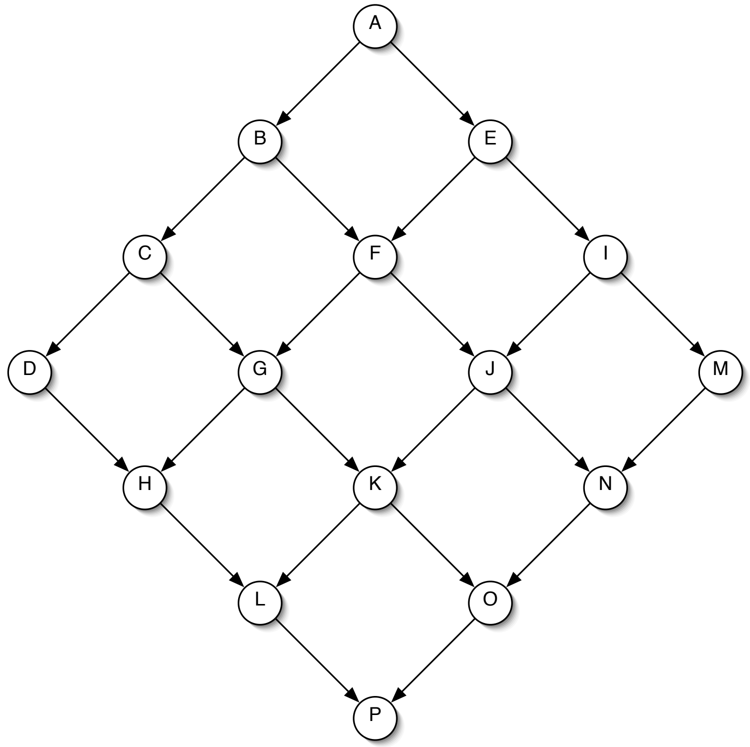
Answer: There are multiple such structures, including at least the following:











(a) What is the size of the largest CPT in this network?

Answer: No node has more than two parents, so the largest CPT is a function of three variables. Assuming the variables are binary, the CPT has $2^3 = 8$ entries. We may choose to only store 4 of these, taking advantage of the fact that the rows in the CPT sum to 1.

(b) What nodes can be ignored while computing $\Pr(H|M)$?

Answer: J, K, L, N, O, P.

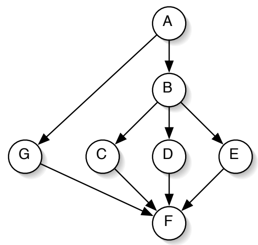
(c) Give a minimal expression for $\Pr(G|A)$ in terms of CPTs stored in the network.

Answer:

$$
\begin{aligned}
\Pr(G|A) &= \frac{\Pr(G,A)}{\Pr(A)} \\
&= \frac{\sum_{B,C,E,F} \Pr(G|C,F)\Pr(C|B)\Pr(F|B,E)\Pr(B|A)\Pr(E|A)\Pr(A))}{\Pr(A)} \\
&= \sum_{B,C,E,F} \Pr(G|C,F)\Pr(C|B)\Pr(F|B,E)\Pr(B|A)\Pr(E|A).
\end{aligned}
$$

1 Bayesian Networks

(22 points) Consider the network shown below:



1. (2 point) Is this a polytree?
2. (3 points) Assuming the nodes are binary, how many parameters are required to specify the CPTs?
3. (2 points) Is F independent of A given B?
4. (2 points) Is G independent of E given A and F?
5. (2 points) Is B independent of F given C, D, and E?
6. (3 points) Give an expression for $\Pr(d|c)$ (where d and c are specific values of variables D and C) in terms of parameters stored in the network?
7. (2 point) Which variables are irrelevant to the query $\Pr(d|c)$?
8. (4 points) What factors are created by variable elimination using order A, B, E, F, G?
9. (2 point) Is there another elimination order with a smaller largest factor?

1 Bayesian Networks

1. No.
2. Sum the sizes of the conditional probability tables of $\Pr(A), \Pr(B|A), \Pr(C|B), \Pr(D|B), \Pr(E|B), \Pr(F|C, D, E, G), \Pr(G|A)$ $= 2 + 4 + 4 + 4 + 4 + 32 + 4 = 54$
3. No.
4. No.
5. No.
6.

$$P(d|c) = P(c, d) \times P(c)$$

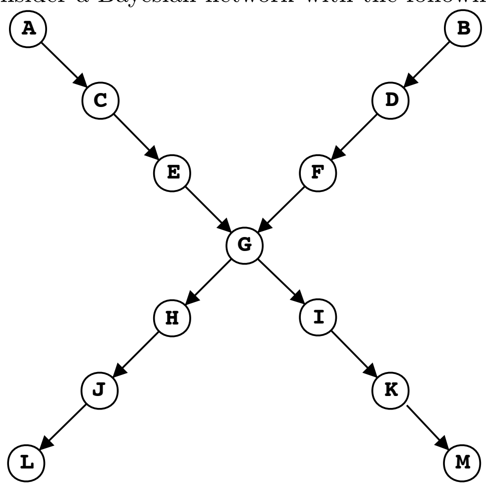
$$P(c, d) = \sum_B Pr(c|b)Pr(d|b)Pr(b) = \sum_B Pr(c|b)Pr(d|b)sum_A Pr(b|a)Pr(a)$$

$$Pr(c) = \sum_B Pr(c|b) \sum_A Pr(b|a)Pr(a)$$

7. Relevant variables: A, B, C, D.
   Irrelevant variables: E, F, G.
8. Factors created: F1(B,G), F2(C,D,E,G), F3(C,D,F,G), F4(C,D,G), F5(C,D)
9. No.

8 Bayesian Network Structure

Consider a Bayesian network with the following structure:



Does computing $P(M|A)$ depend on:

- $P(L|J)$?
- $P(K|I)$?
- $P(D|B)$?
- $P(H|G)$?

In the network above, if we decided not to include G in our network, but still wanted to model the joint distribution of all the other variables, what is the smallest network structure we could use?

8 Bayesian Network Structure

- No
- Yes
- Yes
- No

Remove node G.
Now node I has parents E, F, H.
Node H has parent, E, F.

1 Bayes' Nets

I am a professor trying to predict the performance of my class on an exam. After much thought, it is apparent that the students who do well are those that studied and do not have a headache when they take the exam. My vast medical knowledge leads me to believe that headaches can only be caused by being tired or by having the flu. Studying, the flu, and being tired are pairwise independent.

a) We will model everything with Boolean variables. F indicates the presence of the flu, T indicates being tired, H - having a headache, S - studying, and E - passing the exam. Which of the following three networks best models the relationships described?

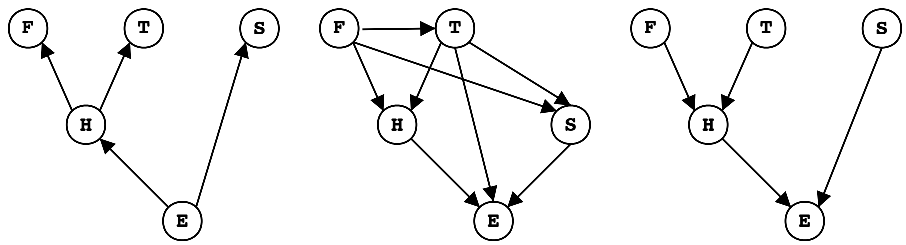


Figure 1: From left to right, models 1, 2, and 3

b) Why were the other two networks unsatisfactory models? Explain the deficiencies of each in terms of the conditional independence and dependence relationships they model. Which one of the remaining models represents an equivalent joint probability table as the best model, given that the description of the relationships was accurate?

c) I found that tiredness and having the flu each have a small impact on the likelihood of studying (small because MIT students are so tough). Draw a network that expresses this connection. Compute its complexity and the complexity of the network you choose in part a). Give two reasons why the original network is superior, despite the *small* improvement this new network gives in predictive power.

d) Leslie got the flu. Using model 3, compute the probability that she will fail the exam, in terms of values that are available in the conditional probability tables stored at each node.

e) Michael passed the exam. Using model 3, compute the probability that he studied, in terms of values that are available in the conditional probability tables stored at each node.

1 Bayes' Nets

I am a professor trying to predict the performance of my class on an exam. After much thought, it is apparent that the students who do well are those that studied and do not have a headache when they take the exam. My vast medical knowledge leads me to believe that headaches can only be caused by being tired or by having the flu. Studying, the flu, and being tired are pairwise independent.

a) Model 3 best models the relationship.

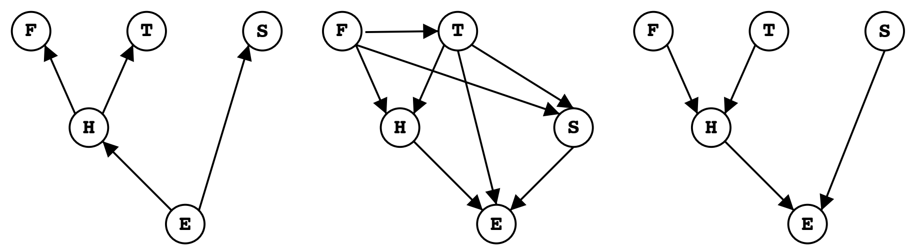
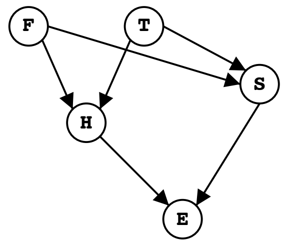


Figure 1: From left to right, models 1, 2, and 3

b) The first model makes flu, tiredness, and studying only conditionally independent (the first two conditioned on H and the third conditioned on E). The second model has the right relations, but many unnecessary dependencies. If the situation is accurately described in the problem, it will produce the same joint distribution as the third model because the dependencies will have no effect (for example, $P(S|F,T)$ will be the same as $P(S)$).

c) If we assume that we need to hold only one value, $P(X = true)$ to represent the apriori probability of a single variable, then we need 2^n entries for a conditional probability table for a node with n parents. So, the original network has a complexity of $1+1+1+4+4 = 11$ and the new network has a complexity of $1+1+4+4+4 = 14$. So, the size of the information

necessary to store the network has increased by about 27%, but we h
gained only a little more accuracy.



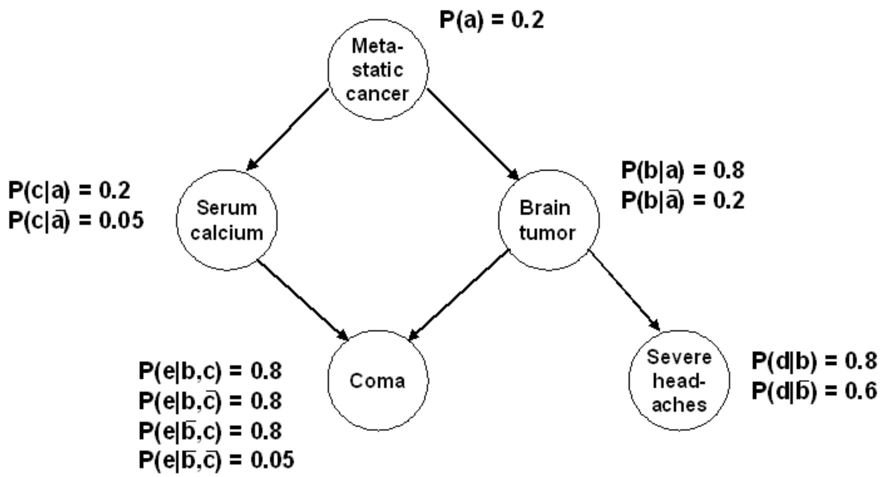
d)

$$
\begin{aligned}
P(\neg E|F) &= \sum_{H,S} P(\neg E|H,S,F)P(H,S|F) \\
&= \sum_{H,S} P(\neg E|H,S)P(H,S|F) \\
&= \sum_{H,S} P(\neg E|H,S)P(H|F,S)P(S|F) \\
&= \sum_{H,S} P(\neg E|H,S)P(H|F,S)P(S) \\
&= \sum_{H,S} P(\neg E|H,S)P(S)\sum_{T} P(H|F,T)P(T)
\end{aligned}
$$

e)

$$
\begin{aligned}
P(S|E) &= \frac{P(E|S)P(S)}{P(E)} \\
P(E|S) &= \sum_{H} P(E|S,H)P(H) \\
&= \sum_{H} P(E|S,H)\sum_{F,T} P(H|F,T)P(F)P(T) \\
P(E) &= \sum_{H,S} P(E|S,H)P(S)\sum_{F,T} P(H|F,T)P(F)P(T)
\end{aligned}
$$

Given is a simplified version of a network that could be used to diagnose patients arriving at a clinic. Each node in the network corresponds to some condition of the patient. This network demonstrates some cuasality links. For example, both brain tumor and serum calcium increase the chances of a coma. A brain tumor can cause severe headaches and a comma, and so on.



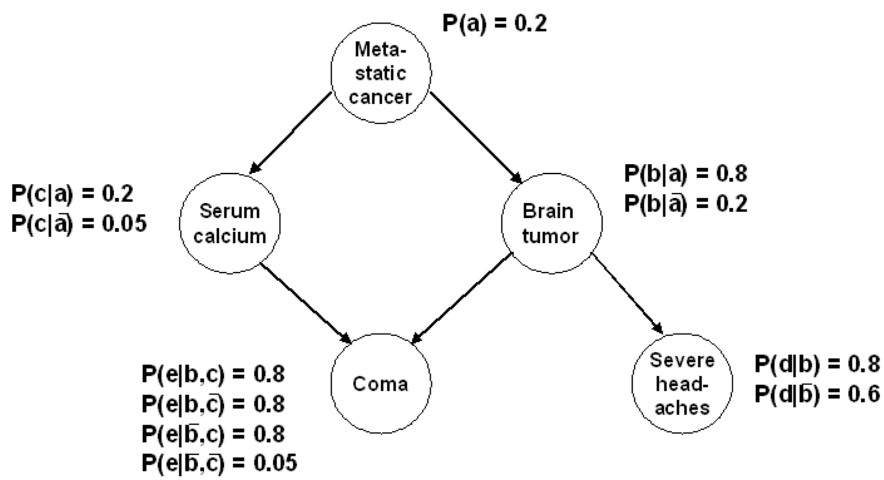
1. (2 points) What is the joint distribution $P(a, b, c, d, e)$? give a factorized expression, according to the network's structure.
2. (3 points) Give an example of 'explaining away' in this Bayes net.
3. (5 points) One of your patients experiences severe headaches, had a comma and serum calcium. What is the probability of him having cancer ? show full dereivation of this probability, as well as the numerical result.
4. (5 points) What is the probability of a positive serum calcium given severe headaches? Derive this expression. Specifically, start from the joint distribution, factorize it and use variable elimination, so as to lower calculation cost. (Note: a numerical result is *not* required here.)
5. (15 points) Write code for sampling joint and conditional probabilities in Bayes nets. Use the standard name "sample" for your code. The command line arguments should be: FILE-NAME (VAR-NAME=value,VAR-NAME=..) (VAR-NAME=value,VAR-NAME=..). (In case you are to use a non-standard language, be sure to included a README file with runtime instructions.)

   The first argument specifies the file describing the Bayes net. The second group of arguments gives the required variables' values, whose probability we'd like to evaluate. Several such values can be specified, using a comma separator within the same round brackets. The last group of arguments gives the variables' values that the requested probability is conditioned on (again, multiple values can be specified, using a comma separator and bounding brackets). If the second group is empty (unfilled brackets), this means we are looking for a joint probability expression, that is conditioned on nothing.

http://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15381-s07/www/hw5/

Problem 3 - Bayes Nets (35 points)

Given is a simplified version of a network that could be used to diagnose patients arriving at a clinic. Each node in the network corresponds to some condition of the patient. This network demonstrates some cuasality links. For example, both brain tumor and serum calcium increase the chances of a coma. A brain tumor can cause severe headaches and a comma, and so on.



1. (2 points) What is the joint distribution $P(a, b, c, d, e)$? give a factorized expression, according to the network's structure.
2. (3 points) Give an example of 'explaining away' in this Bayes net.
3. (5 points) One of your patients experiences severe headaches, had a comma and serum calcium. What is the probability of him having cancer ? show full dereivation of this probability, as well as the numerical result.
4. (5 points) What is the probability of a positive serum calcium given severe headaches? Derive this expression. Specifically, start from the joint distribution, factorize it and use variable elimination, so as to lower calculation cost. (Note: a numerical result is *not* required here.)

http://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15381-s07/www/hw5/

Solution 3

1. $p(a,b,c,d,e) = p(e|a,b,c,d)p(a,b,c,d) = p(e|a,b,c,d)p(d|a,b,c)p(c|b,a)p(b|a) = p(e|a,c)p(d|b)p(c|a)p(b|a)p(a)$

2. e is a common child to b and c. In this structure, two causes “compete” to “explain” the observed data. Hence b and c become conditionally dependent given that their common child, e, is observed, even though they are marginally independent. If we know that e (comma) is true and also b (brain tumor) is true, this reduces the probability that c (serum calcium) is true.

3.

$$P(a|d,e,c) = \frac{P(a,c,d,e)}{P(c,d,e)} = \frac{\sum_B P(a,B,c,d,e)}{\sum_{a,B} P(a,B,c,d,e)} \tag{1}$$

$$\propto \sum_B P(a,B,c,d,e) \tag{2}$$

$$\propto \sum_B P(a)P(B|a)P(c|a)P(d|B)P(e|B,c) \tag{3}$$

$$\propto P(a)P(c|a)\sum_B P(B|a)P(d|B)P(e|B,c) \tag{4}$$

$$\propto P(a)P(c|a)\left(P(b|a)P(d|b)P(e|b,c) + P(\bar{b}|a)P(d|\bar{b}P(e|\bar{b},c))\right) \tag{5}$$

$$\propto 0.2 \times 0.2\,(0.8 \times 0.8 \times 0.8 + 0.2 \times 0.6 \times 0.8) \tag{6}$$

$$\propto 0.02432 \tag{7}$$

$$P(\bar{a}|d,e,c) = \frac{P(\bar{a},c,d,e)}{P(c,d,e)} = \frac{\sum_B P(\bar{a},B,c,d,e)}{\sum_{\bar{a},B} P(\bar{a},B,c,d,e)} \tag{8}$$

$$\propto \sum_B P(\bar{a},B,c,d,e) \tag{9}$$

$$\propto \sum_B P(\bar{a})P(B|\bar{a})P(c|\bar{a})P(d|B)P(e|B,c) \tag{10}$$

$$\propto P(\bar{a})P(c|\bar{a})\sum_B P(e|B,c)P(B|\bar{a})P(d|B) \tag{11}$$

$$\propto P(\bar{a})P(c|\bar{a})\left(P(e|\bar{b},c)P(b|\bar{a})P(d|b) + P(e|b,c)P(\bar{b}|\bar{a})P(d|\bar{b})\right) \tag{12}$$

$$\propto 0.8 \times 0.05\,(0.8 \times 0.2 \times 0.8 + 0.8 \times 0.8 \times 0.6) \tag{13}$$

$$\propto 0.02048 \tag{14}$$

$$P(a|d,e,c) = \frac{\sum_B P(a,B,c,d,e)}{\sum_B P(a,B,c,d,e) + \sum_B P(\bar{a},B,c,d,e)} = \frac{0.02432}{0.02432 + 0.02048} = 0.54 \tag{15}$$

4.

$$P(c|d) = \frac{P(c,d)}{P(d)} = \frac{\sum_{A,B,E} P(A,B,c,d,E)}{\sum_{A,B,c,E} P(A,B,c,d,E)} \tag{16}$$

$$\propto \sum_{A,B,E} P(A,B,c,d,E) \tag{17}$$

$$\propto \sum_{A,B,E} P(A)P(B|A)P(c|A)P(d|B)P(E|B,c) \tag{18}$$

$$\propto \sum_{A} P(A)P(c|A) \sum_{B} P(B|A)P(d|B) \sum_{E} P(E|B,c) \tag{19}$$

$$\propto \sum_{A} P(A)P(c|A) \sum_{B} P(B|A)P(d|B) \tag{20}$$

Likewise,

$$P(\bar{c}|d) \propto \sum_{A} P(A)P(\bar{c}|A) \sum_{B} P(B|A)P(d|B) \tag{21}$$

Thus,

$$P(c|d) = \frac{\sum_A P(A)P(c|A) \sum_B P(B|A)P(d|B)}{\sum_A P(A)P(c|A) \sum_B P(B|A)P(d|B) + \sum_A P(A)P(\bar{c}|A) \sum_B P(B|A)P(d|B)} \tag{22}$$