Question 1 (Homework)

1. 

2. 

\[ P(A|C = c, S = s) = \alpha \sum_h P(A, h, c, s) = \alpha \sum_h P(c)P(h|c)P(s|h)P(A|h, s); \]

\[ \alpha = \frac{1}{\sum_a \sum_h P(c)P(h|c)P(s|h)P(a|h, s)} \]

| A  | \( \frac{1}{\alpha} P(A|c, s) \) | \( P(A|c, s) \) |
|----|---------------------------------|----------------|
| \( a \) | 0.3 \cdot 0.6 \cdot 0.9 \cdot 0.01 + 0.3 \cdot 0.1 \cdot 0.7 \cdot 0.5 + 0.3 \cdot 0.3 \cdot 0.3 \cdot 0.2 | 0.08 |
| \( \neg a \) | 0.3 \cdot 0.6 \cdot 0.9 \cdot 0.99 + 0.3 \cdot 0.1 \cdot 0.7 \cdot 0.5 + 0.3 \cdot 0.3 \cdot 0.3 \cdot 0.8 | 0.92 |

Question 2 (Homework)

a. A, C, E, F, G
b. A, C, D, E, F
c. A,B,E
d. A,D

There were several errors on this question. Please review the base cases below for determining conditional independence in a Bayes’ net.
Question 3 (Homework)

a. Under $G$, $P(x|y) = \alpha \sum_z P(x, y, z) = \alpha \sum_z P(x)P(y|x)P(z|y) = \alpha P(x)P(y|x)\sum_z P(z|y)$.

Since $P(Z|y)$ is a distribution over $Z$, $\sum_z P(z|y) = 1$. So, $P(x|y) = \alpha P(x)P(y|x)$.

Under $G'$, $P(x|y) = \alpha P(x, y) = \alpha P(x)P(y|x)$, which is the same as under $G$.

In both calculations, $\alpha$ is the normalizing factor $\sum_z P(x)P(y|x)$.

b. 

$$P(q_1, \ldots, q_k|e_1, \ldots, e_m) = \frac{\sum_{h_1} \ldots \sum_{h_p} P(q_1, \ldots, q_k, e_1, \ldots, e_m, h_1, \ldots, h_p)}{\sum_{h_1} \ldots \sum_{h_p} \sum_{q_1} \ldots \sum_{q_k} P(q_1, \ldots, q_k, e_1, \ldots, e_m, h_1, \ldots, h_p)}$$

c. Let $\pi(n)$ denote the parents of node $n$ in the Bayes net. Then, we can rewrite

$$P(q_1, \ldots, q_k, e_1, \ldots, e_m, h_1, \ldots, h_p) = \prod_{i=1}^{k} P(q_i|\pi(q_i)) \prod_{j=1}^{m} P(e_j|\pi(e_j)) \prod_{l=1}^{k} P(h_l|\pi(h_l))$$

which just follows from the semantics of Bayes nets. Notice that since $h_1$ is a leaf, none of the factors above contains $h_1$ except $P(h_1|\pi(h_1))$. So, we know that

$$P(q_1, \ldots, q_k|e_1, \ldots, e_m) = \alpha \sum_{h_1} \ldots \sum_{h_p} P(q_1, \ldots, q_k, e_1, \ldots, e_m, h_1, \ldots, h_p)$$

$$= \alpha \sum_{h_1} \ldots \sum_{h_p} \prod_{i=1}^{k} P(q_i|\pi(q_i)) \prod_{j=1}^{m} P(e_j|\pi(e_j)) \prod_{l=1}^{k} P(h_l|\pi(h_l))$$

$$= \alpha \sum_{h_2} \ldots \sum_{h_p} \prod_{i=1}^{k} P(q_i|\pi(q_i)) \prod_{j=1}^{m} P(e_j|\pi(e_j)) \prod_{l=2}^{k} P(h_l|\pi(h_l)) \sum_{h_1} P(h_1|\pi(h_1))$$

$$= \alpha \sum_{h_1} \ldots \sum_{h_p} \prod_{i=1}^{k} P(q_i|\pi(q_i)) \prod_{j=1}^{m} P(e_j|\pi(e_j)) \prod_{l=2}^{k} P(h_l|\pi(h_l))$$

Where the last step follows from the fact that $\sum_{h_1} P(h_1|\pi(h_1)) = 1$, much like in part (1). Note that the first line is just a restatement of part (2), where $\alpha$ is the normalizing factor (denominator).

d. If we successively remove hidden leaf nodes from the graph (which must be dangling), we will leave $P(q_1, \ldots, q_k|e_1, \ldots, e_m)$ unchanged by part (3). Let $d$ be a dangling node. Then, $n$ will eventually be pruned because all its descendants are also dangling. Before $n$ is pruned, there will always be at least one dangling leaf because the graph is acyclic.
Exercise 5

1. Let $M_n$ denote the monkey’s action at time $n$
Let $L_n$ denote the state of the lever at time $n$
Let $C_n$ denote the output of the factory at time $n$

2) We alter the state and set $L$ now equal to a 2-bit state representing the first lever for almonds and the second lever for coconuts. We say that $L_n = [L_A L_C]$. With this we define a new transition matrix from states and exclude the monkey from our directed graphical model.

<table>
<thead>
<tr>
<th>$L_A$</th>
<th>$L_C$</th>
<th>Type of Chocolate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Plain</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Coconut</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>Almond</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Almond+coconut</td>
</tr>
</tbody>
</table>
3) Since we know the emissions transmit with 100% probability, we can simply relabel the output and classify the correct states and determine the probability from these.

\[ L_1 = 00, L_2 = 10, L_3 = 10, L_4 = 11 \]

\[ P(L_1, L_2, L_3, L_4) = P(L_1)P(L_2|L_1)P(L_3|L_2)P(L_4|L_3) \]

We know that \( P(L_1) = 0.25 \) by two independent processes determining the initial state of each bit with probability 50%.

We compute the transition probabilities by factoring in the monkey’s behavior.

- 2 changes results in \( 0.3 \times 0.3 = 0.09 \) probability
- 1 change results in \( 0.7 \times 0.3 = 0.21 \) probability
- 0 changes results in \( 0.7 \times 0.7 = 0.49 \) probability

Thus, \( P(L_1, L_2, L_3, L_4) = (0.25)(0.21)(0.49)(0.21) = 0.005 \)

4) Let \( E_n \) be the emission of seeing a dark or light chocolate.

\[
P(L_1, L_2|E_1, E_2) \propto P(L_1, L_2, E_1, E_2) = \alpha P(L_1)P(L_2|L_1)P(E_1|L_1)P(E_2|L_2)
\]

\[
\alpha = \text{the normalization constant}
\]

\[
\alpha = \frac{1}{\sum_{L_1} \sum_{L_2} P(L_1)P(L_2|L_1)P(E_1|L_1)P(E_2|L_2)}
\]

\[
P(L_1, L_2|E_1, E_2) = \alpha(0.0126)
\]