Question 1

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

In both calculations, $\alpha$ is the normalizing factor $\frac{1}{\sum_x P(x) P(y|x)}$.

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

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

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

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

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

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

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

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

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 $P(q_1, ..., q_k|e_1, ..., e_m)$ 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.
Question 2

You sometimes get colds, which make you sneeze. You also get allergies, which make you sneeze. Sometimes you are well, which doesn’t make you sneeze. You decide to model the process using the following HMM, with hidden states \( X \in \{ \text{well, allergy, cold} \} \) and observations \( E \in \{ \text{well, allergy, cold} \} \):

<table>
<thead>
<tr>
<th>( X_t )</th>
<th>( P(X_1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>well</td>
<td>1</td>
</tr>
<tr>
<td>allergy</td>
<td>0</td>
</tr>
<tr>
<td>cold</td>
<td>0</td>
</tr>
</tbody>
</table>

| \( X_{t-1} = \text{well} \) | \( P(X_t | X_{t-1} = \text{well}) \) |
|----------------------------|----------------------------------|
| well                      | 0.7                              |
| allergy                   | 0.2                              |
| cold                      | 0.1                              |

| \( X_{t-1} = \text{allergy} \) | \( P(X_t | X_{t-1} = \text{allergy}) \) |
|-------------------------------|---------------------------------------|
| well                         | 0.6                                   |
| allergy                      | 0.3                                   |
| cold                         | 0.1                                   |

| \( X_{t-1} = \text{cold} \) | \( P(X_t | X_{t-1} = \text{cold}) \) |
|-----------------------------|-----------------------------------|
| well                        | 0.2                                |
| allergy                     | 0.2                                |
| cold                        | 0.6                                |

Note that colds are “stickier” in that you tend to have them for multiple days, while allergies come and go on a quicker time scale. However, allergies are more frequent. Assume that on the first day, you are well.

a) Imagine that you observe the sequence quiet, sneeze, sneeze. What is the probability that you were well all three days and observed these effects?

\[
P(X_1=\text{well}, X_2=\text{well}, X_3=\text{well} \mid E_1=\text{quiet}, E_2=\text{sneeze}, E_3=\text{sneeze}) = 0
\]

since it is proportional to \( P(E_3=\text{sneeze} \mid X_3=\text{well}) = 0 \) in the product.

b) What is the posterior distribution over your state on day 2 \( X_2 \) if \( E_1 = \text{quiet} \), \( E_2 = \text{sneeze} \)?

This is following: \( P(x_2 | e_{1:2}) \propto P(x_2, e_{1:2}) = \sum_{x_1} P(x_1, x_2, e_{1:2}) \)

\[
= \sum_{x_1} P(x_2 | x_1) P(x_1) P(e_1 | x_1) P(e_2 | x_2) = P(x_2 | x_1=\text{well}) P(x_1=\text{well}) P(E_1=\text{quiet}) P(E_2=\text{sneeze})
\]

\[
\text{since } P(X_1=\text{well}) > P(X_1=\text{allergy}) > 0
\]

[see next page for cont.]


**c) What is the posterior distribution over your state on day 3 ($X_3$) if $E_3 = \text{quiet}$, $E_2 = \text{sneeze}$, $E_1 = \text{sneeze}$?**

\[
P(X_3 | E_1, E_2) \propto P(E_3 = \text{quiet} | X_3) P(X_3 = \text{quiet})
\]

\[
\begin{align*}
\text{constant} & \quad \begin{bmatrix} 0.47 & 0.27 \\
0.02 & 0.27 \end{bmatrix} = \begin{bmatrix} 0.1 \\
0.05 \end{bmatrix} \\
\end{align*}
\]

**d) What is the Viterbi (most likely) sequence for the observation sequence quiet, sneeze, sneeze, sneeze, quiet, quiet?** [Hint: you should not have to do extensive calculations.]


**e) If you observe sneeze, what weight will each of your particles have?**

\[
w(x_6 = \text{well}) = P(x_6 = \text{sneeze} | x_6 = \text{well}) = 0
\]

\[
w(x_6 = \text{cold}) = P(1 | x_6 = \text{cold}) = 1
\]

\[
w(x_6 = \text{allergy}) = P(1 | x_6 = \text{allergy}) = 1
\]

**f) After resampling, what is the expected number of particles you will have on cold?**

- Total weight on cold was $w(x_6 = \text{cold}) \cdot N_{\text{cold}} = 2$
- Total weight of all particles was $0.5 + 2.1 + 3.1 = 5$
- So prob. of sampling cold is $\frac{2}{5}$
- You sample 10 particles, so expected # of particles on cold is $\frac{2}{5} \cdot 10 = 4$