1. (Modified version of Exercise 3.1 in the textbook) Suppose you have three Boolean random variables $X_1, X_2, X_3$. Write out a discrete joint probability distribution $P(X_1, X_2, X_3)$ where for each $i \neq j$, we have that $(X_i \perp X_j) \in I(P)$, but we also have that $((X_1, X_2) \perp X_3) \notin I(P)$. This should be a table with $2^3$ rows (one for each combination of values for the three random variables). Explain your solution. [5 points]

You will need to create a probability distribution where $P(X_i, X_j) = P(X_i)P(X_j)$ and $P(X_1, X_2, X_3) \neq P(X_1, X_2)P(X_3)$. There are many possible solutions. One example is the XOR function in which the joint distribution has probability $1/4$ if there are an odd number of “true”s and 0 otherwise. Note that in the XOR example, knowing all three random variables gives you the value of the joint distribution but knowing only two of them doesn’t.

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$P(X_1, X_2, X_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>false</td>
<td>0</td>
</tr>
<tr>
<td>false</td>
<td>false</td>
<td>true</td>
<td>0.25</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>false</td>
<td>0.25</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>0</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>false</td>
<td>0.25</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>true</td>
<td>0</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>false</td>
<td>0</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>true</td>
<td>0.25</td>
</tr>
</tbody>
</table>

To see that $(X_i \perp X_j) \in I(P)$

\[
P(X_1 = true, X_2 = true) = P(X_1 = true)P(X_2 = true) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}
\]

\[
P(X_2 = true, X_3 = true) = P(X_2 = true)P(X_3 = true) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}
\]

\[
P(X_1 = true, X_3 = true) = P(X_1 = true)P(X_3 = true) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}
\]

To see that $((X_1, X_2) \perp X_3) \notin I(P)$.

\[
P(X_1 = true, X_2 = true, X_3 = true) \neq P(X_1 = true, X_2 = true)P(X_3 = true)
\]

\[
= \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}
\]
2. Answer true or false to the following conditional independence statements using the graph below. For partial credit, show the paths that are blocked or not blocked. [15 points]
   a) C \perp D | \{B,H\}
      False. H is an evidence node that is a descendant of E which is in the v-structure C→E←D
   
   b) I \perp C
      True.
   
   c) I \perp H | F
      True.
   
   d) B \perp E | \{D,G\}
      False. The path B →C →E is still active.
   
   e) A \perp I | \{E\}
      False. E is a descendant of D which is in the v-structure B →D←I.

3. (Exercise 3.7 in the textbook) Show how you could efficiently compute the distribution over a variable \(X_i\) given some assignment to all the other variables in the network:
\[ P(X_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n). \]
Your procedure should not require the construction of the entire joint distribution \(P(X_1, \ldots, X_n)\). Specify the computational complexity of your procedure using Big-O notation. [10 points]
4. (Modified version of Exercise 3.8 in the textbook) Let $B = (G, P)$ be a Bayesian network over some set of variables $V$. Let us denote the nodes observed as evidence as the set $Z$, and let $X$ be all of the ancestors of the nodes in $Z$ (i.e. the parents of $Z$ as well as their ancestors). Let $W = V - X - Z$ (i.e. the remaining nodes not in $X \cup Z$).

To simplify things, suppose the set $W$ consists of a single node $w$. Now pick the node $w$ and remove it from the graph $G$ for Bayesian network $B$. By “remove a node” we mean delete that node $w$ along with all edges to/from it; we leave the conditional probability tables for nodes in $X \cup Z$ as they are. Let us call the remaining graph (after removal of $w$) $G'$ and the resulting Bayesian network $B'$.

The node $w$ is called a barren node relative to $X \cup Z$ because it is irrelevant to probability computations concerning $X \cup Z$. Prove it is irrelevant by showing that $P(X|Z)$ in Bayesian network $B$ is equal to $P(X|Z)$ in Bayesian network $B'$.

[10 points]

Let $w = V - X - Z$ be the barren node.

1) Case 1: we have $X \rightarrow Z \rightarrow w$

From Bayesian network $B$,

$$P(X|Z) = \frac{P(X,Z)}{P(Z)} = \frac{\sum_w P(X,Z,W)}{\sum_{X,W} P(Z)} = \frac{\sum_w P(X)P(Z|X)P(W|Z)}{\sum_X \sum_w P(X)P(Z|X)P(W|Z)}$$

$$= \frac{P(X)P(Z|X)\sum_w P(W|Z)}{\sum_X P(X)P(Z|X)\sum_w P(W|Z)} = \frac{P(X)P(Z|X)}{P(Z)}$$

From Bayesian network $B'$,
\[
P(X|Z) = \frac{P(X,Z)}{P(Z)} = \frac{\sum_w P(X,Z,W)}{P(Z)} = \frac{\sum_w P(X)P(Z|X)P(W|Z)}{P(Z)} = \frac{P(X)P(Z|X)}{P(Z)}
\]

Note: an even easier way is to use conditional independence \(P(X|Z) = \sum_w P(X,W|Z) = \sum_w P(X|Z)P(W|Z) = P(X|Z) \sum_w P(W|Z) = P(X|Z)\). You can see from this that the \(W\) nodes don’t matter.

2) Case 2: we have \(W \leftarrow X \rightarrow Z\)

From Bayesian network \(B\):
\[
P(X|Z) = \frac{\sum_w P(X,W,Z)}{P(Z)} = \frac{\sum_w P(X)P(W|X)P(Z|X)}{P(Z)} = \frac{P(Z|X)P(X) \sum_w P(W|X)}{P(Z)}
\]

From Bayesian network \(B'\):
\[
P(X|Z) = \frac{P(X,Z)}{P(Z)} = \frac{P(Z|X)P(X)}{P(Z)}
\]