1 Variable Elimination

Using the Bayes Net shown below, we want to compute $P(Y \mid +z)$. All variables have binary domains. We run variable elimination, with the following variable elimination ordering: $X, T, U, V, W$.

After inserting evidence, we have the following factors to start out with:

$$P(T), P(U\mid T), P(V\mid T), P(W\mid T), P(X\mid T), P(Y\mid V, W), P(\pmz X)$$

(a) When eliminating $X$ we generate a new factor $f_1$ as follows,

$$f_1(\pmz T) = \sum_x P(x\mid T)P(\pmz x)$$

which leaves us with the factors:

$$P(T), P(U\mid T), P(V\mid T), P(W\mid T), P(Y\mid V, W), f_1(\pmz T)$$

(b) When eliminating $T$ we generate a new factor $f_2$ as follows, which leaves us with the factors:

$$f_2(U, V, W, +z) = \sum_t P(t)P(U\mid t)P(V\mid t)P(W\mid t)f_1(\pmz t) \quad P(Y\mid V, W), f_2(U, V, W, +z)$$

(c) When eliminating $U$ we generate a new factor $f_3$ as follows, which leaves us with the factors:

$$f_3(V, W, +z) = \sum_u f_2(u, V, W, +z) \quad P(Y\mid V, W), f_3(V, W, +z)$$

Note that $U$ could have just been deleted from the original graph, because $\sum_u P(U\mid t) = 1$. We can see this in the graph: we can remove any leaf node that is not a query variable or an evidence variable.

(d) When eliminating $V$ we generate a new factor $f_4$ as follows, which leaves us with the factors:

$$f_4(W, Y, +z) = \sum_v f_3(v, W, +z)P(Y\mid v, W) \quad f_4(W, Y, +z)$$

(e) When eliminating $W$ we generate a new factor $f_5$ as follows, which leaves us with the factors:

$$f_5(Y, +z) = \sum_w f_4(w, Y, +z) \quad f_5(Y, +z)$$

(f) How would you obtain $P(Y \mid +z)$ from the factors left above: Simply renormalize $f_5(Y, +z)$ to obtain $P(Y \mid +z)$. Concretely,

$$P(y \mid +z) = \frac{f_5(y, +z)}{\sum_{y'} f_5(y', +z)}$$
(g) What is the size of the largest factor that gets generated during the above process? $f_2(U, V, W, +z)$. This contains 3 unconditioned variables, so it will have $2^3 = 8$ entries ($U, V, W$ are binary variables, and we only need to store the entries for $+z$ for each possible setting of these variables).

(h) Does there exist a better elimination ordering (one which generates smaller largest factors)? Yes. One such ordering is $X, U, T, V, W$. All factors generated with this ordering contain at most 2 unconditioned variables, so the tables will have at most $2^2 = 4$ entries (as all variables are binary).
2 Sampling and Dynamic Bayes Nets

We would like to analyze people’s ice cream eating habits on sunny and rainy days. Suppose we consider the weather, along with a person’s ice-cream eating, over the span of two days. We’ll have four random variables: $W_1$ and $W_2$ stand for the weather on days 1 and 2, which can either be rainy $R$ or sunny $S$, and the variables $I_1$ and $I_2$ represent whether or not the person ate ice cream on days 1 and 2, and take values $T$ (for truly eating ice cream) or $F$. We can model this as the following Bayes Net with these probabilities.

![Bayes Net Diagram]

| $W_1$ | $P(W_1)$ | $W_2$ | $P(W_2|W_1)$ |
|-------|----------|-------|---------------|
| $S$   | 0.6      | $S$   | 0.7           |
| $R$   | 0.4      | $R$   | 0.3           |
|       |          | $S$   | 0.5           |
|       |          | $R$   | 0.5           |

Suppose we produce the following samples of $(W_1, I_1, W_2, I_2)$ from the ice-cream model:


1. What is $P(W_2 = R)$, the probability that sampling assigns to the event $W_2 = R$?
   Number of samples in which $W_2 = R$: 5. Total number of samples: 10. Answer $5/10 = 0.5$.

2. Cross off samples above which are rejected by rejection sampling if we’re computing $P(W_2| I_1 = T, I_2 = F)$.

   Rejection sampling seems to be wasting a lot of effort, so we decide to switch to likelihood weighting.
   Assume we generate the following six samples given the evidence $I_1 = T$ and $I_2 = F$:

   $(W_1, I_1, W_2, I_2) = \{ (S, T, R, F), (R, T, R, F), (S, T, R, F), (S, T, S, F), (S, T, S, F), (R, T, S, F) \}$

3. What is the weight of the first sample $(S, T, R, F)$ above?

   The weight given to a sample in likelihood weighting is

   $\prod \text{Pr}(e|\text{Parents}(e))$.

   Evidence variables $e$

   In this case, the evidence is $I_1 = T, I_2 = F$. The weight of the first sample is therefore

   $w = \text{Pr}(I_1 = T|W_1 = S) \cdot \text{Pr}(I_2 = F|W_2 = R) = 0.9 \cdot 0.8 = 0.72$

4. Use likelihood weighting to estimate $P(W_2| I_1 = T, I_2 = F)$.

   The sample weights are given by

<table>
<thead>
<tr>
<th>$(W_1, I_1, W_2, I_2)$</th>
<th>$w$</th>
<th>$(W_1, I_1, W_2, I_2)$</th>
<th>$w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S, T, R, F$</td>
<td>0.72</td>
<td>$S, T, S, F$</td>
<td>0.09</td>
</tr>
<tr>
<td>$R, T, R, F$</td>
<td>0.16</td>
<td>$S, T, S, F$</td>
<td>0.09</td>
</tr>
<tr>
<td>$S, T, R, F$</td>
<td>0.72</td>
<td>$R, T, S, F$</td>
<td>0.02</td>
</tr>
</tbody>
</table>

   To compute the probabilities, we thus normalize the weights and find

   $\hat{P}(W_2 = R| I_1 = T, I_2 = F) = \frac{0.72 + 0.16 + 0.72}{0.72 + 0.16 + 0.72 + 0.09 + 0.09 + 0.02} = 0.889$

   $\hat{P}(W_2 = S| I_1 = T, I_2 = F) = 1 - 0.889 = 0.111$. 

3