Q1. Bayes’ Nets Sampling

Assume the following Bayes’ net, and the corresponding distributions over the variables in the Bayes’ net:

(a) You are given the following samples:

\[
\begin{array}{cccc}
+a & +b & -c & -d \\
+a & -b & +c & -d \\
-a & +b & +c & -d \\
-a & -b & +c & -d \\
\end{array}
\]

(i) Assume that these samples came from performing Prior Sampling, and calculate the sample estimate of \( P(\mathcal{+} | \mathcal{b}) \).

\[ \frac{5}{8} \]

(ii) Now we will estimate \( P(\mathcal{+} | \mathcal{a}, \mathcal{-} \mathcal{d}) \). Above, clearly cross out the samples that would not be used when doing Rejection Sampling for this task, and write down the sample estimate of \( P(\mathcal{+} | \mathcal{a}, \mathcal{-} \mathcal{d}) \) below.

\[ \frac{2}{3} \]

(b) Using Likelihood Weighting Sampling to estimate \( P(\mathcal{-} \mathcal{a} | \mathcal{+} \mathcal{b}, \mathcal{-} \mathcal{d}) \), the following samples were obtained. Fill in the weight of each sample in the corresponding row.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-a +b +c -d)</td>
<td>( P(+b</td>
</tr>
<tr>
<td>(+a +b +c -d)</td>
<td>( P(+b</td>
</tr>
<tr>
<td>(+a +b -c -d)</td>
<td>( P(+b</td>
</tr>
<tr>
<td>(-a +b -c -d)</td>
<td>( P(+b</td>
</tr>
</tbody>
</table>

(c) From the weighted samples in the previous question, estimate \( P(\mathcal{-} \mathcal{a} | \mathcal{+} \mathcal{b}, \mathcal{-} \mathcal{d}) \).

\[ \frac{5/18+1/24}{5/18+5/30+1/40+1/24} = 0.625 \]

(d) Which query is better suited for likelihood weighting, \( P(\mathcal{D} | \mathcal{A}) \) or \( P(\mathcal{A} | \mathcal{D}) \)? Justify your answer in one sentence.
$P(D \mid A)$ is better suited for likelihood weighting sampling, because likelihood weighting conditions only on upstream evidence.

(e) Recall that during Gibbs Sampling, samples are generated through an iterative process.

Assume that the only evidence that is available is $A = +a$. Clearly fill in the circle(s) of the sequence(s) below that could have been generated by Gibbs Sampling.

<table>
<thead>
<tr>
<th>Sequence 1</th>
<th>Sequence 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: $+a -b -c +d$</td>
<td>1: $+a -b -c +d$</td>
</tr>
<tr>
<td>2: $+a -b -c +d$</td>
<td>2: $+a -b -c -d$</td>
</tr>
<tr>
<td>3: $+a -b +c +d$</td>
<td>3: $-a -b -c +d$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 3</th>
<th>Sequence 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: $+a -b -c +d$</td>
<td>1: $+a -b -c +d$</td>
</tr>
<tr>
<td>2: $+a -b -c -d$</td>
<td>2: $+a -b -c -d$</td>
</tr>
<tr>
<td>3: $+a +b -c -d$</td>
<td>3: $+a +b -c +d$</td>
</tr>
</tbody>
</table>

Gibbs sampling updates one variable at a time and never changes the evidence.

The first and third sequences have at most one variable change per row, and hence could have been generated from Gibbs sampling. In sequence 2, the evidence variable is changed. In sequence 4, the second and third samples have both $B$ and $D$ changing.
2 Particle Filtering

Let’s use Particle Filtering to estimate the distribution of $P(W_2 | O_1 = a, O_2 = b)$. Here’s the HMM again. $O_1$ and $O_2$ are supposed to be shaded.

We start with two particles representing our distribution for $W_1$.

\[ P_1 : W_1 = 0 \]
\[ P_2 : W_1 = 1 \]

Use the following random numbers to run particle filtering:

\[ [0.22, 0.05, 0.33, 0.20, 0.84, 0.54, 0.79, 0.66, 0.14, 0.96] \]

(a) **Observe**: Compute the weight of the two particles after evidence $O_1 = a$.

\[
w(P_1) = P(O_t = a | W_t = 0) = 0.9
\]
\[
w(P_2) = P(O_t = a | W_t = 1) = 0.5
\]

(b) **Resample**: Using the random numbers, resample $P_1$ and $P_2$ based on the weights.

We now sample from the weighted distribution we found above. Using the first two random samples, we find:

\[
P_1 = \text{sample(weights, 0.22)} = 0
\]
\[
P_2 = \text{sample(weights, 0.05)} = 0
\]

(c) **Predict**: Sample $P_1$ and $P_2$ from applying the time update.

\[
P_1 = \text{sample}(P(W_{t+1} | W_t = 0), 0.33) = 0
\]
\[
P_2 = \text{sample}(P(W_{t+1} | W_t = 0), 0.20) = 0
\]

(d) **Update**: Compute the weight of the two particles after evidence $O_2 = b$.

\[
w(P_1) = P(O_t = b | W_t = 0) = 0.1
\]
\[
w(P_2) = P(O_t = b | W_t = 0) = 0.1
\]

(e) **Resample**: Using the random numbers, resample $P_1$ and $P_2$ based on the weights.

Because both of our particles have $X = 0$, resampling will still leave us with two particles with $X = 0$.

\[ P_1 = 0 \]
\[ P_2 = 0 \]
(f) What is our estimated distribution for $P(W_2|O_1 = a, O_2 = b)$?

\[
P(W_2 = 0|O_1 = a, O_2 = b) = \frac{2}{2} = 1
\]

\[
P(W_2 = 1|O_1 = a, O_2 = b) = \frac{0}{2} = 0
\]
Consider a modified version of the apprenticeship problem. We are observing an agent’s actions in an MDP and are trying to determine which out of a set \( \{\pi_1, \ldots, \pi_n\} \) the agent is following. Let the random variable \( \Pi \) take values in that set and represent the policy that the agent is acting under. We consider only stochastic policies, so that \( A_t \) is a random variable with a distribution conditioned on \( S_t \) and \( \Pi \). As in a typical MDP, \( S_t \) is a random variable with a distribution conditioned on \( S_{t-1} \) and \( A_{t-1} \). The full Bayes net is shown below.

The agent acting in the environment knows what state it is currently in (as is typical in the MDP setting). Unfortunately, however, we, the observer, cannot see the states \( S_t \). Thus we are forced to use an adapted particle filtering algorithm to solve this problem. Concretely, we will develop an efficient algorithm to estimate \( P(\Pi | a_1: t) \).

(a) The Bayes net for part (a) is

![Bayes Net Diagram]

We will compute our estimate for \( P(\Pi | a_1: t) \) by coming up with a recursive algorithm for computing \( P(\Pi, S_t | a_1: t) \). (We can then sum out \( S_t \) to get the desired distribution; in this problem we ignore that step.)

(ii) Write a recursive expression for \( P(\Pi, S_t | a_1: t) \) in terms of the CPTs in the Bayes net above. Hint: Think of the forward algorithm.

\[
P(\Pi, S_t | a_1: t) \propto \sum_{s_{t-1}} P(\Pi, s_{t-1} | a_{1:t-1}) P(a_t | S_t, \Pi) P(S_t | s_{t-1}, a_{t-1})
\]

We now try to adapt particle filtering to approximate this value. Each particle will contain a single state \( s_t \) and a potential policy \( \pi_t \).

(iii) The following is pseudocode for the body of the loop in our adapted particle filtering algorithm. Fill in the boxes with the correct values so that the algorithm will approximate \( P(\Pi, S_t | a_1: t) \).

1. Elapse time: for each particle \((s_t, \pi_t)\), sample a successor \(s_{t+1}\) from
\[
P(S_{t+1} \mid s_t, a_t)
\]
The policy \(\pi'\) in the new particle is \(\pi_t\).

2. Incorporate evidence: To each new particle \((s_{t+1}, \pi')\), assign weight
\[
P(a_{t+1} \mid s_{t+1}, \pi')
\]

3. Resample particles from the weighted particle distribution.

(b) We now observe the acting agent’s actions and rewards at each time step (but we still don’t know the states). Unlike the MDPs in lecture, here we use a stochastic reward function, so that \(R_t\) is a random variable with a distribution conditioned on \(S_t\) and \(A_t\). The new Bayes net is given by

Notice that the observed rewards do in fact give useful information since d-separation does not give that \(R_t \perp \perp \Pi \mid A_{1:t}\). Give an active path connecting \(R_t\) and \(\Pi\) when \(A_{1:t}\) are observed. Your answer should be an ordered list of nodes in the graph, for example “\(S_t, S_{t+1}, A_t, \Pi, A_{t-1}, R_{t-1}\)”.

\(R_t, S_t, A_t, \Pi\). This list reversed is also correct, and many other similar (though more complicated) paths are also correct.
(c) We now observe only the sequence of rewards and no longer observe the sequence of actions. The new Bayes net is:

```
We will compute our estimate for \( P(\Pi | r_1:t) \) by coming up with a recursive algorithm for computing \( P(\Pi, S_t, A_t | r_1:t) \). (We can then sum out \( S_t \) and \( A_t \) to get the desired distribution; in this problem we ignore that step.)

(i) Write a recursive expression for \( P(\Pi, S_t, A_t | r_1:t) \) in terms of the CPTs in the Bayes net above.
```

\[
P(\Pi, S_t, A_t | r_1:t) \propto \sum_{s_{t-1}} \sum_{a_{t-1}} P(\Pi, s_{t-1}, a_{t-1} | r_{1:t-1})P(A_t | S_t, \Pi)P(S_t | s_{t-1}, a_{t-1})P(r_t | S_t, A_t)
\]

We now try to adapt particle filtering to approximate this value. Each particle will contain a single state \( s_t \), a single action \( a_t \), and a potential policy \( \pi_t \).

(ii) The following is pseudocode for the body of the loop in our adapted particle filtering algorithm. Fill in the boxes with the correct values so that the algorithm will approximate \( P(\Pi, S_t, A_t | r_{1:t}) \).

1. Elapse time: for each particle \((s_t, a_t, \pi_t)\), sample a successor state \( s_{t+1} \) from

\[
P(S_{t+1} | s_t, a_t)
\]

Then, sample a successor action \( a_{t+1} \) from

\[
P(A_{t+1} | s_{t+1}, \pi_t)
\]

The policy \( \pi' \) in the new particle is \( \pi_t \).

2. Incorporate evidence: To each new particle \((s_{t+1}, a_{t+1}, \pi')\), assign weight

\[
P(r_{t+1} | s_{t+1}, a_{t+1})
\]

3. Resample particles from the weighted particle distribution.