Assignment 3, Part 2 - Report

MDL

Team : **Room543**
Mihir Bani, 2019113003
Amul Agrawal, 2019101113

Roll Number used : 2019113003

$x = 1 - (3003 \% 30 + 1)/100 = 0.96$

General:

There will be 128 (= 8*8*2) states total. The given state space is a grid of 2x4.
The agent (**A**), and target (**T**) can be on any of these cells.
And the call (**C**) can be *on(1)* or *off (0)*.
Each state is represented by the string :

$$S\_ < A.row,\ A.col > \_ < T.row,\ T.col > \_ < C >$$

Where $< x >$ denotes the value of *x*.
Example : S_01_12_0. When **A** is at (0,1) and **T** is at (1,2) and **C** = *off*

(row, col) = (0,0)	(0,1)	(0,2)	(0,3)
(1,0)	(1,1)	(1,2)	(1,3)

Question 1:

T at (1,0).
Observation = ***o6***. (the target is not in the 1 cell neighbourhood of the agent.)
That means the states that are allowed for **A** will be then (0,1), (0,2), (0,3), (1,2), (1,3).
And with both the values of **C.**
So total such states that have an equal chance of being the required state,$n = 2 * 5 = 10$
Thus the belief value of each of these states will be $1 / n = 1/10 = 0.1$

Thus the initial belief state, rest all (118) states will have value 0.

- S_01_10_0 -> 0.1
- S_01_10_1 -> 0.1
- S_02_10_0 -> 0.1
- S_02_10_1 -> 0.1
- S_03_10_0 -> 0.1
- S_03_10_1 -> 0.1
- S_12_10_0 -> 0.1
- S_12_10_1 -> 0.1
- S_13_10_0 -> 0.1
- S_13_10_1 -> 0.1

Question 2:

A at (1,1).
Observation = the target is in the 1 cell neighbourhood of the agent and not making a call.
That means the states that are allowed for **T** will be then (0,1), (1,0), (1,1), (1,2); (including the self state in the neighbourhood state.).
C = *off*.
So total such states that have an equal chance of being the required state, $n = 4$
Thus the belief value of each of these states will be $1 / n = 1/4 = 0.25$

Thus the initial belief state, rest all (124) states will have value 0.

- S_11_01_0 -> 0.25
- S_11_10_0 -> 0.25
- S_11_11_0 -> 0.25
- S_11_12_0 -> 0.25

Question 3:

We calculated the expected utility value with the help of Sarsop.
By running this Command -
` ./pomdpsim --simLen 100 --simNum 5000 --policy-file q1.policy q1.pomdp `
But we found the value to be always fluctuating between the range given as 95% Confidence Interval, and was never the same on running the same code consecutively.

For Q1.

Expected value: **26.2283** (it fluctuated between 25.8 ~ 26.5)

```
Simulating ...
  action selection :  one-step look ahead

-----------------------------------
 #Simulations   | Exp Total Reward
-----------------------------------
 500              25.4188
 1000             26.3189
 1500             26.5435
 2000             26.4345
 2500             26.3387
 3000             26.3835
 3500             26.2743
 4000             26.1139
 4500             26.1625
 5000             26.2283
-----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations   | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 5000             26.2283            (25.8008, 26.6557)
-------------------------------------------------------------
```

For Q2.

Expected value: **40.4532**
(it fluctuated between 40 ~ 41)

```
Simulating ...
  action selection :  one-step look ahead

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 500            39.9194
 1000           39.6657
 1500           39.6764
 2000           39.87
 2500           39.9783
 3000           40.0784
 3500           40.2792
 4000           40.3324
 4500           40.3849
 5000           40.4532
-----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 5000           40.4532            (40.0442, 40.8622)
-------------------------------------------------------------
```

Question 4:

When agent **A** is at (0,0) with probability *0.4* and target **T** is at (0,1), (0, 2), (1, 1) and (1, 2).
So amongst all the observations, it can be *o2* when **T** is at (0,1) otherwise it has to be *o6.*

o1	***o2***	***o3***	***o4***	***o5***	***o6***
0	0.25	0	0	0	0.75

When agent **A** is at (1,3) with probability *0.6* and target **T** is at (0,1), (0, 2), (1, 1) and (1, 2).
So amongst all the observations, it can be *o4* when **T** is at (1,2) otherwise it has to be *o6.*

o1	***o2***	***o3***	***o4***	***o5***	***o6***
0	0	0	0.25	0	0.75

Now taking the weighted mean for both cases with 0.4 and 0.6 probability.

$$Table3 \ = \ 0.4 \ * \ Table1 \ + \ 0.6 \ * \ Table3$$

Example : $Table3 \ (o2) \ = \ (0.4 \ * \ 0.25) \ + \ (0.6 \ * \ 0) \ = \ 0.1$

o1	***o2***	***o3***	***o4***	***o5***	***o6***
0	0.1	0	0.15	0	0.75

Thus the most probable observation will be ***o6*** with probability 0.75

Other way to calculate:
All states that are possible for the conditions given in the question: state -> probability ->observation

- S_00_01_0 -> 0.05 -> *o2*
- S_00_01_1 -> 0.05 -> *o2*
- S_00_02_0 -> 0.05 -> *o6*
- S_00_02_1 -> 0.05 -> *o6*
- S_00_11_0 -> 0.05 -> *o6*
- S_00_11_1 -> 0.05 -> *o6*
- S_00_12_0 -> 0.05 -> *o6*
- S_00_12_1 -> 0.05 -> *o6*
- S_13_01_0 -> 0.075 -> *o6*
- S_13_01_1 -> 0.075 -> *o6*
- S_13_02_0 -> 0.075 -> *o6*
- S_13_02_1 -> 0.075 -> *o6*
- S_13_11_0 -> 0.075 -> *o6*
- S_13_11_1 -> 0.075 -> *o6*
- S_13_12_0 -> 0.075 -> *o4*
- S_13_12_1 -> 0.075 ->*o4*

We can calculate the expectation for observation:
For o2 = 0.05 * 2 = 0.1
For o4 = 0.075 * 2 = 0.15
For o6 = (0.05 * 6) + (0.075*6) = 0.75

Thus the most probable observation will be ***o6*** with probability **0.75**

Question 5:

While running the **pomdpbsol** with **.pomdp** file to generate **.policy,** it also gives as output.
In this **#Trial** can be used as Time Horizon ***T*** for the POMPD.
For our case, we found it to be **T = 28.**

```
-------------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------------
 0.03    28      215      29.999     29.9999    0.00091268  80       49
-------------------------------------------------------------------------------------
```

Using this in the formula to find the number of nodes, and then finding the number of trees.

$|A| = 5$
$|O| = 6$
$T = 28$

For no. of nodes,
(the sum of geometric progression, with the Time horizon providing the limit till which to sum)

$$N = \frac{|O|^T - 1}{|O| - 1} = \frac{6^{28} - 1}{6 - 1} = 1.2281884428929632e + 21$$

No. of Policy trees that can be made with *n* nodes.

$$N_{PT} = |A|^N = 5^{1.228e+21}$$

which is a very huge number, beyond comprehension.