Assignment 3 Part 2

Snehal Ranjan (2020121003) VJS Pranavasri (2020121001)

Values Used

Rollnumber = 2020121001
x = 1 - (((LastFourDigitsOfRollNumber)%30 + 1) / 100)
Available Actions : [UP, DOWN, LEFT, RIGHT, STAY]
P(Left)=P(Right)=P(Up)=P(Down) = 0.1
P(Stay)=0.6
P(CallOn) = 0.5
P(CallOff) = 0.1

Transition Probabilities
Probability of success of stay = 1
Probability of success of any other action = x
Probability of failure = 1-x (Moves opposite)

Rewards
Reward for aany step = -1
Reward for reaching the target before the call is turned off = (RollNumber%90 + 10)

Grid

0,0	(0,1)	(0,2)	(0,3)
(1,0)	(1,1)	(1,2)	(1,3)

Answers

1.

Initial Belief state
Target is in (0, 1) and Target is not in 1 cell neighbourhood of agent (o6 is observed)
i.e. Possible cells for Agent: (0,1), (0,2), (0,3), (1,2), (1,3)
Possible states for call On(1), Off(0)
Therefore possible states are:

```
[
 (0,1), 0,
 (0,1), 1,
 (0,2), 0,
 (0,2), 1,
 (0,3), 0,
 (0,3), 1,
 (1,2), 0,
 (1,2), 1,
 (1,3), 0,
 (0,3), 1,
]
```

As there is equal probability of agent being in any of those states:
Initial Belief states would be all 0 except for the above states where it will be 1/10

2.

Agent in (1, 1), target is in one neighbourhood
Possibile positions for target: (1, 0), (0, 1), (1, 1), (1, 2)
For each, call will be off (given)

```
[
(1, 0), 0,
(0, 1), 0,
(1, 1), 0,
(1, 2), 0
]
```

For these four states probability will be 1/4 and remaining 0
That is our belief state.

3.

Expected reward for 1 after initial belief states

```
┌──(vjspranav㉿TUF-A15)-[~/Courses/sem4/MDL]
└─$ ./pomdpsim --simLen 100 --simNum 1000 --policy-file 2020121001_2020121003.policy q1.pomdp

Loading the model ...
  input file   : q1.pomdp

Loading the policy ...
  input file   : 2020121001_2020121003.policy

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

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 100            10.8456
 200            10.2012
 300            9.52019
 400            9.58399
 500            9.6753
 600            9.82451
 700            10.0164
 800            10.1327
 900            10.1279
 1000           10.2233
-----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 1000           10.2233            (9.50892, 10.9376)
-------------------------------------------------------------
```

Expected utility = 10.2233 after 1000 iterations

Expected reward for 2 after inital belief states

```
┌──(vjspranav㉿TUF-A15)-[~/Courses/sem4/MDL]
└─$ ./pomdpsim --simLen 100 --simNum 1000 --policy-file q2.policy q2.pomdp

Loading the model ...
  input file   : q2.pomdp

Loading the policy ...
  input file   : q2.policy

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

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 100             30.077
 200             31.3426
 300             31.4987
 400             30.4094
 500             30.3747
 600             30.9127
 700             31.9382
 800             31.4889
 900             31.3839
 1000            31.2598
-----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 1000            31.2598            (29.2013, 33.3183)
-------------------------------------------------------------
```

Expected utility = 31.2598 after 1000 iterations

4.

Given agent position:

Position	Probab
(0,1)	0.4
(1,3)	0.6

Given Target Position

Position	Probab
(0,1)	0.25
(0,2)	0.25
(1,1)	0.25
(1,2)	0.25

Call:

Position	Probab

0	0.5
1	0.5

All possible states and their probabilities:

State	Probability	Observation
((0, 0),(0, 1),0)	0.05	o2
((0, 0),(0, 1),1)	0.05	o2
((0, 0),(0, 2),0)	0.05	o6
((0, 0),(0, 2),1)	0.05	o6
((0, 0),(1, 1),0)	0.05	o6
((0, 0),(1, 1),1)	0.05	o6
((0, 0),(1, 2),0)	0.05	o6
((0, 0),(1, 2),1)	0.05	o6
((1, 3),(0, 1),0)	0.075	o6
((1, 3),(0, 1),1)	0.075	o6
((1, 3),(0, 2),0)	0.075	o6
((1, 3),(0, 2),1)	0.075	o6
((1, 3),(1, 1),0)	0.075	o6
((1, 3),(1, 1),1)	0.075	o6
((1, 3),(1, 2),0)	0.075	o4
((1, 3),(1, 2),1)	0.075	o4

Probability of each observation:

Observation	Probability
o2	2 * 0.05
o4	6 * 0.05 + 6 * 0.075
06	2 * 0.075

o6 has the highest cumulative probability hence we are the most likely to observe o6

5.

decision.

The size of a policy tree depends on the number of possible observations and the horizon. When the horizon is H, the number of nodes in a tree is

$$\sum_{t=0}^{H-1} |O|^t = \frac{|O|^H - 1}{|O| - 1} \qquad (6)$$

where $|O|$ is the size of O. At each node, the number of possible actions is $|A|$. Therefore, the total number of all possible H-horizon policy trees is

$$|A|^{\frac{|O|^H - 1}{|O|-1}}. \qquad (7)$$

Both numbers are exponential.

O is obesvations and A is Actions
|O| = 6
|A| = 5
|T| cannot be calculated without running the policy file, even after running as per the formula we can see that the value for number of policy trees, calculated will be very large as the power of A itself is going to be a very huge number, and the answer will be exponential.
Hence number of policy trees cannot be calculated rather is too big