MDL Assignment 3 Part B

Palash Sharma (2019101082)
Aryan Jain (2019101056)

Roll no used : 2019101082

0 (0, 0)	1 (0, 1)	2 (0, 2)	3 (0, 3)
4 (1, 0)	5 (1, 1)	6 (1, 2)	7 (1, 3)

1. Target is in cell no. 4. Due to the observation o6, agent cannot be in cell 0 or 5. Therefore we have the following 10 states possible:
   (1, 4, 0), (1, 4, 1), (2, 4, 0), (2, 4, 1), (6, 4, 0), (6, 4, 1), (3, 4, 0), (3, 4, 1), (7, 4, 0), (7, 4, 1)
   Each of these states have equal probability of occurring as the initial state, and hence each has a 0.10 value in the initial belief state table and all the other 118 states have 0 value.

2. We(agent) are in cell 5 and the target could be in one of 1, 4 or 6 with equal probability and the state call remains in state 0. Hence the 3 only possible states are:
   (5, 1, 0), (5, 4, 0), (5, 6, 0)
   Each of the 3 states shall have equal probabilities of 0.333 and the rest of the states will have 0 probability for the initial belief state.

3. The expected values are:
   Q1: 7.33356
   Q2. 13.999

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


----------------------------------
 #Simulations  | Exp Total Reward
----------------------------------
 100            8.2018
 200            8.05627
 300            7.54689
 400            7.53189
 500            7.38043
 600            7.30276
 700            7.22493
 800            7.30995
 900            7.23358
 1000           7.33356
----------------------------------

Finishing ...

------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
------------------------------------------------------------
 1000           7.33356           (6.90193, 7.7652)
------------------------------------------------------------
```

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


----------------------------------
 #Simulations  | Exp Total Reward
----------------------------------
 100            13.6564
 200            14.3065
 300            14.0618
 400            14.0038
 500            13.9463
 600            14.0376
 700            14.0147
 800            14.0466
 900            14.0733
 1000           13.999
----------------------------------

Finishing ...

------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
------------------------------------------------------------
 1000           13.999            (13.7032, 14.2947)
------------------------------------------------------------
```

4. The agent can be at cell 0 or 7 and the target can be in one of cells 1, 2, 5 or 6 with equal probability. So lets see at the corresponding belief state table

(0, 1, 0) with probability of 0.05	observation is o2
(0, 2, 0) with probability of 0.05	observation is o6
(0, 5, 0) with probability of 0.05	observation is o6
(0, 6, 0) with probability of 0.05	observation is o6
(0, 1, 1) with probability of 0.05	observation is o2
(0, 2, 1) with probability of 0.05	observation is o6
(0, 5, 1) with probability of 0.05	observation is o6
(0, 6, 1) with probability of 0.05	observation is o6

(7, 1, 0) with probability of 0.075 observation is o6
(7, 2, 0) with probability of 0.075 observation is o6
(7, 5, 0) with probability of 0.075 observation is o6
(7, 6, 0) with probability of 0.075 observation is o4
(7, 1, 1) with probability of 0.075 observation is o6
(7, 2, 1) with probability of 0.075 observation is o6
(7, 5, 1) with probability of 0.075 observation is o6
(7, 6, 1) with probability of 0.075 observation is o4

o2 has total probability of 0.10 of occurrence, o4 has probability of 0.15 whereas o6 has 0.75

Hence we will observe o6, i.e. agent and target will not be in each others' 1 neighbourhood with most likeliness.

5. If |A| are the total number of actions, |O| be the total number of observations, and T be the number of horizons (basically the number of trials), then total possible number of policy trees are given by the formula :

   |A|^N

   where N is the total number of nodes in a tree.

   N is given by the formula :
   N = summation(i=0 to T-1) |O|^i
   therefore,
   N = 1 + |O| + |O|^2 + |O|^3 + …+ |O|^(T-1)

   This forms a Geometric Progression, which can be compressed and can be written as:
   N = (|O|^T - 1) / (|O| - 1).

   Hence, the overall formula for calculating the number of policy trees becomes :

   Ans = |A| ^ ((|O|^T - 1) / (|O| - 1))

   For the question 4, we have |O| = 6, |A| = 5, T = 35

   Ans = 5 ^ ((6^35 -1) / (5)), which is approximately infinite! :D