MDL Assignment 3 Part 2

Ahana Datta (2019111007)
Tanvi Narsapur (2019111005)

The roll number used is 2019111005.
The value of x can be given by
x = ((1005%30)+1)/100 = 0.84
The reward value is (2019111005%90)+10 = 75

The coordinates of a position (x, y) are encoded as
(x, y): 2*x + y

Total number of states: 8*8*2 = 128
The state is represented as (agent position, target position, call)
The states are encoded as
(a, t, c) : a*16 + t*2 + c
a and t can have values {0,1,...,6,7}
c can have the value 0 indicating the call is off and 1 indicating the call is on.

The positions are mapped as:

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

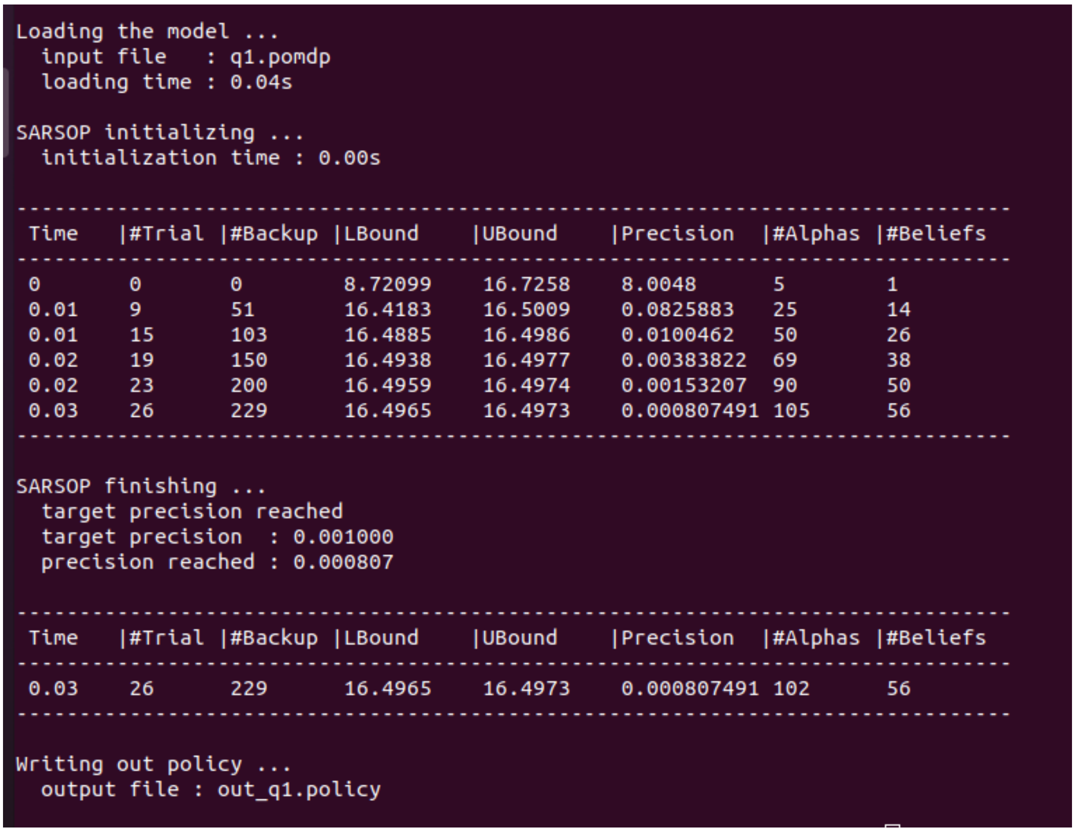
The mapping used for actions is as follows:
Stay: 0
Up: 1
Down: 2
Left : 3
Right: 4

The observations are mapped as follows:
O1: 0
O2: 1
O3: 2
O4: 3
O5: 4
O6: 5

Question 1:
Given that target is at (1,0) and observation o6 is observed.
According to the convention used for positions, the target is at (0,0), encoded as 0.
The agent can have the positions - 3,4,5,6,7 and the call can be either on or off.
Thus the possible states are
(3,0,0), (3,0,1), (4,0,0), (4,0,1), (5,0,0), (5,0,1), (6,0,0), (6,0,1), (7,0,0), (7,0,1)

For the initial belief state, the above states will have the same probability that is equal to 1/10. Rest all the states will have a probability value of 0 in the initial belief state.

For generating the policy file, the possible start states are mapped to a single integer and included in the pomdp file as
start include: 48 49 64 65 80 81 96 97 112 113



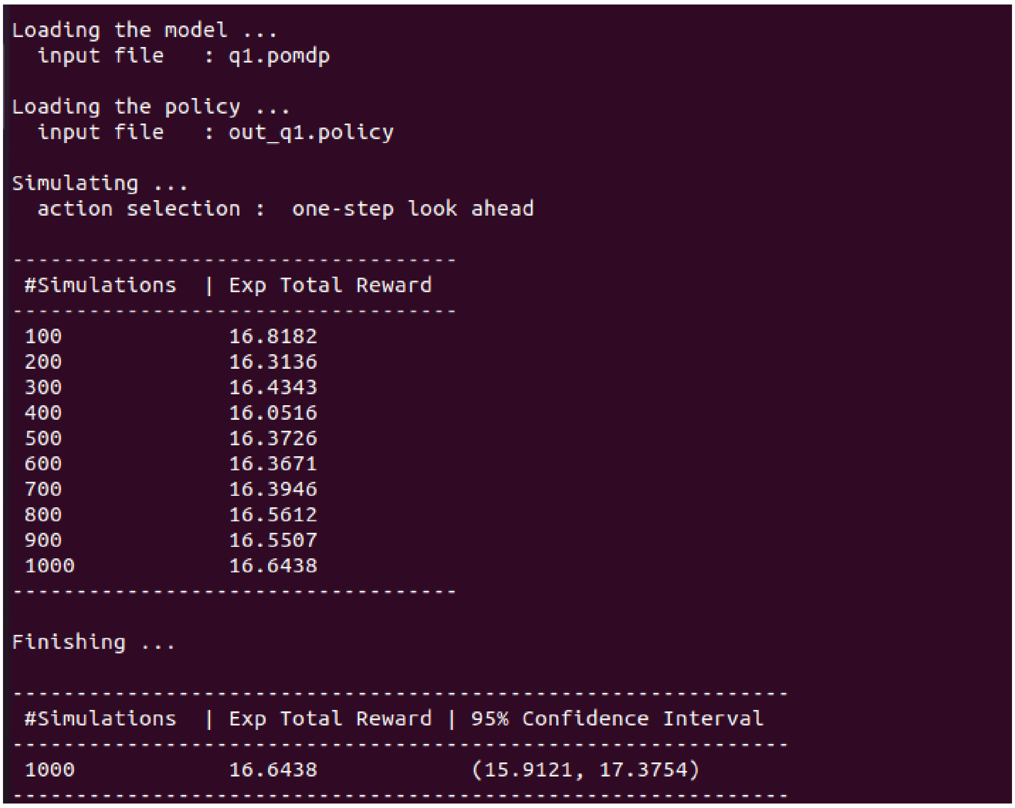
```
Loading the model ...
  input file   : q1.pomdp
  loading time : 0.04s

SARSOP initializing ...
  initialization time : 0.00s

-------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------
 0       0       0        8.72099    16.7258    8.0048      5        1
 0.01    9       51       16.4183    16.5009    0.0825883   25       14
 0.01    15      103      16.4885    16.4986    0.0100462   50       26
 0.02    19      150      16.4938    16.4977    0.00383822  69       38
 0.02    23      200      16.4959    16.4974    0.00153207  90       50
 0.03    26      229      16.4965    16.4973    0.000807491 105      56
-------------------------------------------------------------------------------

SARSOP finishing ...
  target precision reached
  target precision  : 0.001000
  precision reached : 0.000807

-------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------
 0.03    26      229      16.4965    16.4973    0.000807491 102      56
-------------------------------------------------------------------------------

Writing out policy ...
  output file : out_q1.policy
```

Question 2:
Given that the agent is located at (1,1) and the target is present in a one-cell neighbourhood, without making a call.
Since the target doesn't make a call, we use call=0.
Based on the mapping for positions, the agent is located at (1,0), encoded as 2.
Thus the target can be at 0,2,3,4 and the possible states can be given by
(2,0,0), (2,2,0), (2,3,0), (2,4,0)

In the initial belief state, the above states will have a probability value of ¼ and the rest all the states will have a probability value of 0.
This initial belief state is taken into account by mapping the above states to a single integer as follows
(2,0,0): 32
(2,2,0): 36
(2,3,0): 38
(2,4,0): 40

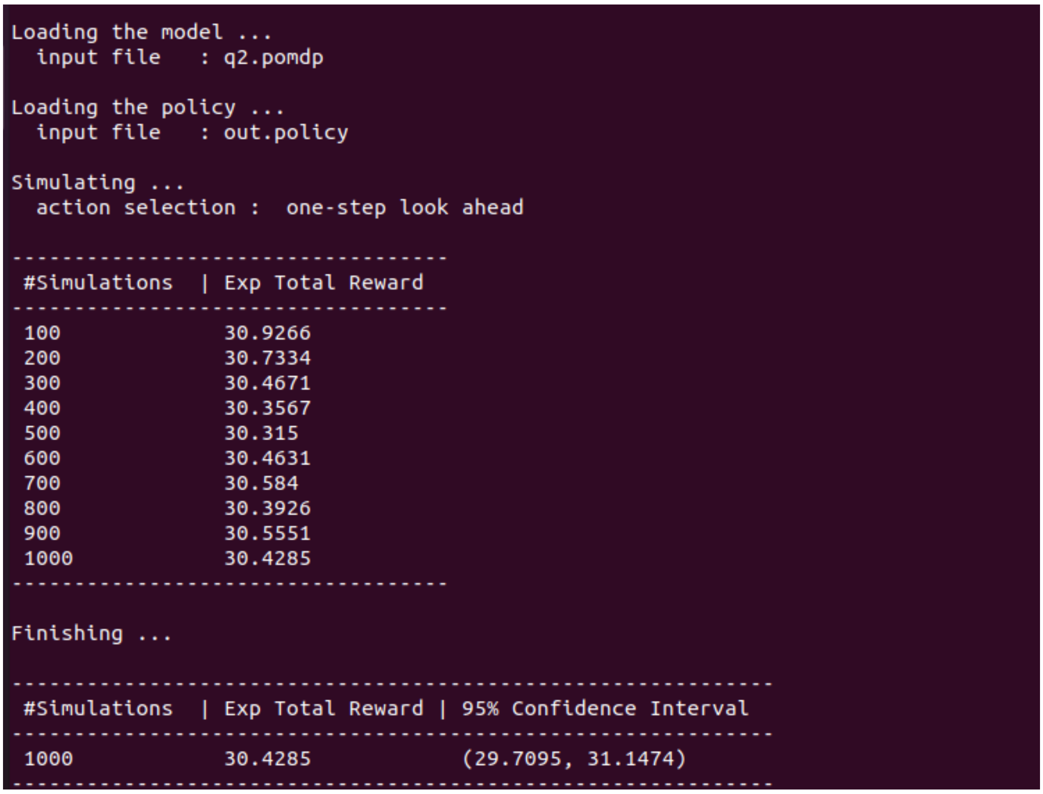
Question 3:

The command used to calculate expected utility for initial belief states is -
./pomdpsim pomdpFilename --policy-file policyFilename --simLen 100 --simNum 1000
Expected utility for initial belief state for q1 = 16.6438
Expected utility for initial belief state for q2 = 30.2845
q1 pomdpsim output:



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

Loading the policy ...
  input file   : out_q1.policy

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

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 100             16.8182
 200             16.3136
 300             16.4343
 400             16.0516
 500             16.3726
 600             16.3671
 700             16.3946
 800             16.5612
 900             16.5507
 1000            16.6438
-----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 1000            16.6438            (15.9121, 17.3754)
-------------------------------------------------------------
```

q2 pomdpsim output:



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

Loading the policy ...
  input file   : out.policy

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

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 100            30.9266
 200            30.7334
 300            30.4671
 400            30.3567
 500            30.315
 600            30.4631
 700            30.584
 800            30.3926
 900            30.5551
 1000           30.4285
-----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 1000           30.4285            (29.7095, 31.1474)
-------------------------------------------------------------
```

Question 4:
The agent can be located at (0,0) with probability 0.4 and (1,3) with a probability of 0.6 According to the position mapping, the agent can be at (0,1), encoded as 1 and (3,0) encoded as 6. The target can be at (0,1), (0,2), (1,1) and (1,2). According to the conventions used for mapping positions, the target positions are (1,1) encoded as 3, (2,1) encoded as 5, (1,0) encoded as 2, (2,0) encoded as 4.

(a, t, c)	Probability	Observation
(1, 2, 0)	0.05	O6
(1, 2, 1)	0.05	O6
(1, 3, 0)	0.05	O2
(1, 3, 1)	0.05	O2
(1, 4, 0)	0.05	O6
(1, 4, 1)	0.05	O6
(1, 5, 0)	0.05	O6
(1, 5, 1)	0.05	O6
(6, 2, 0)	0.075	O6
(6, 2, 1)	0.075	O6
(6, 3, 0)	0.075	O6
(6, 3, 1)	0.075	O6
(6, 4, 0)	0.075	O4
(6, 4, 1)	0.075	O4
(6, 5, 0)	0.075	O6
(6, 5, 1)	0.075	O6

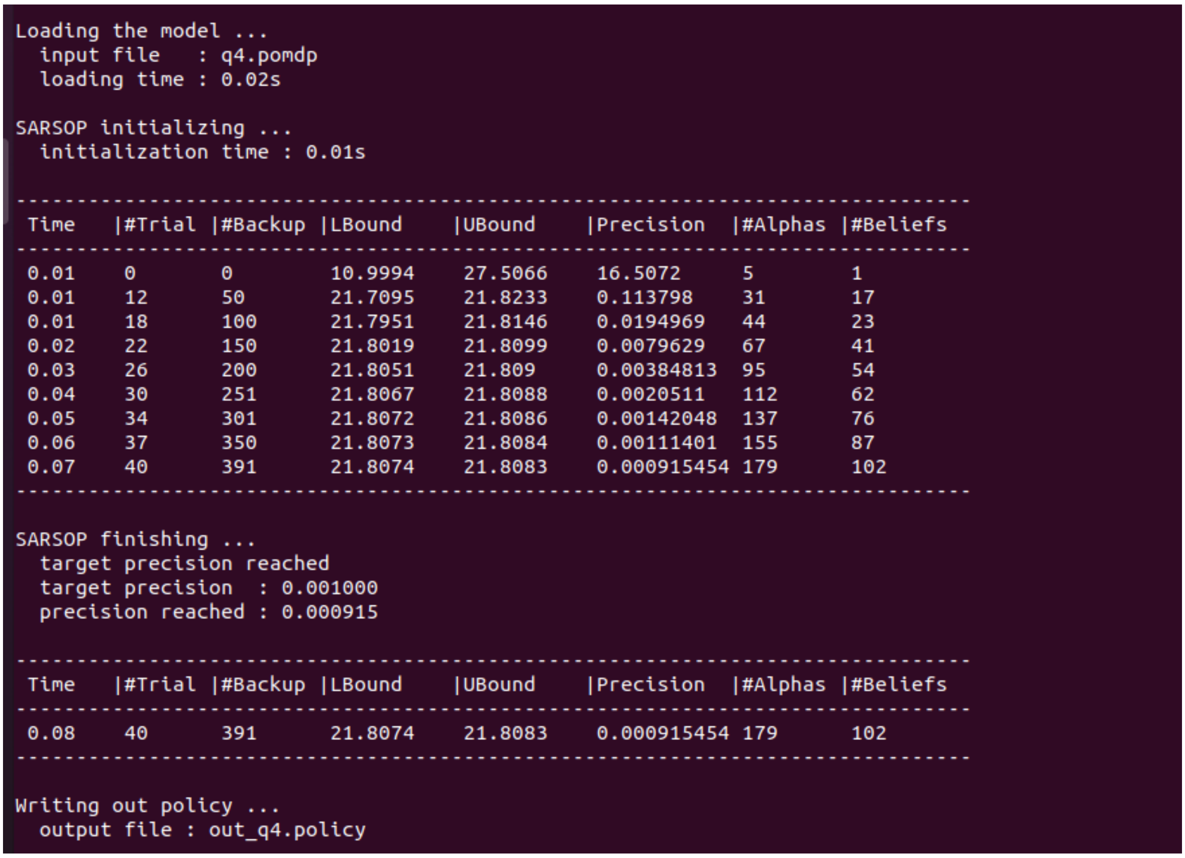
The probability of observing -

1) O6 is 6*0.05 + 6*0.075 = 0.75
2) O2 is 2*0.05 = 0.10
3) O4 is 2*0.075 = 0.15

Thus we are most likely to observe o6.

Question 5:

On running pomdpsol for Question 4:



```
Loading the model ...
  input file   : q4.pomdp
  loading time : 0.02s

SARSOP initializing ...
  initialization time : 0.01s

-------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------
 0.01    0       0        10.9994    27.5066    16.5072     5        1
 0.01    12      50       21.7095    21.8233    0.113798    31       17
 0.01    18      100      21.7951    21.8146    0.0194969   44       23
 0.02    22      150      21.8019    21.8099    0.0079629   67       41
 0.03    26      200      21.8051    21.809     0.00384813  95       54
 0.04    30      251      21.8067    21.8088    0.0020511   112      62
 0.05    34      301      21.8072    21.8086    0.00142048  137      76
 0.06    37      350      21.8073    21.8084    0.00111401  155      87
 0.07    40      391      21.8074    21.8083    0.000915454 179      102
-------------------------------------------------------------------------------

SARSOP finishing ...
  target precision reached
  target precision  : 0.001000
  precision reached : 0.000915

-------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------
 0.08    40      391      21.8074    21.8083    0.000915454 179      102
-------------------------------------------------------------------------------

Writing out policy ...
  output file : out_q4.policy
```

We will use the #Trial as T value for calculation

How many trees:

$$N = \sum_{i=0}^{T-1} |O|^{i} = (|O|^{T} - 1) / (|O| - 1)$$

$$|A|^{N}$$

A denotes the number of actions, O is the number of observations. T is taken as the #Trial value. For the given pomdp -
|A|=5, |O|=6, T=40.

Calculating the value of N -

N = (6^40 -1)/(6-1)
    = 2.6734989e*10^30

Thus the number of trees can be given by -
|A|^N = 5^(2.6734989*10^30)

This is a very large number.