MDL ASSIGNMENT 3 PART 2

Arushi Mittal (2019101120), Meghna Mishra (2019111030)

Variables Used:

```
Roll Number : 2019101120

Reward : 90

x : 0.79
```

Q1

The initial belief state would include all the positions except `(1, 0)`, `(0, 0)` and `(1, 1)` with call turned on and call turned off.

```
State Set: s01 s01c s02 s02c s03 s03c s12 s12c s13 s13c
```

This was generated by adding the states, transition matrix and reward matrix to the POMDP file and running it to generate a policy.

Time	#Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs
0.01	0	0	14.1782	27.4251	13.247	5	1
0.01	9	51	26.962	27.0847	0.122705	21	11
0.01	15	100	27.0472	27.0804	0.0331575	29	20
0.01	20	150	27.0682	27.0795	0.0113555	53	30
0.02	24	201	27.0754	27.0788	0.00341124	68	41
0.02	28	250	27.0772	27.0785	0.00129391	94	51
0.03	31	293	27.0773	27.0783	0.000994653	120	60

Time	Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs

0.03	31	293	27.0773	27.0783	0.000994653	120	60

Q2

The initial state would include all the states where the call is not active, and the states that do not have positions within the following states: (1, 0), (1, 1), (1, 2), (0, 1)

```
State Set: s13, s00, s02, s03
```

This was generated by adding the states, transition matrix and reward matrix to the POMDP file and running it to generate a policy.

Time	#Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs
0.01	0	0	22.5333	63.3519	40.8186	5	1
0.01	12	51	44.9633	45.0758	0.112574	42	14
0.01	20	103	45.0578	45.0696	0.0118304	86	27
0.02	24	150	45.0642	45.0692	0.00493364	93	37
0.02	28	200	45.0659	45.0686	0.00272805	133	47
0.03	32	255	45.0666	45.0683	0.00169059	162	63
0.03	36	299	45.067	45.068	0.000924471	184	70

Time	#Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs
0.03	36	299	45.067	45.068	0.000924471	184	70

Q3

After generating the policy files, we run the POMDP simulator in order to generate the expected utilities for the given states.

This is done by giving a simLen of 100 and simNum of 1000 where simLen is the number of steps taken and simNum is the number of times the simulation is run.

Q1

#Simulations	Exp Total Reward
100	27.5817
200	27.4375
300	27.4995
400	27.5089
500	27.8025
600	27.6536
700	27.342
800	27.6365
900	27.5673

#Simulations	Exp Total Reward	95% Confidence Interval
1000	27.4623	(26.4069, 28.5176)

Q2

#Simulations	Exp Total Reward
100	44.4628
200	43.916
300	44.1887
400	44.1797
500	44.2635
600	44.3647
700	44.3327
800	44.528
900	44.7364
1000	44.9339

#Simulations	Exp Total Reward	95% Confidence Interval
1000	44.9339	(44.0072, 45.8606)

Q4

According to the POMDP file, and the corresponding policy file, we can see the most likely observation would be `o4`, where the target is to the left of the agent. We know this because the only possible observations are `o2` and `o4` since none of the other observations would make sense in this context. However, since the probability of being in position `(1, 3)` is higher and in this case the target is to the left, this is the most likely observation. In the case of `(0, 0)` the probability is lower and the target is to the right, therefore the likelihood of `o2` is lower.

Q5

Running the POMDP solver yields the following results:

Time	#Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs
0.01	0	0	14.2777	42.3953	28.1176	5	1
0.01	11	51	32.8868	33.0694	0.182576	27	15
0.01	18	100	33.0248	33.0563	0.0315073	39	21
0.02	24	150	33.0419	33.0516	0.00966891	50	29
0.02	28	200	33.0462	33.0501	0.00391652	61	41
0.02	32	250	33.0474	33.0499	0.00246231	76	50
0.03	36	301	33.0478	33.0496	0.00188166	91	58
0.04	40	350	33.048	33.0495	0.00145474	116	76
0.04	44	405	33.0482	33.0493	0.00110486	135	90
0.05	47	447	33.0483	33.0492	0.000954413	162	98

Time	#Trial	#Backup	LBound	UBound	Precision	#Alphas	#Beliefs
0.05	47	447	33.0483	33.0492	0.000954413	153	98

The number of `actions` is `5` , the number of `observations` is `6` , `horizon` is 47

Therefore, the number of policy trees is $A^{\frac{O^H-1}{O-1}}$

This is equal to $5^{\frac{6^{47}-1}{6-1}}$ policy trees.