```
--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 0]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

--------------------------------------
Value of Behaviour policy:55.239
O_threshold = 100
MC for this TARGET:[60.446, 0.085]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.2, 0.12, -0.78]][[-0.31, -60.45, -5.21]]
std:[[0.56, 0.57, 0.37]][[0.34, 0.0, 0.23]]
MSE:[[0.59, 0.58, 0.86]][[0.46, 60.45, 5.22]]
MSE(-DR):[[0.0, -0.01, 0.27]][[-0.13, 59.86, 4.63]]
==============
O_threshold = 105
MC for this TARGET:[61.202, 0.07]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.28, -2.33, -3.31]][[-3.73, -61.2, -5.96]]
std:[[0.57, 0.57, 0.37]][[0.35, 0.0, 0.23]]
MSE:[[2.35, 2.4, 3.33]][[3.75, 61.2, 5.96]]
MSE(-DR):[[0.0, 0.05, 0.98]][[1.4, 58.85, 3.61]]
***
==============
O_threshold = 110
MC for this TARGET:[60.33, 0.063]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.26, -2.3, -2.98]][[-4.89, -60.33, -5.09]]
std:[[0.52, 0.52, 0.44]][[0.36, 0.0, 0.23]]
MSE:[[2.32, 2.36, 3.01]][[4.9, 60.33, 5.1]]
MSE(-DR):[[0.0, 0.04, 0.69]][[2.58, 58.01, 2.78]]
***
==============
O_threshold = 115
MC for this TARGET:[63.126, 0.061]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.68, -7.67, -7.69]][[-13.14, -63.13, -7.89]]
std:[[0.7, 0.7, 0.49]][[0.37, 0.0, 0.23]]
MSE:[[7.71, 7.7, 7.71]][[13.15, 63.13, 7.89]]
MSE(-DR):[[0.0, -0.01, 0.0]][[5.44, 55.42, 0.18]]
***
==============
[[ 0.59  0.58  0.86  0.46 60.45  5.22]
 [ 2.35  2.4   3.33  3.75 61.2   5.96]
 [ 2.32  2.36  3.01  4.9  60.33  5.1 ]
 [ 7.71  7.7   7.71 13.15 63.13  7.89]]


time spent until now: 39.6 mins

20:56, 04/11

--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 10]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
```

```
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:55.225
O_threshold = 100
MC for this TARGET:[60.461, 0.14]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.22, 0.14, -0.85]][[-0.35, -60.46, -5.24]]
std:[[0.57, 0.58, 0.48]][[0.31, 0.0, 0.19]]
MSE:[[0.61, 0.6, 0.98]][[0.47, 60.46, 5.24]]
MSE(-DR):[[0.0, -0.01, 0.37]][[-0.14, 59.85, 4.63]]
==============
O_threshold = 105
MC for this TARGET:[61.218, 0.131]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.11, -2.17, -3.34]][[-3.8, -61.22, -5.99]]
std:[[0.66, 0.66, 0.42]][[0.34, 0.0, 0.19]]
MSE:[[2.21, 2.27, 3.37]][[3.82, 61.22, 5.99]]
MSE(-DR):[[0.0, 0.06, 1.16]][[1.61, 59.01, 3.78]]
***
==============
O_threshold = 110
MC for this TARGET:[60.346, 0.129]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.19, -2.24, -3.02]][[-4.93, -60.35, -5.12]]
std:[[0.54, 0.53, 0.49]][[0.37, 0.0, 0.19]]
MSE:[[2.26, 2.3, 3.06]][[4.94, 60.35, 5.12]]
MSE(-DR):[[0.0, 0.04, 0.8]][[2.68, 58.09, 2.86]]
***
==============
O_threshold = 115
MC for this TARGET:[63.142, 0.133]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.67, -7.66, -7.72]][[-13.17, -63.14, -7.92]]
std:[[0.77, 0.75, 0.56]][[0.37, 0.0, 0.19]]
MSE:[[7.71, 7.7, 7.74]][[13.18, 63.14, 7.92]]
MSE(-DR):[[0.0, -0.01, 0.03]][[5.47, 55.43, 0.21]]
***
==============
[[ 0.59  0.58  0.86  0.46 60.45  5.22]
 [ 2.35  2.4   3.33  3.75 61.2   5.96]
 [ 2.32  2.36  3.01  4.9  60.33  5.1 ]
 [ 7.71  7.7   7.71 13.15 63.13  7.89]]


[[ 0.61  0.6   0.98  0.47 60.46  5.24]
 [ 2.21  2.27  3.37  3.82 61.22  5.99]
 [ 2.26  2.3   3.06  4.94 60.35  5.12]
 [ 7.71  7.7   7.74 13.18 63.14  7.92]]


time spent until now: 79.9 mins

21:37, 04/11

---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 20]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:55.211
O_threshold = 100
MC for this TARGET:[60.477, 0.242]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[0.25, 0.17, -0.92]][[-0.41, -60.48, -5.27]]
std:[[0.92, 0.93, 0.67]][[0.4, 0.0, 0.22]]
MSE:[[0.95, 0.95, 1.14]][[0.57, 60.48, 5.27]]
MSE(-DR):[[0.0, 0.0, 0.19]][[-0.38, 59.53, 4.32]]
==============
O_threshold = 105
MC for this TARGET:[61.234, 0.237]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.95, -2.02, -3.49]][[-3.87, -61.23, -6.02]]
std:[[1.15, 1.16, 0.62]][[0.45, 0.0, 0.22]]
MSE:[[2.26, 2.33, 3.54]][[3.9, 61.23, 6.02]]
MSE(-DR):[[0.0, 0.07, 1.28]][[1.64, 58.97, 3.76]]
***
==============
O_threshold = 110
MC for this TARGET:[60.362, 0.237]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.06, -2.16, -3.03]][[-4.97, -60.36, -5.15]]
std:[[0.91, 0.89, 0.64]][[0.49, 0.0, 0.22]]
MSE:[[2.25, 2.34, 3.1]][[4.99, 60.36, 5.15]]
MSE(-DR):[[0.0, 0.09, 0.85]][[2.74, 58.11, 2.9]]
***
==============
O_threshold = 115
MC for this TARGET:[63.158, 0.241]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.66, -7.69, -7.71]][[-13.2, -63.16, -7.95]]
std:[[1.18, 1.13, 0.89]][[0.49, 0.0, 0.22]]
MSE:[[7.75, 7.77, 7.76]][[13.21, 63.16, 7.95]]
MSE(-DR):[[0.0, 0.02, 0.01]][[5.46, 55.41, 0.2]]
***
==============
[[ 0.59  0.58  0.86  0.46 60.45  5.22]
 [ 2.35  2.4   3.33  3.75 61.2   5.96]
 [ 2.32  2.36  3.01  4.9  60.33  5.1 ]
 [ 7.71  7.7   7.71 13.15 63.13  7.89]]


[[ 0.61  0.6   0.98  0.47 60.46  5.24]
 [ 2.21  2.27  3.37  3.82 61.22  5.99]
 [ 2.26  2.3   3.06  4.94 60.35  5.12]
 [ 7.71  7.7   7.74 13.18 63.14  7.92]]


[[ 0.95  0.95  1.14  0.57 60.48  5.27]
 [ 2.26  2.33  3.54  3.9  61.23  6.02]
 [ 2.25  2.34  3.1   4.99 60.36  5.15]
 [ 7.75  7.77  7.76 13.21 63.16  7.95]]


time spent until now: 120.0 mins

22:17, 04/11
ubuntu@ip-172-31-4-195:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
22:25, 04/11; num of cores:16
simple_old

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, sd_R_range, t_func] = [16, None, None, 20, 0.5, 2, 0,
[0, 10, 20], None]


---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 0]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
 target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!
```

```
--------------------------------------
Value of Behaviour policy:47.183
O_threshold = 100
MC for this TARGET:[49.616, 0.082]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.89, 1.8, 0.91]][[1.92, -49.62, -2.43]]
std:[[0.56, 0.57, 0.38]][[0.35, 0.0, 0.27]]
MSE:[[1.97, 1.89, 0.99]][[1.95, 49.62, 2.44]]
MSE(-DR):[[0.0, -0.08, -0.98]][[-0.02, 47.65, 0.47]]
==============
O_threshold = 105
MC for this TARGET:[48.966, 0.065]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.36, 0.3, -0.69]][[-0.81, -48.97, -1.78]]
std:[[0.63, 0.63, 0.38]][[0.34, 0.0, 0.27]]
MSE:[[0.73, 0.7, 0.79]][[0.88, 48.97, 1.8]]
MSE(-DR):[[0.0, -0.03, 0.06]][[0.15, 48.24, 1.07]]
***
==============
O_threshold = 110
MC for this TARGET:[47.587, 0.061]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.85, 0.82, 0.03]][[-1.73, -47.59, -0.4]]
std:[[0.6, 0.59, 0.45]][[0.36, 0.0, 0.27]]
MSE:[[1.04, 1.01, 0.45]][[1.77, 47.59, 0.48]]
MSE(-DR):[[0.0, -0.03, -0.59]][[0.73, 46.55, -0.56]]
**
==============
O_threshold = 115
MC for this TARGET:[51.09, 0.044]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.92, -5.9, -6.07]][[-11.73, -51.09, -3.91]]
std:[[0.7, 0.72, 0.51]][[0.36, 0.0, 0.27]]
MSE:[[5.96, 5.94, 6.09]][[11.74, 51.09, 3.92]]
MSE(-DR):[[0.0, -0.02, 0.13]][[5.78, 45.13, -2.04]]
***
==============
[[ 1.97  1.89  0.99  1.95 49.62  2.44]
 [ 0.73  0.7   0.79  0.88 48.97  1.8 ]
 [ 1.04  1.01  0.45  1.77 47.59  0.48]
 [ 5.96  5.94  6.09 11.74 51.09  3.92]]


time spent until now: 40.1 mins

23:05, 04/11

--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 10]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

--------------------------------------
Value of Behaviour policy:47.169
O_threshold = 100
MC for this TARGET:[49.632, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.92, 1.84, 0.83]][[1.89, -49.63, -2.46]]
std:[[0.63, 0.66, 0.49]][[0.33, 0.0, 0.25]]
MSE:[[2.02, 1.95, 0.96]][[1.92, 49.63, 2.47]]
MSE(-DR):[[0.0, -0.07, -1.06]][[-0.1, 47.61, 0.45]]
==============
O_threshold = 105
MC for this TARGET:[48.982, 0.124]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.49, 0.43, -0.75]][[-0.87, -48.98, -1.81]]
```

```
std:[[0.7, 0.7, 0.47]][[0.37, 0.0, 0.25]]
MSE:[[0.85, 0.82, 0.89]][[0.95, 48.98, 1.83]]
MSE(-DR):[[0.0, -0.03, 0.04]][[0.1, 48.13, 0.98]]
***
==============
O_threshold = 110
MC for this TARGET:[47.603, 0.123]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.9, 0.85, -0.02]][[-1.76, -47.6, -0.43]]
std:[[0.57, 0.55, 0.53]][[0.39, 0.0, 0.25]]
MSE:[[1.07, 1.01, 0.53]][[1.8, 47.6, 0.5]]
MSE(-DR):[[0.0, -0.06, -0.54]][[0.73, 46.53, -0.57]]
**
==============
O_threshold = 115
MC for this TARGET:[51.106, 0.125]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.94, -5.93, -6.11]][[-11.75, -51.11, -3.94]]
std:[[0.92, 0.92, 0.57]][[0.4, 0.0, 0.25]]
MSE:[[6.01, 6.0, 6.14]][[11.76, 51.11, 3.95]]
MSE(-DR):[[0.0, -0.01, 0.13]][[5.75, 45.1, -2.06]]
***
==============
[[ 1.97  1.89  0.99  1.95 49.62  2.44]
 [ 0.73  0.7   0.79  0.88 48.97  1.8 ]
 [ 1.04  1.01  0.45  1.77 47.59  0.48]
 [ 5.96  5.94  6.09 11.74 51.09  3.92]]


[[ 2.02  1.95  0.96  1.92 49.63  2.47]
 [ 0.85  0.82  0.89  0.95 48.98  1.83]
 [ 1.07  1.01  0.53  1.8  47.6   0.5 ]
 [ 6.01  6.    6.14 11.76 51.11  3.95]]


time spent until now: 80.7 mins

23:46, 04/11

--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 20]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
```