ubuntu@ip-172-31-0-7:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
10:01, 04/16; num of cores:96
sd_u_O_20_full_T

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D_range, t_func] = [16, None, None, 20, 0.5, 1.5, [20, 10], None]

[thre_range, sd_R_range, day_range, penalty_range]: [[100, 100.5, 105, 110, 110.5, 111], [15], [2, 4, 6, 8], [[0.0003, 0.0001, 5e-05], [0.0003, 0.0001, 5e-05]]]

[pattern_seed, day, sd_R, u_O_u_D] = [2, 2, 15, 20]

max(u_O) = 145.8
O_threshold = 100
means of Order:

91.7 98.9 57.3 132.8 64.1

83.2 110.1 75.1 78.8 81.8

111.0 145.8 100.8 77.6 110.8

88.1 99.6 123.5 85.0 100.2

82.4 96.9 105.1 80.2 93.2

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 1

0 0 1 0 0

number of reward locations: 9
O_threshold = 100.5
number of reward locations: 8
O_threshold = 105
number of reward locations: 7
O_threshold = 110
number of reward locations: 6
O_threshold = 110.5
number of reward locations: 5
O_threshold = 111
number of reward locations: 4
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

```
––––––––––––––––––––––––––––––––––––––––
Value of Behaviour policy:49.444
O_threshold = 100
MC for this TARGET:[54.85, 0.309]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.58, -2.73, -2.42]][[-3.37, -54.85, -5.41]]
std:[[1.18, 1.21, 1.22]][[0.77, 0.0, 0.46]]
MSE:[[2.84, 2.99, 2.71]][[3.46, 54.85, 5.43]]
MSE(-DR):[[0.0, 0.15, -0.13]][[0.62, 52.01, 2.59]]
**
==============
O_threshold = 100.5
MC for this TARGET:[52.122, 0.304]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.87, -1.01, -1.18]][[-2.51, -52.12, -2.68]]
std:[[1.31, 1.33, 1.07]][[0.72, 0.0, 0.46]]
MSE:[[1.57, 1.67, 1.59]][[2.61, 52.12, 2.72]]
MSE(-DR):[[0.0, 0.1, 0.02]][[1.04, 50.55, 1.15]]
***
==============
O_threshold = 105
MC for this TARGET:[53.588, 0.311]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.27, -4.36, -4.12]][[-5.73, -53.59, -4.14]]
std:[[1.91, 1.91, 1.17]][[0.71, 0.0, 0.46]]
MSE:[[4.68, 4.76, 4.28]][[5.77, 53.59, 4.17]]
MSE(-DR):[[0.0, 0.08, -0.4]][[1.09, 48.91, -0.51]]
**
==============
O_threshold = 110
MC for this TARGET:[52.207, 0.307]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.98, -3.06, -3.85]][[-6.45, -52.21, -2.76]]
std:[[1.83, 1.83, 1.23]][[0.76, 0.0, 0.46]]
MSE:[[3.5, 3.57, 4.04]][[6.49, 52.21, 2.8]]
MSE(-DR):[[0.0, 0.07, 0.54]][[2.99, 48.71, -0.7]]
***
==============
O_threshold = 110.5
MC for this TARGET:[54.59, 0.302]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.3, -6.39, -6.38]][[-10.79, -54.59, -5.15]]
std:[[1.88, 1.8, 1.51]][[0.82, 0.0, 0.46]]
MSE:[[6.57, 6.64, 6.56]][[10.82, 54.59, 5.17]]
MSE(-DR):[[0.0, 0.07, -0.01]][[4.25, 48.02, -1.4]]
**
==============
O_threshold = 111
MC for this TARGET:[52.543, 0.302]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.25, -5.27, -6.06]][[-10.76, -52.54, -3.1]]
std:[[2.15, 2.08, 1.36]][[0.76, 0.0, 0.46]]
MSE:[[5.67, 5.67, 6.21]][[10.79, 52.54, 3.13]]
MSE(-DR):[[0.0, 0.0, 0.54]][[5.12, 46.87, -2.54]]
***
==============
[[ 2.84  2.99  2.71  3.46 54.85  5.43]
 [ 1.57  1.67  1.59  2.61 52.12  2.72]
 [ 4.68  4.76  4.28  5.77 53.59  4.17]
 [ 3.5   3.57  4.04  6.49 52.21  2.8 ]
 [ 6.57  6.64  6.56 10.82 54.59  5.17]
 [ 5.67  5.67  6.21 10.79 52.54  3.13]]


time spent until now: 15.8 mins

10:17, 04/16

––––––––––––––––––––––––––––––––––––––––
[pattern_seed, day, sd_R, u_O_u_D] = [2, 4, 15, 20]

max(u_O) =  145.8
O_threshold = 100
means of Order:

91.7 98.9 57.3 132.8 64.1

83.2 110.1 75.1 78.8 81.8

111.0 145.8 100.8 77.6 110.8

88.1 99.6 123.5 85.0 100.2

82.4 96.9 105.1 80.2 93.2

target policy:

0 0 0 1 0
```

```
0 1 0 0 0

1 1 1 0 1

0 0 1 0 1

0 0 1 0 0
```

```
number of reward locations:  9
O_threshold = 100.5
number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 110.5
number of reward locations:  5
O_threshold = 111
number of reward locations:  4
```