```
Last login: Thu Apr  2 11:24:49 on ttys001
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-35-168-113-18.compute-1.amazonaws.com
ssh: connect to host ec2-35-168-113-18.compute-1.amazonaws.com port 22: Connection refused
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-35-168-113-18.compute-1.amazonaws.com
Warning: Permanently added the ED25519 host key for IP address '35.168.113.18' to the list of known hosts.
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1063-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Thu Apr  2 15:50:57 UTC 2020

  System load:  1.02               Processes:            810
  Usage of /:   56.9% of 15.45GB   Users logged in:      0
  Memory usage: 0%                 IP address for ens5: 172.31.78.245
  Swap usage:   0%

 * Kubernetes 1.18 GA is now available! See https://microk8s.io for docs or
   install it with:

     sudo snap install microk8s --channel=1.18 --classic

 * Multipass 1.1 adds proxy support for developers behind enterprise
   firewalls. Rapid prototyping for cloud operations just got easier.

     https://multipass.run/

 * Canonical Livepatch is available for installation.
   - Reduce system reboots and improve kernel security. Activate at:
     https://ubuntu.com/livepatch

53 packages can be updated.
0 updates are security updates.


Last login: Wed Apr  1 20:30:39 2020 from 107.13.161.147
ubuntu@ip-172-31-78-245:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
11:52, 04/02; num of cores:96

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, simple, u_O_u_D]] = [None, 96, 10, 10,
None, 0.3, 0.5, 1, [True, False, True, False, 10]]


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

max(u_O) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
```

O_threshold = 100
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations: 8
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations: 6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!

```
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!


----------------------------------------
Value of Behaviour policy:60.786
O_threshold = 80
MC for this TARGET:[70.884, 0.141]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.88, 0.13, -70.88]][[-70.88, -70.88, -10.1]]
std:[[0.0, 0.63, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.88, 0.64, 70.88]][[70.88, 70.88, 10.1]]
MSE(-DR):[[0.0, -70.24, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.371, 0.133]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -0.28, -69.37]][[-69.37, -69.37, -8.59]]
std:[[0.0, 0.65, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[69.37, 0.71, 69.37]][[69.37, 69.37, 8.59]]
MSE(-DR):[[0.0, -68.66, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 100
MC for this TARGET:[68.94, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -2.88, -68.94]][[-68.94, -68.94, -8.15]]
std:[[0.0, 0.66, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[68.94, 2.95, 68.94]][[68.94, 68.94, 8.15]]
MSE(-DR):[[0.0, -65.99, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 110
MC for this TARGET:[70.484, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.48, -6.53, -70.48]][[-70.48, -70.48, -9.7]]
std:[[0.0, 0.66, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.48, 6.56, 70.48]][[70.48, 70.48, 9.7]]
MSE(-DR):[[0.0, -63.92, 0.0]][[0.0, 0.0, -60.78]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.88  0.64 70.88 70.88 70.88 10.1 ]
 [69.37  0.71 69.37 69.37 69.37  8.59]
```

```
 [68.94  2.95 68.94 68.94 68.94  8.15]
 [70.48  6.56 70.48 70.48 70.48  9.7 ]]
```

time spent until now: 0.6 mins

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

max(u_O) = 197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations: 15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations: 12
O_threshold = 100
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations: 8
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations: 6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!

```
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
```

25 -th region DONE!

Value of Behaviour policy:60.792
O_threshold = 80
MC for this TARGET:[70.887, 0.092]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, 0.37, -70.89]][[-70.89, -70.89, -10.09]]
std:[[0.0, 0.56, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.89, 0.67, 70.89]][[70.89, 70.89, 10.09]]
MSE(-DR):[[0.0, -70.22, 0.0]][[0.0, 0.0, -60.8]]
***
==============

O_threshold = 90
MC for this TARGET:[69.373, 0.094]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -0.15, -69.37]][[-69.37, -69.37, -8.58]]
std:[[0.0, 0.52, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[69.37, 0.54, 69.37]][[69.37, 69.37, 8.58]]
MSE(-DR):[[0.0, -68.83, 0.0]][[0.0, 0.0, -60.79]]
***
==============

O_threshold = 100
MC for this TARGET:[68.936, 0.097]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -2.96, -68.94]][[-68.94, -68.94, -8.14]]
std:[[0.0, 0.57, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[68.94, 3.01, 68.94]][[68.94, 68.94, 8.14]]
MSE(-DR):[[0.0, -65.93, 0.0]][[0.0, 0.0, -60.8]]
***
==============

O_threshold = 110
MC for this TARGET:[70.474, 0.102]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -6.61, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.64, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.47, 6.64, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -63.83, 0.0]][[0.0, 0.0, -60.79]]
***
==============

******************* THIS SETTING IS GOOD *******************
[[70.88 0.64 70.88 70.88 70.88 10.1]
 [69.37 0.71 69.37 69.37 69.37 8.59]
 [68.94 2.95 68.94 68.94 68.94 8.15]
 [70.48 6.56 70.48 70.48 70.48 9.7]]

[[70.89 0.67 70.89 70.89 70.89 10.09]
 [69.37 0.54 69.37 69.37 69.37 8.58]
 [68.94 3.01 68.94 68.94 68.94 8.14]
 [70.47 6.64 70.47 70.47 70.47 9.68]]

time spent until now: 1.7 mins

[pattern_seed, day, sd_R] = [2, 14, 10]

max(u_O) = 197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

```
1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
```
`O_threshold = 90`
```
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
```
`O_threshold = 100`
```
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations:  8
```
`O_threshold = 110`
```
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
```

```
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!


----------------------------------------
Value of Behaviour policy:60.789
O_threshold = 80
MC for this TARGET:[70.894, 0.091]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, 0.55, -70.89]][[-70.89, -70.89, -10.11]]
std:[[0.0, 0.53, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.89, 0.76, 70.89]][[70.89, 70.89, 10.11]]
MSE(-DR):[[0.0, -70.13, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.377, 0.097]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.38, -0.1, -69.38]][[-69.38, -69.38, -8.59]]
std:[[0.0, 0.41, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[69.38, 0.42, 69.38]][[69.38, 69.38, 8.59]]
MSE(-DR):[[0.0, -68.96, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.925, 0.09]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.92, -3.02, -68.92]][[-68.92, -68.92, -8.14]]
std:[[0.0, 0.46, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[68.92, 3.05, 68.92]][[68.92, 68.92, 8.14]]
```

MSE(-DR):[[0.0, -65.87, 0.0]][[0.0, 0.0, -60.78]]
***
==============

O_threshold = 110
MC for this TARGET:[70.467, 0.083]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -6.73, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.55, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.47, 6.75, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -63.72, 0.0]][[0.0, 0.0, -60.79]]
***
==============

******************* THIS SETTING IS GOOD *******************
[[70.88 0.64 70.88 70.88 70.88 10.1]
 [69.37 0.71 69.37 69.37 69.37 8.59]
 [68.94 2.95 68.94 68.94 68.94 8.15]
 [70.48 6.56 70.48 70.48 70.48 9.7]]

[[70.89 0.67 70.89 70.89 70.89 10.09]
 [69.37 0.54 69.37 69.37 69.37 8.58]
 [68.94 3.01 68.94 68.94 68.94 8.14]
 [70.47 6.64 70.47 70.47 70.47 9.68]]

[[70.89 0.76 70.89 70.89 70.89 10.11]
 [69.38 0.42 69.38 69.38 69.38 8.59]
 [68.92 3.05 68.92 68.92 68.92 8.14]
 [70.47 6.75 70.47 70.47 70.47 9.68]]

time spent until now: 3.9 mins

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

max(u_O) = 197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations: 15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations: 12
O_threshold = 100
target policy:

0 0 0 1 0

0 1 0 0 0

```
1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations:  8
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
```

```
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!


----------------------------------------
Value of Behaviour policy:60.787
O_threshold = 80
MC for this TARGET:[70.887, 0.239]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, 0.12, -70.89]][[-70.89, -70.89, -10.1]]
std:[[0.0, 1.0, 0.0]][[0.0, 0.0, 0.31]]
MSE:[[70.89, 1.01, 70.89]][[70.89, 70.89, 10.1]]
MSE(-DR):[[0.0, -69.88, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 90
MC for this TARGET:[69.375, 0.232]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.38, -0.3, -69.38]][[-69.38, -69.38, -8.59]]
std:[[0.0, 1.09, 0.0]][[0.0, 0.0, 0.31]]
MSE:[[69.38, 1.13, 69.38]][[69.38, 69.38, 8.6]]
MSE(-DR):[[0.0, -68.25, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 100
MC for this TARGET:[68.943, 0.229]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -2.87, -68.94]][[-68.94, -68.94, -8.16]]
std:[[0.0, 1.08, 0.0]][[0.0, 0.0, 0.31]]
MSE:[[68.94, 3.07, 68.94]][[68.94, 68.94, 8.17]]
MSE(-DR):[[0.0, -65.87, 0.0]][[0.0, 0.0, -60.77]]
***
==============


O_threshold = 110
MC for this TARGET:[70.487, 0.229]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.49, -6.56, -70.49]][[-70.49, -70.49, -9.7]]
std:[[0.0, 1.07, 0.0]][[0.0, 0.0, 0.31]]
MSE:[[70.49, 6.65, 70.49]][[70.49, 70.49, 9.7]]
MSE(-DR):[[0.0, -63.84, 0.0]][[0.0, 0.0, -60.79]]
***
==============


******************* THIS SETTING IS GOOD *******************
[[70.88  0.64 70.88 70.88 70.88 10.1 ]
 [69.37  0.71 69.37 69.37 69.37  8.59]
 [68.94  2.95 68.94 68.94 68.94  8.15]
 [70.48  6.56 70.48 70.48 70.48  9.7 ]]


[[70.89  0.67 70.89 70.89 70.89 10.09]
 [69.37  0.54 69.37 69.37 69.37  8.58]
 [68.94  3.01 68.94 68.94 68.94  8.14]
```

```
 [70.47  6.64 70.47 70.47 70.47  9.68]]


[[70.89  0.76 70.89 70.89 70.89 10.11]
 [69.38  0.42 69.38 69.38 69.38  8.59]
 [68.92  3.05 68.92 68.92 68.92  8.14]
 [70.47  6.75 70.47 70.47 70.47  9.68]]


[[70.89  1.01 70.89 70.89 70.89 10.1 ]
 [69.38  1.13 69.38 69.38 69.38  8.6 ]
 [68.94  3.07 68.94 68.94 68.94  8.17]
 [70.49  6.65 70.49 70.49 70.49  9.7 ]]


time spent until now: 4.5 mins


_____
[pattern_seed, day, sd_R] = [2, 10, 20]

max(u_0) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 100
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations:  8
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0
```

```
number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
```

```
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!


_____
Value of Behaviour policy:60.789
O_threshold = 80
MC for this TARGET:[70.881, 0.169]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.88, 0.34, -70.88]][[-70.88, -70.88, -10.09]]
std:[[0.0, 0.86, 0.0]][[0.0, 0.0, 0.27]]
MSE:[[70.88, 0.92, 70.88]][[70.88, 70.88, 10.09]]
MSE(-DR):[[0.0, -69.96, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 90
MC for this TARGET:[69.368, 0.17]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -0.19, -69.37]][[-69.37, -69.37, -8.58]]
std:[[0.0, 0.83, 0.0]][[0.0, 0.0, 0.27]]
MSE:[[69.37, 0.85, 69.37]][[69.37, 69.37, 8.58]]
MSE(-DR):[[0.0, -68.52, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.931, 0.172]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.93, -2.94, -68.93]][[-68.93, -68.93, -8.14]]
std:[[0.0, 0.91, 0.0]][[0.0, 0.0, 0.27]]
MSE:[[68.93, 3.08, 68.93]][[68.93, 68.93, 8.14]]
MSE(-DR):[[0.0, -65.85, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 110
MC for this TARGET:[70.469, 0.175]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -6.55, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 1.02, 0.0]][[0.0, 0.0, 0.27]]
MSE:[[70.47, 6.63, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -63.84, 0.0]][[0.0, 0.0, -60.79]]
***
==============


****************** THIS SETTING IS GOOD ******************
[[70.88  0.64 70.88 70.88 70.88 10.1 ]
 [69.37  0.71 69.37 69.37 69.37  8.59]
 [68.94  2.95 68.94 68.94 68.94  8.15]
 [70.48  6.56 70.48 70.48 70.48  9.7 ]]


[[70.89  0.67 70.89 70.89 70.89 10.09]
 [69.37  0.54 69.37 69.37 69.37  8.58]
 [68.94  3.01 68.94 68.94 68.94  8.14]
 [70.47  6.64 70.47 70.47 70.47  9.68]]


[[70.89  0.76 70.89 70.89 70.89 10.11]
 [69.38  0.42 69.38 69.38 69.38  8.59]
 [68.92  3.05 68.92 68.92 68.92  8.14]
 [70.47  6.75 70.47 70.47 70.47  9.68]]


[[70.89  1.01 70.89 70.89 70.89 10.1 ]
 [69.38  1.13 69.38 69.38 69.38  8.6 ]
 [68.94  3.07 68.94 68.94 68.94  8.17]
 [70.49  6.65 70.49 70.49 70.49  9.7 ]]


[[70.88  0.92 70.88 70.88 70.88 10.09]
 [69.37  0.85 69.37 69.37 69.37  8.58]
 [68.93  3.08 68.93 68.93 68.93  8.14]
 [70.47  6.63 70.47 70.47 70.47  9.68]]
```

time spent until now: 5.7 mins

max(u_O) = 197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations: 15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations: 12
O_threshold = 100
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations: 8
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations: 6
1 -th region DONE!