```
Last login: Tue Mar 31 13:04:51 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" root@ec2-34-205-247-180.compute-1.amazonaws.com
Warning: Permanently added the ED25519 host key for IP address '34.205.247.180' to the list of known hosts.
Please login as the user "ubuntu" rather than the user "root".

Connection to ec2-34-205-247-180.compute-1.amazonaws.com closed.
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-34-205-247-180.compute-1.amazonaws.com
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1060-aws x86_64)

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

  System information as of Tue Mar 31 17:43:08 UTC 2020

  System load:  1.2               Processes:           225
  Usage of /:   55.9% of 15.45GB  Users logged in:     0
  Memory usage: 1%                IP address for ens5: 172.31.7.138
  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

93 packages can be updated.
43 updates are security updates.


Last login: Thu Mar  5 21:23:34 2020 from 107.13.161.147
ubuntu@ip-172-31-7-138:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1
ubuntu@ip-172-31-7-138:~$ python EC2.py
13:45, 03/31; num of cores:16

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, lam, simple, M_in_R, u_O_u_D, mean_reversion, day_range, thre_range] = [None, 10,
10, 5, 0.2, 0.5, 1, 0.0001, False, True, 10, False, [3, 7, 14], [80, 90, 100, 110, 120, 130]]


----------------------------------------
[pattern_seed, T, sd_R, spatial_strength] = [0, 672, 5, 0]

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0
```

```
0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
O_threshold = 130
target policy:

1 0 0 1 1

0 0 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  6
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:80.115
O_threshold = 80
MC for this TARGET:[80.068, 0.077]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.01, nan]][[nan, 0.06, nan]][[-80.07, -80.07, -80.07]][0.05]
std:[[nan, 0.3, nan]][[nan, 0.01, nan]][[0.0, 0.0, 0.0]][0.02]
MSE:[[nan, 0.3, nan]][[nan, 0.06, nan]][[80.07, 80.07, 80.07]][0.05]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


O_threshold = 90
MC for this TARGET:[80.068, 0.077]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.01, nan]][[nan, 0.11, nan]][[-80.07, -80.07, -80.07]][0.05]
std:[[nan, 0.17, nan]][[nan, 0.02, nan]][[0.0, 0.0, 0.0]][0.02]
MSE:[[nan, 0.17, nan]][[nan, 0.11, nan]][[80.07, 80.07, 80.07]][0.05]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.0, nan]][[nan, 0.05, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.14, nan]][[nan, 0.03, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.14, nan]][[nan, 0.06, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============
```

```
O_threshold = 100
MC for this TARGET:[80.068, 0.077]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.02, nan]][[nan, 0.16, nan]][[-80.07, -80.07, -80.07]][0.05]
std:[[nan, 0.07, nan]][[nan, 0.02, nan]][[0.0, 0.0, 0.0]][0.02]
MSE:[[nan, 0.07, nan]][[nan, 0.16, nan]][[80.07, 80.07, 80.07]][0.05]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.03, nan]][[nan, 0.1, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.36, nan]][[nan, 0.04, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.36, nan]][[nan, 0.11, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


O_threshold = 110
MC for this TARGET:[80.068, 0.077]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.05, nan]][[nan, 0.09, nan]][[-80.07, -80.07, -80.07]][0.05]
std:[[nan, 0.07, nan]][[nan, 0.05, nan]][[0.0, 0.0, 0.0]][0.02]
MSE:[[nan, 0.09, nan]][[nan, 0.1, nan]][[80.07, 80.07, 80.07]][0.05]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.06, nan]][[nan, 0.03, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.37, nan]][[nan, 0.06, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.37, nan]][[nan, 0.07, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


O_threshold = 120
MC for this TARGET:[80.068, 0.077]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.05, nan]][[nan, 0.1, nan]][[-80.07, -80.07, -80.07]][0.05]
std:[[nan, 0.14, nan]][[nan, 0.01, nan]][[0.0, 0.0, 0.0]][0.02]
MSE:[[nan, 0.15, nan]][[nan, 0.1, nan]][[80.07, 80.07, 80.07]][0.05]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.06, nan]][[nan, 0.04, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.36, nan]][[nan, 0.02, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.36, nan]][[nan, 0.04, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


O_threshold = 130
MC for this TARGET:[80.068, 0.077]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.04, nan]][[nan, 0.11, nan]][[-80.07, -80.07, -80.07]][0.05]
std:[[nan, 0.19, nan]][[nan, 0.0, nan]][[0.0, 0.0, 0.0]][0.02]
MSE:[[nan, 0.19, nan]][[nan, 0.11, nan]][[80.07, 80.07, 80.07]][0.05]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.03, nan]][[nan, 0.05, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.41, nan]][[nan, 0.02, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.41, nan]][[nan, 0.05, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


time spent until now: 14.0 mins


----------------------------------------
[pattern_seed, T, sd_R, spatial_strength] = [0, 672, 5, 0.5]

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1
```

```
1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
O_threshold = 130
target policy:

1 0 0 1 1

0 0 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  6
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

--------------------------------------
Value of Behaviour policy:80.962
O_threshold = 80
MC for this TARGET:[83.719, 0.068]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

bias:[[2.03, 1.98, 2.15]][[3.7, 3.61, 3.52]][[-83.72, -83.72, -83.72]][-2.76]
std:[[0.25, 0.26, 0.27]][[0.2, 0.2, 0.14]][[0.0, 0.0, 0.0]][0.07]
MSE:[[2.05, 2.0, 2.17]][[3.71, 3.62, 3.52]][[83.72, 83.72, 83.72]][2.76]
MSE(-DR):[[0.0, -0.05, 0.12]][[1.66, 1.57, 1.47]][[81.67, 81.67, 81.67]][0.71]
***
==========

O_threshold = 90
MC for this TARGET:[82.878, 0.062]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[2.15, 2.1, 1.73]][[3.91, 3.81, 3.72]][[-82.88, -82.88, -82.88]][-1.92]
std:[[0.08, 0.07, 0.25]][[0.27, 0.26, 0.22]][[0.0, 0.0, 0.0]][0.07]
MSE:[[2.15, 2.1, 1.75]][[3.92, 3.82, 3.73]][[82.88, 82.88, 82.88]][1.92]
MSE(-DR):[[0.0, -0.05, -0.4]][[1.77, 1.67, 1.58]][[80.73, 80.73, 80.73]][-0.23]
**
MC-based ATE = -0.84
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.13, 0.12, -0.43]][[0.21, 0.2, 0.21]][[0.84, 0.84, 0.84]][0.84]
std:[[0.21, 0.22, 0.15]][[0.07, 0.07, 0.09]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.25, 0.25, 0.46]][[0.22, 0.21, 0.23]][[0.84, 0.84, 0.84]][0.84]
MSE(-DR):[[0.0, 0.0, 0.21]][[-0.03, -0.04, -0.02]][[0.59, 0.59, 0.59]][0.59]
==========

O_threshold = 100
MC for this TARGET:[86.564, 0.062]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.21, 0.14, -0.44]][[1.32, 1.19, 0.97]][[-86.56, -86.56, -86.56]][-5.6]
std:[[0.19, 0.19, 0.24]][[0.21, 0.22, 0.17]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.28, 0.24, 0.5]][[1.34, 1.21, 0.98]][[86.56, 86.56, 86.56]][5.6]
MSE(-DR):[[0.0, -0.04, 0.22]][[1.06, 0.93, 0.7]][[86.28, 86.28, 86.28]][5.32]
***
MC-based ATE = 2.84
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.82, -1.84, -2.59]][[-2.38, -2.42, -2.54]][[-2.84, -2.84, -2.84]][-2.84]
std:[[0.45, 0.44, 0.35]][[0.09, 0.09, 0.03]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.87, 1.89, 2.61]][[2.38, 2.42, 2.54]][[2.84, 2.84, 2.84]][2.84]
MSE(-DR):[[0.0, 0.02, 0.74]][[0.51, 0.55, 0.67]][[0.97, 0.97, 0.97]][0.97]
*
==========

O_threshold = 110
MC for this TARGET:[85.173, 0.058]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.15, -1.2, -1.21]][[-1.36, -1.43, -1.68]][[-85.17, -85.17, -85.17]][-4.21]
std:[[0.03, 0.04, 0.1]][[0.14, 0.13, 0.14]][[0.0, 0.0, 0.0]][0.07]
MSE:[[1.15, 1.2, 1.21]][[1.37, 1.44, 1.69]][[85.17, 85.17, 85.17]][4.21]
MSE(-DR):[[0.0, 0.05, 0.06]][[0.22, 0.29, 0.54]][[84.02, 84.02, 84.02]][3.06]
***
MC-based ATE = 1.45
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.18, -3.18, -3.36]][[-5.06, -5.04, -5.19]][[-1.45, -1.45, -1.45]][-1.45]
std:[[0.27, 0.27, 0.22]][[0.08, 0.09, 0.0]][[0.0, 0.0, 0.0]][0.0]
MSE:[[3.19, 3.19, 3.37]][[5.06, 5.04, 5.19]][[1.45, 1.45, 1.45]][1.45]
MSE(-DR):[[0.0, 0.0, 0.18]][[1.87, 1.85, 2.0]][[-1.74, -1.74, -1.74]][-1.74]
*
==========

O_threshold = 120
MC for this TARGET:[85.445, 0.06]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.34, -3.35, -3.28]][[-3.82, -3.85, -4.13]][[-85.44, -85.44, -85.44]][-4.48]
std:[[0.36, 0.35, 0.11]][[0.08, 0.07, 0.09]][[0.0, 0.0, 0.0]][0.07]
MSE:[[3.36, 3.37, 3.28]][[3.82, 3.85, 4.13]][[85.44, 85.44, 85.44]][4.48]
MSE(-DR):[[0.0, 0.01, -0.08]][[0.46, 0.49, 0.77]][[82.08, 82.08, 82.08]][1.12]
**
MC-based ATE = 1.73
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.37, -5.33, -5.43]][[-7.52, -7.46, -7.64]][[-1.73, -1.73, -1.73]][-1.73]
std:[[0.45, 0.43, 0.18]][[0.13, 0.13, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[5.39, 5.35, 5.43]][[7.52, 7.46, 7.64]][[1.73, 1.73, 1.73]][1.73]
MSE(-DR):[[0.0, -0.04, 0.04]][[2.13, 2.07, 2.25]][[-3.66, -3.66, -3.66]][-3.66]
*
==========

O_threshold = 130
MC for this TARGET:[86.088, 0.059]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.65, -4.65, -4.06]][[-5.66, -5.68, -5.93]][[-86.09, -86.09, -86.09]][-5.13]
std:[[0.26, 0.25, 0.03]][[0.04, 0.02, 0.09]][[0.0, 0.0, 0.0]][0.07]
MSE:[[4.66, 4.66, 4.06]][[5.66, 5.68, 5.93]][[86.09, 86.09, 86.09]][5.13]
MSE(-DR):[[0.0, 0.0, -0.6]][[1.0, 1.02, 1.27]][[81.43, 81.43, 81.43]][0.47]
**
MC-based ATE = 2.37
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]

bias:[[-6.68, -6.63, -6.22]][[-9.36, -9.29, -9.44]][[-2.37, -2.37, -2.37]][-2.37]
std:[[0.33, 0.33, 0.24]][[0.17, 0.18, 0.09]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.69, 6.64, 6.22]][[9.36, 9.29, 9.44]][[2.37, 2.37, 2.37]][2.37]
MSE(-DR):[[0.0, -0.05, -0.47]][[2.67, 2.6, 2.75]][[-4.32, -4.32, -4.32]][-4.32]
**
==========

time spent until now: 27.1 mins

[pattern_seed, T, sd_R, spatial_strength] = [0, 672, 5, 1]

max(u_O) = 156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations: 24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations: 21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations: 18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations: 11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

```
0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
O_threshold = 130
target policy:

1 0 0 1 1

0 0 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  6
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

--------------------------------------
Value of Behaviour policy:74.942
O_threshold = 80
MC for this TARGET:[83.927, 0.068]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.35, 1.26, 0.73]][[3.25, 3.04, 2.62]][[-83.93, -83.93, -83.93]][-8.98]
std:[[0.22, 0.22, 0.19]][[0.24, 0.22, 0.17]][[0.0, 0.0, 0.0]][0.1]
MSE:[[1.37, 1.28, 0.75]][[3.26, 3.05, 2.63]][[83.93, 83.93, 83.93]][8.98]
MSE(-DR):[[0.0, -0.09, -0.62]][[1.89, 1.68, 1.26]][[82.56, 82.56, 82.56]][7.61]
**
==============


O_threshold = 90
MC for this TARGET:[82.089, 0.062]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.53, 1.46, 0.23]][[3.7, 3.49, 2.99]][[-82.09, -82.09, -82.09]][-7.15]
std:[[0.04, 0.05, 0.2]][[0.31, 0.28, 0.29]][[0.0, 0.0, 0.0]][0.1]
MSE:[[1.53, 1.46, 0.3]][[3.71, 3.5, 3.0]][[82.09, 82.09, 82.09]][7.15]
MSE(-DR):[[0.0, -0.07, -1.23]][[2.18, 1.97, 1.47]][[80.56, 80.56, 80.56]][5.62]
**
MC-based ATE = -1.84
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.18, 0.2, -0.5]][[0.44, 0.45, 0.37]][[1.84, 1.84, 1.84]][1.84]
std:[[0.18, 0.21, 0.14]][[0.1, 0.09, 0.12]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.25, 0.29, 0.52]][[0.45, 0.46, 0.39]][[1.84, 1.84, 1.84]][1.84]
MSE(-DR):[[0.0, 0.04, 0.27]][[0.2, 0.21, 0.14]][[1.59, 1.59, 1.59]][1.59]
*
==============


O_threshold = 100
MC for this TARGET:[85.631, 0.063]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.03, -1.14, -3.1]][[0.75, 0.51, -0.32]][[-85.63, -85.63, -85.63]][-10.69]
std:[[0.42, 0.45, 0.18]][[0.22, 0.22, 0.23]][[0.0, 0.0, 0.0]][0.1]
MSE:[[1.11, 1.23, 3.11]][[0.78, 0.56, 0.39]][[85.63, 85.63, 85.63]][10.69]
MSE(-DR):[[0.0, 0.12, 2.0]][[-0.33, -0.55, -0.72]][[84.52, 84.52, 84.52]][9.58]
MC-based ATE = 1.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.38, -2.4, -3.83]][[-2.51, -2.53, -2.94]][[-1.7, -1.7, -1.7]][-1.7]
std:[[0.62, 0.65, 0.28]][[0.07, 0.06, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.46, 2.49, 3.84]][[2.51, 2.53, 2.94]][[1.7, 1.7, 1.7]][1.7]
MSE(-DR):[[0.0, 0.03, 1.38]][[0.05, 0.07, 0.48]][[-0.76, -0.76, -0.76]][-0.76]
*
==============


O_threshold = 110
MC for this TARGET:[83.146, 0.054]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.76, -2.85, -3.43]][[-3.01, -3.15, -4.03]][[-83.15, -83.15, -83.15]][-8.2]
std:[[0.05, 0.05, 0.2]][[0.16, 0.14, 0.22]][[0.0, 0.0, 0.0]][0.1]
MSE:[[2.76, 2.85, 3.44]][[3.01, 3.15, 4.04]][[83.15, 83.15, 83.15]][8.2]
MSE(-DR):[[0.0, 0.09, 0.68]][[0.25, 0.39, 1.28]][[80.39, 80.39, 80.39]][5.44]
***
MC-based ATE = -0.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.11, -4.11, -4.16]][[-6.26, -6.19, -6.65]][[0.78, 0.78, 0.78]][0.78]
std:[[0.2, 0.2, 0.22]][[0.08, 0.1, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[4.11, 4.11, 4.17]][[6.26, 6.19, 6.65]][[0.78, 0.78, 0.78]][0.78]
MSE(-DR):[[0.0, 0.0, 0.06]][[2.15, 2.08, 2.54]][[-3.33, -3.33, -3.33]][-3.33]
*
```

==========

O_threshold = 120
MC for this TARGET:[83.838, 0.053]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.85, -6.88, -6.88]][[-7.03, -7.13, -7.97]][[-83.84, -83.84, -83.84]][-8.9]
std:[[0.34, 0.32, 0.17]][[0.13, 0.11, 0.22]][[0.0, 0.0, 0.0]][0.1]
MSE:[[6.86, 6.89, 6.88]][[7.03, 7.13, 7.97]][[83.84, 83.84, 83.84]][8.9]
MSE(-DR):[[0.0, 0.03, 0.02]][[0.17, 0.27, 1.11]][[76.98, 76.98, 76.98]][2.04]
***
MC-based ATE = -0.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.2, -8.14, -7.61]][[-10.29, -10.17, -10.59]][[0.09, 0.09, 0.09]][0.09]
std:[[0.43, 0.41, 0.09]][[0.1, 0.11, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.21, 8.15, 7.61]][[10.29, 10.17, 10.59]][[0.09, 0.09, 0.09]][0.09]
MSE(-DR):[[0.0, -0.06, -0.6]][[2.08, 1.96, 2.38]][[-8.12, -8.12, -8.12]][-8.12]
**
==========

O_threshold = 130
MC for this TARGET:[86.09, 0.057]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.6, -10.59, -9.5]][[-11.27, -11.33, -12.17]][[-86.09, -86.09, -86.09]][-11.15]
std:[[0.13, 0.1, 0.02]][[0.07, 0.06, 0.16]][[0.0, 0.0, 0.0]][0.1]
MSE:[[10.6, 10.59, 9.5]][[11.27, 11.33, 12.17]][[86.09, 86.09, 86.09]][11.15]
MSE(-DR):[[0.0, -0.01, -1.1]][[0.67, 0.73, 1.57]][[75.49, 75.49, 75.49]][0.55]
**
MC-based ATE = 2.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.95, -11.85, -10.23]][[-14.52, -14.38, -14.79]][[-2.16, -2.16, -2.16]][-2.16]
std:[[0.27, 0.27, 0.19]][[0.16, 0.16, 0.03]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.95, 11.85, 10.23]][[14.52, 14.38, 14.79]][[2.16, 2.16, 2.16]][2.16]
MSE(-DR):[[0.0, -0.1, -1.72]][[2.57, 2.43, 2.84]][[-9.79, -9.79, -9.79]][-9.79]
**
==========

time spent until now: 40.6 mins

[pattern_seed, T, sd_R, spatial_strength] = [0, 672, 5, 1.5]

max(u_O) = 156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations: 24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations: 21
O_threshold = 100
target policy:

```
1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
O_threshold = 130
target policy:

1 0 0 1 1

0 0 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  6
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:68.417
O_threshold = 80
MC for this TARGET:[83.485, 0.069]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.04, -0.07, -1.6]][[2.02, 1.69, 0.89]][[-83.48, -83.48, -83.48]][-15.07]
std:[[0.3, 0.3, 0.11]][[0.24, 0.22, 0.14]][[0.0, 0.0, 0.0]][0.09]
MSE:[[0.3, 0.31, 1.6]][[2.03, 1.7, 0.9]][[83.48, 83.48, 83.48]][15.07]
MSE(-DR):[[0.0, 0.01, 1.3]][[1.73, 1.4, 0.6]][[83.18, 83.18, 83.18]][14.77]
***
==============


O_threshold = 90
MC for this TARGET:[80.827, 0.064]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.36, 0.23, -1.94]][[2.56, 2.26, 1.32]][[-80.83, -80.83, -80.83]][-12.41]
std:[[0.16, 0.15, 0.09]][[0.3, 0.28, 0.27]][[0.0, 0.0, 0.0]][0.09]
MSE:[[0.39, 0.27, 1.94]][[2.58, 2.28, 1.35]][[80.83, 80.83, 80.83]][12.41]
MSE(-DR):[[0.0, -0.12, 1.55]][[2.19, 1.89, 0.96]][[80.44, 80.44, 80.44]][12.02]
***
MC-based ATE = -2.66
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.31, 0.3, -0.34]][[0.54, 0.57, 0.43]][[2.66, 2.66, 2.66]][2.66]
std:[[0.16, 0.19, 0.12]][[0.12, 0.11, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.35, 0.36, 0.36]][[0.55, 0.58, 0.45]][[2.66, 2.66, 2.66]][2.66]
MSE(-DR):[[0.0, 0.01, 0.01]][[0.2, 0.23, 0.1]][[2.31, 2.31, 2.31]][2.31]
*
==============


O_threshold = 100
```

```
MC for this TARGET:[83.409, 0.066]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.42, -2.54, -5.46]][[-0.37, -0.7, -2.14]][[-83.41, -83.41, -83.41]][-14.99]
std:[[0.52, 0.54, 0.27]][[0.27, 0.25, 0.3]][[0.0, 0.0, 0.0]][0.09]
MSE:[[2.48, 2.6, 5.47]][[0.46, 0.74, 2.16]][[83.41, 83.41, 83.41]][14.99]
MSE(-DR):[[0.0, 0.12, 2.99]][[-2.02, -1.74, -0.32]][[80.93, 80.93, 80.93]][12.51]
MC-based ATE = -0.08
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.46, -2.48, -3.86]][[-2.39, -2.38, -3.04]][[0.08, 0.08, 0.08]][0.08]
std:[[0.72, 0.74, 0.38]][[0.06, 0.06, 0.16]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.56, 2.59, 3.88]][[2.39, 2.38, 3.04]][[0.08, 0.08, 0.08]][0.08]
MSE(-DR):[[0.0, 0.03, 1.32]][[-0.17, -0.18, 0.48]][[-2.48, -2.48, -2.48]][-2.48]
==============


O_threshold = 110
MC for this TARGET:[79.01, 0.055]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.88, -3.95, -5.08]][[-4.02, -4.22, -5.73]][[-79.01, -79.01, -79.01]][-10.59]
std:[[0.13, 0.11, 0.16]][[0.19, 0.15, 0.27]][[0.0, 0.0, 0.0]][0.09]
MSE:[[3.88, 3.95, 5.08]][[4.02, 4.22, 5.74]][[79.01, 79.01, 79.01]][10.59]
MSE(-DR):[[0.0, 0.07, 1.2]][[0.14, 0.34, 1.86]][[75.13, 75.13, 75.13]][6.71]
***
MC-based ATE = -4.47
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.92, -3.88, -3.48]][[-6.05, -5.91, -6.62]][[4.47, 4.47, 4.47]][4.47]
std:[[0.22, 0.21, 0.23]][[0.06, 0.08, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[3.93, 3.89, 3.49]][[6.05, 5.91, 6.62]][[4.47, 4.47, 4.47]][4.47]
MSE(-DR):[[0.0, -0.04, -0.44]][[2.12, 1.98, 2.69]][[0.54, 0.54, 0.54]][0.54]
**
==============


O_threshold = 120
MC for this TARGET:[79.861, 0.054]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.25, -9.26, -9.57]][[-9.08, -9.2, -10.57]][[-79.86, -79.86, -79.86]][-11.44]
std:[[0.32, 0.31, 0.14]][[0.16, 0.14, 0.26]][[0.0, 0.0, 0.0]][0.09]
MSE:[[9.26, 9.27, 9.57]][[9.08, 9.2, 10.57]][[79.86, 79.86, 79.86]][11.44]
MSE(-DR):[[0.0, 0.01, 0.31]][[-0.18, -0.06, 1.31]][[70.6, 70.6, 70.6]][2.18]
MC-based ATE = -3.62
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.29, -9.19, -7.97]][[-11.1, -10.88, -11.47]][[3.62, 3.62, 3.62]][3.62]
std:[[0.44, 0.44, 0.16]][[0.09, 0.09, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[9.3, 9.2, 7.97]][[11.1, 10.88, 11.47]][[3.62, 3.62, 3.62]][3.62]
MSE(-DR):[[0.0, -0.1, -1.33]][[1.8, 1.58, 2.17]][[-5.68, -5.68, -5.68]][-5.68]
**
==============


O_threshold = 130
MC for this TARGET:[83.595, 0.061]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-15.28, -15.27, -14.24]][[-15.41, -15.5, -16.86]][[-83.6, -83.6, -83.6]][-15.18]
std:[[0.19, 0.19, 0.03]][[0.11, 0.09, 0.26]][[0.0, 0.0, 0.0]][0.09]
MSE:[[15.28, 15.27, 14.24]][[15.41, 15.5, 16.86]][[83.6, 83.6, 83.6]][15.18]
MSE(-DR):[[0.0, -0.01, -1.04]][[0.13, 0.22, 1.58]][[68.32, 68.32, 68.32]][-0.1]
**
MC-based ATE = 0.11
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-15.32, -15.2, -12.64]][[-17.43, -17.19, -17.75]][[-0.11, -0.11, -0.11]][-0.11]
std:[[0.26, 0.25, 0.1]][[0.15, 0.14, 0.14]][[0.0, 0.0, 0.0]][0.0]
MSE:[[15.32, 15.2, 12.64]][[17.43, 17.19, 17.75]][[0.11, 0.11, 0.11]][0.11]
MSE(-DR):[[0.0, -0.12, -2.68]][[2.11, 1.87, 2.43]][[-15.21, -15.21, -15.21]][-15.21]
**
==============


time spent until now: 53.9 mins


----------------------------------------
[pattern_seed, T, sd_R, spatial_strength] = [1, 672, 5, 0]

max(u_O) =  141.0
O_threshold = 80
means of Order:

137.7 88.0 89.5 80.3 118.3

62.8 141.0 85.4 106.0 94.6

133.3 65.9 93.3 92.1 124.8

79.8 96.1 83.5 100.3 111.8

79.8 125.1 119.1 110.0 119.1

target policy:
```

```
1 1 1 1 1

0 1 1 1 1

1 0 1 1 1

0 1 1 1 1

0 1 1 1 1

number of reward locations:  21
O_threshold = 90
target policy:

1 0 0 0 1

0 1 0 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

number of reward locations:  16
O_threshold = 100
target policy:

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 1 1

0 1 1 1 1

number of reward locations:  12
O_threshold = 110
target policy:

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1

number of reward locations:  10
O_threshold = 120
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 0 0 0

number of reward locations:  5
O_threshold = 130
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  3
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

--------------------------------------
Value of Behaviour policy:67.894
O_threshold = 80
```

```
MC for this TARGET:[67.887, 0.066]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.13, nan]][[nan, 0.02, nan]][[-67.89, -67.89, -67.89]][0.01]
std:[[nan, 0.36, nan]][[nan, 0.09, nan]][[0.0, 0.0, 0.0]][0.04]
MSE:[[nan, 0.38, nan]][[nan, 0.09, nan]][[67.89, 67.89, 67.89]][0.04]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


O_threshold = 90
MC for this TARGET:[67.887, 0.066]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.14, nan]][[nan, 0.03, nan]][[-67.89, -67.89, -67.89]][0.01]
std:[[nan, 0.2, nan]][[nan, 0.11, nan]][[0.0, 0.0, 0.0]][0.04]
MSE:[[nan, 0.24, nan]][[nan, 0.11, nan]][[67.89, 67.89, 67.89]][0.04]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.01, nan]][[nan, 0.01, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.16, nan]][[nan, 0.09, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.16, nan]][[nan, 0.09, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


O_threshold = 100
MC for this TARGET:[67.887, 0.066]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, 0.04, nan]][[nan, -0.04, nan]][[-67.89, -67.89, -67.89]][0.01]
std:[[nan, 0.21, nan]][[nan, 0.09, nan]][[0.0, 0.0, 0.0]][0.04]
MSE:[[nan, 0.21, nan]][[nan, 0.1, nan]][[67.89, 67.89, 67.89]][0.04]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.09, nan]][[nan, -0.06, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.15, nan]][[nan, 0.02, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.17, nan]][[nan, 0.06, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


O_threshold = 110
MC for this TARGET:[67.887, 0.066]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.06, nan]][[nan, -0.06, nan]][[-67.89, -67.89, -67.89]][0.01]
std:[[nan, 0.14, nan]][[nan, 0.08, nan]][[0.0, 0.0, 0.0]][0.04]
MSE:[[nan, 0.15, nan]][[nan, 0.1, nan]][[67.89, 67.89, 67.89]][0.04]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.19, nan]][[nan, -0.07, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.3, nan]][[nan, 0.06, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.36, nan]][[nan, 0.09, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


O_threshold = 120
MC for this TARGET:[67.887, 0.066]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.06, nan]][[nan, -0.04, nan]][[-67.89, -67.89, -67.89]][0.01]
std:[[nan, 0.12, nan]][[nan, 0.06, nan]][[0.0, 0.0, 0.0]][0.04]
MSE:[[nan, 0.13, nan]][[nan, 0.07, nan]][[67.89, 67.89, 67.89]][0.04]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.19, nan]][[nan, -0.06, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.47, nan]][[nan, 0.11, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.51, nan]][[nan, 0.13, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============


O_threshold = 130
MC for this TARGET:[67.887, 0.066]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.02, nan]][[nan, -0.03, nan]][[-67.89, -67.89, -67.89]][0.01]
std:[[nan, 0.09, nan]][[nan, 0.06, nan]][[0.0, 0.0, 0.0]][0.04]
MSE:[[nan, 0.09, nan]][[nan, 0.07, nan]][[67.89, 67.89, 67.89]][0.04]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
MC-based ATE = 0.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[nan, -0.15, nan]][[nan, -0.05, nan]][[0.0, 0.0, 0.0]][0.0]
std:[[nan, 0.46, nan]][[nan, 0.11, nan]][[0.0, 0.0, 0.0]][0.0]
MSE:[[nan, 0.48, nan]][[nan, 0.12, nan]][[0.0, 0.0, 0.0]][0.0]
MSE(-DR):[[nan, nan, nan]][[nan, nan, nan]][[nan, nan, nan]][nan]
==============
```

time spent until now: 68.0 mins

[pattern_seed, T, sd_R, spatial_strength] = [1, 672, 5, 0.5]

max(u_O) = 141.0
O_threshold = 80
means of Order:

137.7 88.0 89.5 80.3 118.3

62.8 141.0 85.4 106.0 94.6

133.3 65.9 93.3 92.1 124.8

79.8 96.1 83.5 100.3 111.8

79.8 125.1 119.1 110.0 119.1

target policy:

1 1 1 1 1

0 1 1 1 1

1 0 1 1 1

0 1 1 1 1

0 1 1 1 1

number of reward locations: 21
O_threshold = 90
target policy:

1 0 0 0 1

0 1 0 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

number of reward locations: 16
O_threshold = 100
target policy:

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 1 1

0 1 1 1 1

number of reward locations: 12
O_threshold = 110
target policy:

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1

number of reward locations: 10
O_threshold = 120
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 0 0 0

number of reward locations: 5

```
O_threshold = 130
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  3
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:71.101
O_threshold = 80
MC for this TARGET:[73.215, 0.062]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.66, 1.62, 1.77]][[3.18, 3.09, 3.0]][[-73.22, -73.22, -73.22]][-2.11]
std:[[0.53, 0.53, 0.38]][[0.09, 0.1, 0.12]][[0.0, 0.0, 0.0]][0.13]
MSE:[[1.74, 1.7, 1.81]][[3.18, 3.09, 3.0]][[73.22, 73.22, 73.22]][2.11]
MSE(-DR):[[0.0, -0.04, 0.07]][[1.44, 1.35, 1.26]][[71.48, 71.48, 71.48]][0.37]
***
==============


O_threshold = 90
MC for this TARGET:[74.158, 0.059]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.18, 0.15, -0.15]][[1.04, 0.96, 0.84]][[-74.16, -74.16, -74.16]][-3.06]
std:[[0.4, 0.4, 0.28]][[0.08, 0.08, 0.08]][[0.0, 0.0, 0.0]][0.13]
MSE:[[0.44, 0.43, 0.32]][[1.04, 0.96, 0.84]][[74.16, 74.16, 74.16]][3.06]
MSE(-DR):[[0.0, -0.01, -0.12]][[0.6, 0.52, 0.4]][[73.72, 73.72, 73.72]][2.62]
**
MC-based ATE = 0.94
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.48, -1.47, -1.92]][[-2.13, -2.13, -2.16]][[-0.94, -0.94, -0.94]][-0.94]
std:[[0.15, 0.13, 0.11]][[0.04, 0.04, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.49, 1.48, 1.92]][[2.13, 2.13, 2.16]][[0.94, 0.94, 0.94]][0.94]
MSE(-DR):[[0.0, -0.01, 0.43]][[0.64, 0.64, 0.67]][[-0.55, -0.55, -0.55]][-0.55]
*
==============


O_threshold = 100
MC for this TARGET:[77.598, 0.057]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.22, -3.27, -3.1]][[-2.77, -2.87, -3.1]][[-77.6, -77.6, -77.6]][-6.5]
std:[[0.32, 0.33, 0.18]][[0.08, 0.08, 0.11]][[0.0, 0.0, 0.0]][0.13]
MSE:[[3.24, 3.29, 3.11]][[2.77, 2.87, 3.1]][[77.6, 77.6, 77.6]][6.5]
MSE(-DR):[[0.0, 0.05, -0.13]][[-0.47, -0.37, -0.14]][[74.36, 74.36, 74.36]][3.26]
MC-based ATE = 4.38
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.89, -4.89, -4.86]][[-5.95, -5.96, -6.09]][[-4.38, -4.38, -4.38]][-4.38]
std:[[0.3, 0.29, 0.2]][[0.06, 0.07, 0.03]][[0.0, 0.0, 0.0]][0.0]
MSE:[[4.9, 4.9, 4.86]][[5.95, 5.96, 6.09]][[4.38, 4.38, 4.38]][4.38]
MSE(-DR):[[0.0, 0.0, -0.04]][[1.05, 1.06, 1.19]][[-0.52, -0.52, -0.52]][-0.52]
**
==============


O_threshold = 110
MC for this TARGET:[79.815, 0.053]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.5, -4.56, -4.18]][[-4.97, -5.09, -5.34]][[-79.82, -79.82, -79.82]][-8.71]
std:[[0.19, 0.23, 0.21]][[0.05, 0.06, 0.11]][[0.0, 0.0, 0.0]][0.13]
MSE:[[4.5, 4.57, 4.19]][[4.97, 5.09, 5.34]][[79.82, 79.82, 79.82]][8.71]
MSE(-DR):[[0.0, 0.07, -0.31]][[0.47, 0.59, 0.84]][[75.32, 75.32, 75.32]][4.21]
**
MC-based ATE = 6.6
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.16, -6.18, -5.95]][[-8.15, -8.18, -8.34]][[-6.6, -6.6, -6.6]][-6.6]
std:[[0.45, 0.44, 0.42]][[0.06, 0.06, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.18, 6.2, 5.96]][[8.15, 8.18, 8.34]][[6.6, 6.6, 6.6]][6.6]
MSE(-DR):[[0.0, 0.02, -0.22]][[1.97, 2.0, 2.16]][[0.42, 0.42, 0.42]][0.42]
**
==============


O_threshold = 120
MC for this TARGET:[76.991, 0.058]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.52, -4.54, -4.27]][[-6.01, -6.04, -6.26]][[-76.99, -76.99, -76.99]][-5.89]
std:[[0.09, 0.09, 0.27]][[0.02, 0.02, 0.04]][[0.0, 0.0, 0.0]][0.13]
```

```
MSE:[[4.52, 4.54, 4.28]][[6.01, 6.04, 6.26]][[76.99, 76.99, 76.99]][5.89]
MSE(-DR):[[0.0, 0.02, -0.24]][[1.49, 1.52, 1.74]][[72.47, 72.47, 72.47]][1.37]
**
MC-based ATE = 3.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.19, -6.16, -6.04]][[-9.19, -9.13, -9.26]][[-3.78, -3.78, -3.78]][-3.78]
std:[[0.57, 0.55, 0.53]][[0.1, 0.11, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.22, 6.18, 6.06]][[9.19, 9.13, 9.26]][[3.78, 3.78, 3.78]][3.78]
MSE(-DR):[[0.0, -0.04, -0.16]][[2.97, 2.91, 3.04]][[-2.44, -2.44, -2.44]][-2.44]
**
==============


O_threshold = 130
MC for this TARGET:[75.247, 0.059]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.18, -4.17, -4.21]][[-6.03, -6.03, -6.22]][[-75.25, -75.25, -75.25]][-4.15]
std:[[0.05, 0.05, 0.21]][[0.06, 0.05, 0.07]][[0.0, 0.0, 0.0]][0.13]
MSE:[[4.18, 4.17, 4.22]][[6.03, 6.03, 6.22]][[75.25, 75.25, 75.25]][4.15]
MSE(-DR):[[0.0, -0.01, 0.04]][[1.85, 1.85, 2.04]][[71.07, 71.07, 71.07]][-0.03]
***
MC-based ATE = 2.03
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.84, -5.79, -5.97]][[-9.21, -9.13, -9.22]][[-2.03, -2.03, -2.03]][-2.03]
std:[[0.56, 0.53, 0.52]][[0.16, 0.15, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[5.87, 5.81, 5.99]][[9.21, 9.13, 9.22]][[2.03, 2.03, 2.03]][2.03]
MSE(-DR):[[0.0, -0.06, 0.12]][[3.34, 3.26, 3.35]][[-3.84, -3.84, -3.84]][-3.84]
*
==============


time spent until now: 82.1 mins


---------------------------------------
[pattern_seed, T, sd_R, spatial_strength] = [1, 672, 5, 1]

max(u_O) =  141.0
O_threshold = 80
means of Order:

137.7 88.0 89.5 80.3 118.3

62.8 141.0 85.4 106.0 94.6

133.3 65.9 93.3 92.1 124.8

79.8 96.1 83.5 100.3 111.8

79.8 125.1 119.1 110.0 119.1

target policy:

1 1 1 1 1

0 1 1 1 1

1 0 1 1 1

0 1 1 1 1

0 1 1 1 1

number of reward locations:  21
O_threshold = 90
target policy:

1 0 0 0 1

0 1 0 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

number of reward locations:  16
O_threshold = 100
target policy:

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 1 1
```

```
0 1 1 1 1

number of reward locations:  12
O_threshold = 110
target policy:

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1

number of reward locations:  10
O_threshold = 120
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 0 0 0

number of reward locations:  5
O_threshold = 130
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  3
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:66.742
O_threshold = 80
MC for this TARGET:[73.149, 0.062]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.56, 1.48, 0.88]][[3.61, 3.43, 2.96]][[-73.15, -73.15, -73.15]][-6.41]
std:[[0.5, 0.49, 0.42]][[0.12, 0.13, 0.13]][[0.0, 0.0, 0.0]][0.15]
MSE:[[1.64, 1.56, 0.98]][[3.61, 3.43, 2.96]][[73.15, 73.15, 73.15]][6.41]
MSE(-DR):[[0.0, -0.08, -0.66]][[1.97, 1.79, 1.32]][[71.51, 71.51, 71.51]][4.77]
**
==============


O_threshold = 90
MC for this TARGET:[73.513, 0.059]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.29, -0.38, -1.44]][[1.03, 0.84, 0.26]][[-73.51, -73.51, -73.51]][-6.77]
std:[[0.36, 0.34, 0.31]][[0.05, 0.05, 0.11]][[0.0, 0.0, 0.0]][0.15]
MSE:[[0.46, 0.51, 1.47]][[1.03, 0.84, 0.28]][[73.51, 73.51, 73.51]][6.77]
MSE(-DR):[[0.0, 0.05, 1.01]][[0.57, 0.38, -0.18]][[73.05, 73.05, 73.05]][6.31]
***
MC-based ATE = 0.36
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.86, -1.85, -2.32]][[-2.58, -2.6, -2.7]][[-0.36, -0.36, -0.36]][-0.36]
std:[[0.2, 0.19, 0.13]][[0.08, 0.08, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.87, 1.86, 2.32]][[2.58, 2.6, 2.7]][[0.36, 0.36, 0.36]][0.36]
MSE(-DR):[[0.0, -0.01, 0.45]][[0.71, 0.73, 0.83]][[-1.51, -1.51, -1.51]][-1.51]
*
==============


O_threshold = 100
MC for this TARGET:[77.165, 0.059]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.97, -5.05, -5.37]][[-4.1, -4.28, -5.12]][[-77.17, -77.17, -77.17]][-10.42]
std:[[0.32, 0.34, 0.22]][[0.12, 0.12, 0.18]][[0.0, 0.0, 0.0]][0.15]
MSE:[[4.98, 5.06, 5.37]][[4.1, 4.28, 5.12]][[77.17, 77.17, 77.17]][10.42]
MSE(-DR):[[0.0, 0.08, 0.39]][[-0.88, -0.7, 0.14]][[72.19, 72.19, 72.19]][5.44]
MC-based ATE = 4.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

```
bias:[[-6.53, -6.53, -6.25]][[-7.71, -7.72, -8.08]][[-4.02, -4.02, -4.02]][-4.02]
std:[[0.52, 0.51, 0.21]][[0.1, 0.11, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.55, 6.55, 6.25]][[7.71, 7.72, 8.08]][[4.02, 4.02, 4.02]][4.02]
MSE(-DR):[[0.0, 0.0, -0.3]][[1.16, 1.17, 1.53]][[-2.53, -2.53, -2.53]][-2.53]
**
==============


O_threshold = 110
MC for this TARGET:[80.266, 0.057]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.47, -7.56, -7.46]][[-7.78, -7.97, -9.01]][[-80.27, -80.27, -80.27]][-13.52]
std:[[0.3, 0.31, 0.33]][[0.14, 0.15, 0.2]][[0.0, 0.0, 0.0]][0.15]
MSE:[[7.48, 7.57, 7.47]][[7.78, 7.97, 9.01]][[80.27, 80.27, 80.27]][13.52]
MSE(-DR):[[0.0, 0.09, -0.01]][[0.3, 0.49, 1.53]][[72.79, 72.79, 72.79]][6.04]
**
MC-based ATE = 7.12
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.03, -9.04, -8.35]][[-11.39, -11.4, -11.97]][[-7.12, -7.12, -7.12]][-7.12]
std:[[0.55, 0.56, 0.44]][[0.11, 0.12, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[9.05, 9.06, 8.36]][[11.39, 11.4, 11.97]][[7.12, 7.12, 7.12]][7.12]
MSE(-DR):[[0.0, 0.01, -0.69]][[2.34, 2.35, 2.92]][[-1.93, -1.93, -1.93]][-1.93]
**
==============


O_threshold = 120
MC for this TARGET:[78.017, 0.061]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.6, -9.6, -9.54]][[-11.13, -11.21, -12.04]][[-78.02, -78.02, -78.02]][-11.27]
std:[[0.22, 0.23, 0.16]][[0.08, 0.08, 0.13]][[0.0, 0.0, 0.0]][0.15]
MSE:[[9.6, 9.6, 9.54]][[11.13, 11.21, 12.04]][[78.02, 78.02, 78.02]][11.27]
MSE(-DR):[[0.0, 0.0, -0.06]][[1.53, 1.61, 2.44]][[68.42, 68.42, 68.42]][1.67]
**
MC-based ATE = 4.87
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.16, -11.08, -10.42]][[-14.74, -14.64, -14.99]][[-4.87, -4.87, -4.87]][-4.87]
std:[[0.33, 0.35, 0.45]][[0.15, 0.15, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.16, 11.09, 10.43]][[14.74, 14.64, 14.99]][[4.87, 4.87, 4.87]][4.87]
MSE(-DR):[[0.0, -0.07, -0.73]][[3.58, 3.48, 3.83]][[-6.29, -6.29, -6.29]][-6.29]
**
==============


O_threshold = 130
MC for this TARGET:[75.728, 0.062]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.45, -9.44, -9.69]][[-11.49, -11.51, -12.24]][[-75.73, -75.73, -75.73]][-8.99]
std:[[0.23, 0.24, 0.16]][[0.12, 0.12, 0.16]][[0.0, 0.0, 0.0]][0.15]
MSE:[[9.45, 9.44, 9.69]][[11.49, 11.51, 12.24]][[75.73, 75.73, 75.73]][8.99]
MSE(-DR):[[0.0, -0.01, 0.24]][[2.04, 2.06, 2.79]][[66.28, 66.28, 66.28]][-0.46]
***
MC-based ATE = 2.58
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.01, -10.91, -10.57]][[-15.11, -14.94, -15.2]][[-2.58, -2.58, -2.58]][-2.58]
std:[[0.3, 0.3, 0.43]][[0.21, 0.21, 0.18]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.01, 10.91, 10.58]][[15.11, 14.94, 15.2]][[2.58, 2.58, 2.58]][2.58]
MSE(-DR):[[0.0, -0.1, -0.43]][[4.1, 3.93, 4.19]][[-8.43, -8.43, -8.43]][-8.43]
**
==============


time spent until now: 96.2 mins


----------------------------------------
[pattern_seed, T, sd_R, spatial_strength] = [1, 672, 5, 1.5]

max(u_O) =  141.0
O_threshold = 80
means of Order:

137.7 88.0 89.5 80.3 118.3

62.8 141.0 85.4 106.0 94.6

133.3 65.9 93.3 92.1 124.8

79.8 96.1 83.5 100.3 111.8

79.8 125.1 119.1 110.0 119.1

target policy:

1 1 1 1 1

0 1 1 1 1

1 0 1 1 1
```

```
0 1 1 1 1

0 1 1 1 1

number of reward locations:  21
O_threshold = 90
target policy:

1 0 0 0 1

0 1 0 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

number of reward locations:  16
O_threshold = 100
target policy:

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 1 1

0 1 1 1 1

number of reward locations:  12
O_threshold = 110
target policy:

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1

number of reward locations:  10
O_threshold = 120
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 0 0 0

number of reward locations:  5
O_threshold = 130
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  3
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target;
```