```
Last login: Tue Mar 31 16:44:36 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-221-170-144.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 21:22:36 UTC 2020

  System load:  0.03               Processes:           209
  Usage of /:   56.9% of 15.45GB   Users logged in:     0
  Memory usage: 1%                 IP address for ens5: 172.31.10.67
  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.


*** System restart required ***
Last login: Tue Mar 31 20:46:34 2020 from 107.13.161.147
ubuntu@ip-172-31-10-67:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
17:22, 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, 1, 1, 0.0001, False, True, 10, False, [3, 7, 14], [80, 90, 100, 110, 120, 130]]


--------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 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:74.925
O_threshold = 80
MC for this TARGET:[83.929, 0.059]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.44, 1.37, 0.73]][[3.2, 3.06, 2.5]][[-83.93, -83.93, -83.93]][-9.0]
std:[[0.21, 0.21, 0.12]][[0.22, 0.21, 0.19]][[0.0, 0.0, 0.0]][0.09]
MSE:[[1.46, 1.39, 0.74]][[3.21, 3.07, 2.51]][[83.93, 83.93, 83.93]][9.0]
MSE(-DR):[[0.0, -0.07, -0.72]][[1.75, 1.61, 1.05]][[82.47, 82.47, 82.47]][7.54]
**
==============


O_threshold = 90
MC for this TARGET:[82.09, 0.054]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.55, 1.49, 0.13]][[3.65, 3.5, 2.83]][[-82.09, -82.09, -82.09]][-7.16]
std:[[0.14, 0.16, 0.21]][[0.23, 0.23, 0.23]][[0.0, 0.0, 0.0]][0.09]
MSE:[[1.56, 1.5, 0.25]][[3.66, 3.51, 2.84]][[82.09, 82.09, 82.09]][7.16]
MSE(-DR):[[0.0, -0.06, -1.31]][[2.1, 1.95, 1.28]][[80.53, 80.53, 80.53]][5.6]
**
MC-based ATE = -1.84
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.11, 0.12, -0.59]][[0.45, 0.44, 0.32]][[1.84, 1.84, 1.84]][1.84]
std:[[0.16, 0.16, 0.12]][[0.04, 0.04, 0.04]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.19, 0.2, 0.6]][[0.45, 0.44, 0.32]][[1.84, 1.84, 1.84]][1.84]
MSE(-DR):[[0.0, 0.01, 0.41]][[0.26, 0.25, 0.13]][[1.65, 1.65, 1.65]][1.65]
*
==============
O_threshold = 100
MC for this TARGET:[85.633, 0.052]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.0, -1.08, -3.19]][[0.61, 0.47, -0.68]][[-85.63, -85.63, -85.63]][-10.71]
std:[[0.45, 0.46, 0.21]][[0.2, 0.18, 0.24]][[0.0, 0.0, 0.0]][0.09]
MSE:[[1.1, 1.17, 3.2]][[0.64, 0.5, 0.72]][[85.63, 85.63, 85.63]][10.71]
MSE(-DR):[[0.0, 0.07, 2.1]][[-0.46, -0.6, -0.38]][[84.53, 84.53, 84.53]][9.61]
MC-based ATE = 1.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.44, -2.45, -3.92]][[-2.6, -2.59, -3.18]][[-1.7, -1.7, -1.7]][-1.7]
std:[[0.59, 0.59, 0.27]][[0.03, 0.03, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.51, 2.52, 3.93]][[2.6, 2.59, 3.18]][[1.7, 1.7, 1.7]][1.7]
MSE(-DR):[[0.0, 0.01, 1.42]][[0.09, 0.08, 0.67]][[-0.81, -0.81, -0.81]][-0.81]
*
==============


O_threshold = 110
MC for this TARGET:[83.148, 0.043]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.75, -2.8, -3.6]][[-3.1, -3.18, -4.34]][[-83.15, -83.15, -83.15]][-8.22]
std:[[0.09, 0.08, 0.19]][[0.15, 0.12, 0.24]][[0.0, 0.0, 0.0]][0.09]
MSE:[[2.75, 2.8, 3.61]][[3.1, 3.18, 4.35]][[83.15, 83.15, 83.15]][8.22]
MSE(-DR):[[0.0, 0.05, 0.86]][[0.35, 0.43, 1.6]][[80.4, 80.4, 80.4]][5.47]
***
MC-based ATE = -0.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.19, -4.17, -4.32]][[-6.3, -6.24, -6.84]][[0.78, 0.78, 0.78]][0.78]
std:[[0.29, 0.3, 0.07]][[0.07, 0.09, 0.09]][[0.0, 0.0, 0.0]][0.0]
MSE:[[4.2, 4.18, 4.32]][[6.3, 6.24, 6.84]][[0.78, 0.78, 0.78]][0.78]
MSE(-DR):[[0.0, -0.02, 0.12]][[2.1, 2.04, 2.64]][[-3.42, -3.42, -3.42]][-3.42]
*
==============


O_threshold = 120
MC for this TARGET:[83.839, 0.044]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.8, -6.82, -6.97]][[-7.12, -7.18, -8.23]][[-83.84, -83.84, -83.84]][-8.91]
std:[[0.35, 0.35, 0.12]][[0.1, 0.09, 0.17]][[0.0, 0.0, 0.0]][0.09]
MSE:[[6.81, 6.83, 6.97]][[7.12, 7.18, 8.23]][[83.84, 83.84, 83.84]][8.91]
MSE(-DR):[[0.0, 0.02, 0.16]][[0.31, 0.37, 1.42]][[77.03, 77.03, 77.03]][2.1]
***
MC-based ATE = -0.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.24, -8.19, -7.7]][[-10.33, -10.24, -10.73]][[0.09, 0.09, 0.09]][0.09]
std:[[0.52, 0.53, 0.01]][[0.12, 0.13, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.26, 8.21, 7.7]][[10.33, 10.24, 10.73]][[0.09, 0.09, 0.09]][0.09]
MSE(-DR):[[0.0, -0.05, -0.56]][[2.07, 1.98, 2.47]][[-8.17, -8.17, -8.17]][-8.17]
**
==============


O_threshold = 130
MC for this TARGET:[86.092, 0.046]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.56, -10.57, -9.45]][[-11.36, -11.39, -12.47]][[-86.09, -86.09, -86.09]][-11.17]
std:[[0.22, 0.21, 0.01]][[0.06, 0.05, 0.15]][[0.0, 0.0, 0.0]][0.09]
MSE:[[10.56, 10.57, 9.45]][[11.36, 11.39, 12.47]][[86.09, 86.09, 86.09]][11.17]
MSE(-DR):[[0.0, 0.01, -1.11]][[0.8, 0.83, 1.91]][[75.53, 75.53, 75.53]][0.61]
**
MC-based ATE = 2.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-12.0, -11.94, -10.17]][[-14.56, -14.45, -14.98]][[-2.16, -2.16, -2.16]][-2.16]
std:[[0.42, 0.42, 0.13]][[0.16, 0.16, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[12.01, 11.95, 10.17]][[14.56, 14.45, 14.98]][[2.16, 2.16, 2.16]][2.16]
MSE(-DR):[[0.0, -0.06, -1.84]][[2.55, 2.44, 2.97]][[-9.85, -9.85, -9.85]][-9.85]
**
==============


time spent until now: 10.2 mins


--------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 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: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:[[1.33, 1.26, 0.67]][[3.19, 3.04, 2.49]][[-83.93, -83.93, -83.93]][-8.98]
std:[[0.23, 0.22, 0.18]][[0.23, 0.22, 0.17]][[0.0, 0.0, 0.0]][0.1]
MSE:[[1.35, 1.28, 0.69]][[3.2, 3.05, 2.5]][[83.93, 83.93, 83.93]][8.98]
MSE(-DR):[[0.0, -0.07, -0.66]][[1.85, 1.7, 1.15]][[82.58, 82.58, 82.58]][7.63]
**
==============


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:[[1.52, 1.46, 0.11]][[3.63, 3.49, 2.83]][[-82.09, -82.09, -82.09]][-7.15]
std:[[0.04, 0.05, 0.17]][[0.28, 0.28, 0.26]][[0.0, 0.0, 0.0]][0.1]
MSE:[[1.52, 1.46, 0.2]][[3.64, 3.5, 2.84]][[82.09, 82.09, 82.09]][7.15]
MSE(-DR):[[0.0, -0.06, -1.32]][[2.12, 1.98, 1.32]][[80.57, 80.57, 80.57]][5.63]
**
MC-based ATE = -1.84
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.19, 0.2, -0.57]][[0.45, 0.45, 0.34]][[1.84, 1.84, 1.84]][1.84]
std:[[0.2, 0.21, 0.15]][[0.08, 0.09, 0.09]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.28, 0.29, 0.59]][[0.46, 0.46, 0.35]][[1.84, 1.84, 1.84]][1.84]
MSE(-DR):[[0.0, 0.01, 0.31]][[0.18, 0.18, 0.07]][[1.56, 1.56, 1.56]][1.56]
*
==============


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:[[-1.06, -1.14, -3.23]][[0.68, 0.51, -0.55]][[-85.63, -85.63, -85.63]][-10.69]
std:[[0.45, 0.45, 0.25]][[0.22, 0.22, 0.2]][[0.0, 0.0, 0.0]][0.1]
MSE:[[1.15, 1.23, 3.24]][[0.71, 0.56, 0.59]][[85.63, 85.63, 85.63]][10.69]
MSE(-DR):[[0.0, 0.08, 2.09]][[-0.44, -0.59, -0.56]][[84.48, 84.48, 84.48]][9.54]
MC-based ATE = 1.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.39, -2.4, -3.9]][[-2.51, -2.53, -3.04]][[-1.7, -1.7, -1.7]][-1.7]
std:[[0.65, 0.65, 0.36]][[0.05, 0.06, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.48, 2.49, 3.92]][[2.51, 2.53, 3.04]][[1.7, 1.7, 1.7]][1.7]
MSE(-DR):[[0.0, 0.01, 1.44]][[0.03, 0.05, 0.56]][[-0.78, -0.78, -0.78]][-0.78]
*
==============


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:[[-2.79, -2.85, -3.59]][[-3.05, -3.15, -4.26]][[-83.15, -83.15, -83.15]][-8.2]
std:[[0.06, 0.05, 0.14]][[0.16, 0.14, 0.21]][[0.0, 0.0, 0.0]][0.1]
MSE:[[2.79, 2.85, 3.59]][[3.05, 3.15, 4.27]][[83.15, 83.15, 83.15]][8.2]
MSE(-DR):[[0.0, 0.06, 0.8]][[0.26, 0.36, 1.48]][[80.36, 80.36, 80.36]][5.41]
***
MC-based ATE = -0.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.12, -4.11, -4.26]][[-6.24, -6.19, -6.75]][[0.78, 0.78, 0.78]][0.78]
std:[[0.22, 0.2, 0.15]][[0.1, 0.1, 0.05]][[0.0, 0.0, 0.0]][0.0]
MSE:[[4.13, 4.11, 4.26]][[6.24, 6.19, 6.75]][[0.78, 0.78, 0.78]][0.78]
MSE(-DR):[[0.0, -0.02, 0.13]][[2.11, 2.06, 2.62]][[-3.35, -3.35, -3.35]][-3.35]
*
==============


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:[[-6.86, -6.88, -7.0]][[-7.08, -7.13, -8.2]][[-83.84, -83.84, -83.84]][-8.9]
std:[[0.33, 0.32, 0.08]][[0.12, 0.11, 0.21]][[0.0, 0.0, 0.0]][0.1]
MSE:[[6.87, 6.89, 7.0]][[7.08, 7.13, 8.2]][[83.84, 83.84, 83.84]][8.9]
MSE(-DR):[[0.0, 0.02, 0.13]][[0.21, 0.26, 1.33]][[76.97, 76.97, 76.97]][2.03]
***
MC-based ATE = -0.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.19, -8.14, -7.68]][[-10.26, -10.17, -10.69]][[0.09, 0.09, 0.09]][0.09]
std:[[0.42, 0.41, 0.11]][[0.12, 0.11, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.2, 8.15, 7.68]][[10.26, 10.17, 10.69]][[0.09, 0.09, 0.09]][0.09]
MSE(-DR):[[0.0, -0.05, -0.52]][[2.06, 1.97, 2.49]][[-8.11, -8.11, -8.11]][-8.11]
**
==============


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:[[-10.59, -10.59, -9.51]][[-11.3, -11.33, -12.39]][[-86.09, -86.09, -86.09]][-11.15]
std:[[0.12, 0.1, 0.02]][[0.07, 0.06, 0.18]][[0.0, 0.0, 0.0]][0.1]
```

```
MSE:[[10.59, 10.59, 9.51]][[11.3, 11.33, 12.39]][[86.09, 86.09, 86.09]][11.15]
MSE(-DR):[[0.0, 0.0, -1.08]][[0.71, 0.74, 1.8]][[75.5, 75.5, 75.5]][0.56]
**
MC-based ATE = 2.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.92, -11.85, -10.18]][[-14.48, -14.38, -14.88]][[-2.16, -2.16, -2.16]][-2.16]
std:[[0.27, 0.27, 0.19]][[0.16, 0.16, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.92, 11.85, 10.18]][[14.48, 14.38, 14.88]][[2.16, 2.16, 2.16]][2.16]
MSE(-DR):[[0.0, -0.07, -1.74]][[2.56, 2.46, 2.96]][[-9.76, -9.76, -9.76]][-9.76]
**
==============


time spent until now: 20.4 mins


-------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 10]

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:74.96
O_threshold = 80
MC for this TARGET:[83.925, 0.091]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.22, 1.14, 0.63]][[3.17, 3.02, 2.52]][[-83.92, -83.92, -83.92]][-8.96]
std:[[0.28, 0.29, 0.32]][[0.24, 0.24, 0.19]][[0.0, 0.0, 0.0]][0.12]
MSE:[[1.25, 1.18, 0.71]][[3.18, 3.03, 2.53]][[83.92, 83.92, 83.92]][8.96]
MSE(-DR):[[0.0, -0.07, -0.54]][[1.93, 1.78, 1.28]][[82.67, 82.67, 82.67]][7.71]
**
==============


O_threshold = 90
MC for this TARGET:[82.087, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.5, 1.43, 0.1]][[3.62, 3.48, 2.86]][[-82.09, -82.09, -82.09]][-7.13]
std:[[0.11, 0.09, 0.2]][[0.35, 0.34, 0.34]][[0.0, 0.0, 0.0]][0.12]
MSE:[[1.5, 1.43, 0.22]][[3.64, 3.5, 2.88]][[82.09, 82.09, 82.09]][7.13]
MSE(-DR):[[0.0, -0.07, -1.28]][[2.14, 2.0, 1.38]][[80.59, 80.59, 80.59]][5.63]
**
MC-based ATE = -1.84
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.28, 0.3, -0.53]][[0.44, 0.45, 0.34]][[1.84, 1.84, 1.84]][1.84]
std:[[0.25, 0.26, 0.2]][[0.13, 0.14, 0.17]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.38, 0.4, 0.57]][[0.46, 0.47, 0.38]][[1.84, 1.84, 1.84]][1.84]
MSE(-DR):[[0.0, 0.02, 0.19]][[0.08, 0.09, 0.0]][[1.46, 1.46, 1.46]][1.46]
*
==============


O_threshold = 100
MC for this TARGET:[85.629, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.12, -1.2, -3.17]][[0.72, 0.56, -0.51]][[-85.63, -85.63, -85.63]][-10.67]
std:[[0.43, 0.45, 0.27]][[0.28, 0.27, 0.32]][[0.0, 0.0, 0.0]][0.12]
MSE:[[1.2, 1.28, 3.18]][[0.77, 0.62, 0.6]][[85.63, 85.63, 85.63]][10.67]
MSE(-DR):[[0.0, 0.08, 1.98]][[-0.43, -0.58, -0.6]][[84.43, 84.43, 84.43]][9.47]
MC-based ATE = 1.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.34, -2.34, -3.8]][[-2.46, -2.46, -3.03]][[-1.7, -1.7, -1.7]][-1.7]
std:[[0.71, 0.73, 0.47]][[0.1, 0.11, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.45, 2.45, 3.83]][[2.46, 2.46, 3.03]][[1.7, 1.7, 1.7]][1.7]
MSE(-DR):[[0.0, 0.0, 1.38]][[0.01, 0.01, 0.58]][[-0.75, -0.75, -0.75]][-0.75]
*
==============


O_threshold = 110
MC for this TARGET:[83.145, 0.082]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.83, -2.9, -3.57]][[-3.01, -3.11, -4.22]][[-83.14, -83.14, -83.14]][-8.18]
std:[[0.2, 0.19, 0.15]][[0.19, 0.17, 0.22]][[0.0, 0.0, 0.0]][0.12]
MSE:[[2.84, 2.91, 3.57]][[3.02, 3.11, 4.23]][[83.14, 83.14, 83.14]][8.18]
MSE(-DR):[[0.0, 0.07, 0.73]][[0.18, 0.27, 1.39]][[80.3, 80.3, 80.3]][5.34]
***
MC-based ATE = -0.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

```
bias:[[-4.05, -4.04, -4.2]][[-6.18, -6.13, -6.74]][[0.78, 0.78, 0.78]][0.78]
std:[[0.27, 0.27, 0.28]][[0.08, 0.11, 0.05]][[0.0, 0.0, 0.0]][0.0]
MSE:[[4.06, 4.05, 4.21]][[6.18, 6.13, 6.74]][[0.78, 0.78, 0.78]][0.78]
MSE(-DR):[[0.0, -0.01, 0.15]][[2.12, 2.07, 2.68]][[-3.28, -3.28, -3.28]][-3.28]
*
==============


O_threshold = 120
MC for this TARGET:[83.836, 0.079]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.92, -6.96, -7.04]][[-7.02, -7.07, -8.15]][[-83.84, -83.84, -83.84]][-8.88]
std:[[0.31, 0.31, 0.09]][[0.15, 0.14, 0.21]][[0.0, 0.0, 0.0]][0.12]
MSE:[[6.93, 6.97, 7.04]][[7.02, 7.07, 8.15]][[83.84, 83.84, 83.84]][8.88]
MSE(-DR):[[0.0, 0.04, 0.11]][[0.09, 0.14, 1.22]][[76.91, 76.91, 76.91]][1.95]
***
MC-based ATE = -0.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.14, -8.09, -7.67]][[-10.19, -10.09, -10.67]][[0.09, 0.09, 0.09]][0.09]
std:[[0.29, 0.29, 0.24]][[0.09, 0.1, 0.12]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.15, 8.1, 7.67]][[10.19, 10.09, 10.67]][[0.09, 0.09, 0.09]][0.09]
MSE(-DR):[[0.0, -0.05, -0.48]][[2.04, 1.94, 2.52]][[-8.06, -8.06, -8.06]][-8.06]
**
==============


O_threshold = 130
MC for this TARGET:[86.088, 0.084]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.62, -10.63, -9.53]][[-11.21, -11.27, -12.34]][[-86.09, -86.09, -86.09]][-11.13]
std:[[0.26, 0.25, 0.07]][[0.1, 0.08, 0.19]][[0.0, 0.0, 0.0]][0.12]
MSE:[[10.62, 10.63, 9.53]][[11.21, 11.27, 12.34]][[86.09, 86.09, 86.09]][11.13]
MSE(-DR):[[0.0, 0.01, -1.09]][[0.59, 0.65, 1.72]][[75.47, 75.47, 75.47]][0.51]
**
MC-based ATE = 2.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.83, -11.76, -10.16]][[-14.39, -14.29, -14.86]][[-2.16, -2.16, -2.16]][-2.16]
std:[[0.11, 0.1, 0.28]][[0.14, 0.16, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.83, 11.76, 10.16]][[14.39, 14.29, 14.86]][[2.16, 2.16, 2.16]][2.16]
MSE(-DR):[[0.0, -0.07, -1.67]][[2.56, 2.46, 3.03]][[-9.67, -9.67, -9.67]][-9.67]
**
==============


time spent until now: 30.9 mins


--------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 15]

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:74.979
O_threshold = 80
MC for this TARGET:[83.923, 0.12]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.09, 1.01, 0.59]][[3.13, 3.0, 2.42]][[-83.92, -83.92, -83.92]][-8.94]
std:[[0.38, 0.37, 0.41]][[0.27, 0.25, 0.22]][[0.0, 0.0, 0.0]][0.13]
MSE:[[1.15, 1.08, 0.72]][[3.14, 3.01, 2.43]][[83.92, 83.92, 83.92]][8.94]
MSE(-DR):[[0.0, -0.07, -0.43]][[1.99, 1.86, 1.28]][[82.77, 82.77, 82.77]][7.79]
**
==============


O_threshold = 90
MC for this TARGET:[82.085, 0.116]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.48, 1.41, 0.11]][[3.59, 3.46, 2.85]][[-82.08, -82.08, -82.08]][-7.11]
std:[[0.22, 0.22, 0.21]][[0.42, 0.4, 0.38]][[0.0, 0.0, 0.0]][0.13]
MSE:[[1.5, 1.43, 0.24]][[3.61, 3.48, 2.88]][[82.08, 82.08, 82.08]][7.11]
MSE(-DR):[[0.0, -0.07, -1.26]][[2.11, 1.98, 1.38]][[80.58, 80.58, 80.58]][5.61]
**
MC-based ATE = -1.84
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.39, 0.39, -0.48]][[0.47, 0.46, 0.43]][[1.84, 1.84, 1.84]][1.84]
std:[[0.32, 0.31, 0.27]][[0.21, 0.2, 0.16]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.5, 0.5, 0.55]][[0.51, 0.5, 0.46]][[1.84, 1.84, 1.84]][1.84]
MSE(-DR):[[0.0, 0.0, 0.05]][[0.01, 0.0, -0.04]][[1.34, 1.34, 1.34]][1.34]
```

```
*
==============


O_threshold = 100
MC for this TARGET:[85.627, 0.119]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.22, -1.27, -3.18]][[0.78, 0.61, -0.45]][[-85.63, -85.63, -85.63]][-10.65]
std:[[0.42, 0.44, 0.29]][[0.32, 0.32, 0.34]][[0.0, 0.0, 0.0]][0.13]
MSE:[[1.29, 1.34, 3.19]][[0.84, 0.69, 0.56]][[85.63, 85.63, 85.63]][10.65]
MSE(-DR):[[0.0, 0.05, 1.9]][[-0.45, -0.6, -0.73]][[84.34, 84.34, 84.34]][9.36]
MC-based ATE = 1.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.3, -2.28, -3.76]][[-2.35, -2.39, -2.87]][[-1.7, -1.7, -1.7]][-1.7]
std:[[0.8, 0.81, 0.59]][[0.14, 0.15, 0.12]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.44, 2.42, 3.81]][[2.35, 2.39, 2.87]][[1.7, 1.7, 1.7]][1.7]
MSE(-DR):[[0.0, -0.02, 1.37]][[-0.09, -0.05, 0.43]][[-0.74, -0.74, -0.74]][-0.74]
==============


O_threshold = 110
MC for this TARGET:[83.143, 0.114]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.88, -2.96, -3.57]][[-2.98, -3.07, -4.2]][[-83.14, -83.14, -83.14]][-8.16]
std:[[0.34, 0.34, 0.15]][[0.2, 0.19, 0.25]][[0.0, 0.0, 0.0]][0.13]
MSE:[[2.9, 2.98, 3.57]][[2.99, 3.08, 4.21]][[83.14, 83.14, 83.14]][8.16]
MSE(-DR):[[0.0, 0.08, 0.67]][[0.09, 0.18, 1.31]][[80.24, 80.24, 80.24]][5.26]
***
MC-based ATE = -0.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.97, -3.97, -4.16]][[-6.11, -6.07, -6.63]][[0.78, 0.78, 0.78]][0.78]
std:[[0.45, 0.44, 0.47]][[0.12, 0.13, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[4.0, 3.99, 4.19]][[6.11, 6.07, 6.63]][[0.78, 0.78, 0.78]][0.78]
MSE(-DR):[[0.0, -0.01, 0.19]][[2.11, 2.07, 2.63]][[-3.22, -3.22, -3.22]][-3.22]
*
==============


O_threshold = 120
MC for this TARGET:[83.834, 0.11]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.01, -7.03, -7.13]][[-6.96, -7.01, -8.11]][[-83.83, -83.83, -83.83]][-8.86]
std:[[0.33, 0.32, 0.06]][[0.19, 0.17, 0.24]][[0.0, 0.0, 0.0]][0.13]
MSE:[[7.02, 7.04, 7.13]][[6.96, 7.01, 8.11]][[83.83, 83.83, 83.83]][8.86]
MSE(-DR):[[0.0, 0.02, 0.11]][[-0.06, -0.01, 1.09]][[76.81, 76.81, 76.81]][1.84]
MC-based ATE = -0.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.1, -8.04, -7.72]][[-10.09, -10.01, -10.53]][[0.09, 0.09, 0.09]][0.09]
std:[[0.16, 0.18, 0.35]][[0.08, 0.1, 0.05]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.1, 8.04, 7.73]][[10.09, 10.01, 10.53]][[0.09, 0.09, 0.09]][0.09]
MSE(-DR):[[0.0, -0.06, -0.37]][[1.99, 1.91, 2.43]][[-8.01, -8.01, -8.01]][-8.01]
**
==============


O_threshold = 130
MC for this TARGET:[86.086, 0.115]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.66, -10.66, -9.6]][[-11.18, -11.21, -12.31]][[-86.09, -86.09, -86.09]][-11.11]
std:[[0.46, 0.45, 0.13]][[0.11, 0.11, 0.19]][[0.0, 0.0, 0.0]][0.13]
MSE:[[10.67, 10.67, 9.6]][[11.18, 11.21, 12.31]][[86.09, 86.09, 86.09]][11.11]
MSE(-DR):[[0.0, 0.0, -1.07]][[0.51, 0.54, 1.64]][[75.42, 75.42, 75.42]][0.44]
**
MC-based ATE = 2.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.75, -11.67, -10.19]][[-14.3, -14.21, -14.74]][[-2.16, -2.16, -2.16]][-2.16]
std:[[0.08, 0.09, 0.3]][[0.16, 0.17, 0.05]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.75, 11.67, 10.19]][[14.3, 14.21, 14.74]][[2.16, 2.16, 2.16]][2.16]
MSE(-DR):[[0.0, -0.08, -1.56]][[2.55, 2.46, 2.99]][[-9.59, -9.59, -9.59]][-9.59]
**
==============


time spent until now: 41.0 mins


--------------------------------------
[pattern_seed, T, sd_R] = [1, 672, 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:66.725
O_threshold = 80
MC for this TARGET:[73.15, 0.051]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.54, 1.47, 0.76]][[3.56, 3.42, 2.86]][[-73.15, -73.15, -73.15]][-6.42]
std:[[0.38, 0.37, 0.35]][[0.08, 0.1, 0.09]][[0.0, 0.0, 0.0]][0.14]
MSE:[[1.59, 1.52, 0.84]][[3.56, 3.42, 2.86]][[73.15, 73.15, 73.15]][6.42]
MSE(-DR):[[0.0, -0.07, -0.75]][[1.97, 1.83, 1.27]][[71.56, 71.56, 71.56]][4.83]
**
==============


O_threshold = 90
MC for this TARGET:[73.515, 0.047]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.32, -0.38, -1.62]][[0.92, 0.79, 0.07]][[-73.52, -73.52, -73.52]][-6.79]
std:[[0.17, 0.17, 0.27]][[0.02, 0.04, 0.05]][[0.0, 0.0, 0.0]][0.14]
MSE:[[0.36, 0.42, 1.64]][[0.92, 0.79, 0.09]][[73.52, 73.52, 73.52]][6.79]
MSE(-DR):[[0.0, 0.06, 1.28]][[0.56, 0.43, -0.27]][[73.16, 73.16, 73.16]][6.43]
***
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, -2.38]][[-2.64, -2.63, -2.8]][[-0.36, -0.36, -0.36]][-0.36]
std:[[0.27, 0.27, 0.14]][[0.06, 0.06, 0.04]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.88, 1.87, 2.38]][[2.64, 2.63, 2.8]][[0.36, 0.36, 0.36]][0.36]
MSE(-DR):[[0.0, -0.01, 0.5]][[0.76, 0.75, 0.92]][[-1.52, -1.52, -1.52]][-1.52]
*
==============


O_threshold = 100
MC for this TARGET:[77.167, 0.048]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.95, -5.01, -5.47]][[-4.23, -4.36, -5.36]][[-77.17, -77.17, -77.17]][-10.44]
std:[[0.2, 0.21, 0.16]][[0.09, 0.09, 0.17]][[0.0, 0.0, 0.0]][0.14]
MSE:[[4.95, 5.01, 5.47]][[4.23, 4.36, 5.36]][[77.17, 77.17, 77.17]][10.44]
MSE(-DR):[[0.0, 0.06, 0.52]][[-0.72, -0.59, 0.41]][[72.22, 72.22, 72.22]][5.49]
MC-based ATE = 4.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.49, -6.48, -6.22]][[-7.79, -7.78, -8.22]][[-4.02, -4.02, -4.02]][-4.02]
std:[[0.36, 0.36, 0.2]][[0.05, 0.05, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.5, 6.49, 6.22]][[7.79, 7.78, 8.22]][[4.02, 4.02, 4.02]][4.02]
MSE(-DR):[[0.0, -0.01, -0.28]][[1.29, 1.28, 1.72]][[-2.48, -2.48, -2.48]][-2.48]
**
==============


O_threshold = 110
MC for this TARGET:[80.267, 0.047]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.45, -7.53, -7.47]][[-7.89, -8.03, -9.23]][[-80.27, -80.27, -80.27]][-13.54]
std:[[0.16, 0.16, 0.26]][[0.13, 0.13, 0.18]][[0.0, 0.0, 0.0]][0.14]
MSE:[[7.45, 7.53, 7.47]][[7.89, 8.03, 9.23]][[80.27, 80.27, 80.27]][13.54]
MSE(-DR):[[0.0, 0.08, 0.02]][[0.44, 0.58, 1.78]][[72.82, 72.82, 72.82]][6.09]
***
MC-based ATE = 7.12
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.99, -9.0, -8.22]][[-11.45, -11.46, -12.09]][[-7.12, -7.12, -7.12]][-7.12]
std:[[0.36, 0.36, 0.34]][[0.1, 0.09, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[9.0, 9.01, 8.23]][[11.45, 11.46, 12.09]][[7.12, 7.12, 7.12]][7.12]
MSE(-DR):[[0.0, 0.01, -0.77]][[2.45, 2.46, 3.09]][[-1.88, -1.88, -1.88]][-1.88]
**
==============


O_threshold = 120
MC for this TARGET:[78.019, 0.048]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.67, -9.68, -9.5]][[-11.21, -11.26, -12.31]][[-78.02, -78.02, -78.02]][-11.29]
std:[[0.32, 0.33, 0.14]][[0.08, 0.07, 0.17]][[0.0, 0.0, 0.0]][0.14]
MSE:[[9.68, 9.69, 9.5]][[11.21, 11.26, 12.31]][[78.02, 78.02, 78.02]][11.29]
MSE(-DR):[[0.0, 0.01, -0.18]][[1.53, 1.58, 2.63]][[68.34, 68.34, 68.34]][1.61]
**
MC-based ATE = 4.87
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.2, -11.15, -10.26]][[-14.77, -14.68, -15.17]][[-4.87, -4.87, -4.87]][-4.87]
std:[[0.06, 0.07, 0.34]][[0.11, 0.11, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.2, 11.15, 10.27]][[14.77, 14.68, 15.17]][[4.87, 4.87, 4.87]][4.87]
MSE(-DR):[[0.0, -0.05, -0.93]][[3.57, 3.48, 3.97]][[-6.33, -6.33, -6.33]][-6.33]
**
==============


O_threshold = 130
MC for this TARGET:[75.73, 0.05]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

```
bias:[[-9.59, -9.59, -9.72]][[-11.58, -11.57, -12.51]][[-75.73, -75.73, -75.73]][-9.0]
std:[[0.37, 0.36, 0.16]][[0.1, 0.1, 0.14]][[0.0, 0.0, 0.0]][0.14]
MSE:[[9.6, 9.6, 9.72]][[11.58, 11.57, 12.51]][[75.73, 75.73, 75.73]][9.0]
MSE(-DR):[[0.0, 0.0, 0.12]][[1.98, 1.97, 2.91]][[66.13, 66.13, 66.13]][-0.6]
***
MC-based ATE = 2.58
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.13, -11.06, -10.48]][[-15.14, -14.99, -15.37]][[-2.58, -2.58, -2.58]][-2.58]
std:[[0.05, 0.07, 0.36]][[0.15, 0.16, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.13, 11.06, 10.49]][[15.14, 14.99, 15.37]][[2.58, 2.58, 2.58]][2.58]
MSE(-DR):[[0.0, -0.07, -0.64]][[4.01, 3.86, 4.24]][[-8.55, -8.55, -8.55]][-8.55]
**
==============


time spent until now: 51.4 mins


______________________________________
[pattern_seed, T, sd_R] = [1, 672, 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: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:[[1.55, 1.48, 0.8]][[3.56, 3.43, 2.86]][[-73.15, -73.15, -73.15]][-6.41]
std:[[0.51, 0.49, 0.4]][[0.11, 0.13, 0.1]][[0.0, 0.0, 0.0]][0.15]
MSE:[[1.63, 1.56, 0.89]][[3.56, 3.43, 2.86]][[73.15, 73.15, 73.15]][6.41]
MSE(-DR):[[0.0, -0.07, -0.74]][[1.93, 1.8, 1.23]][[71.52, 71.52, 71.52]][4.78]
**
==============


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:[[-0.33, -0.38, -1.57]][[0.98, 0.84, 0.11]][[-73.51, -73.51, -73.51]][-6.77]
std:[[0.34, 0.34, 0.3]][[0.05, 0.05, 0.07]][[0.0, 0.0, 0.0]][0.15]
MSE:[[0.47, 0.51, 1.6]][[0.98, 0.84, 0.13]][[73.51, 73.51, 73.51]][6.77]
MSE(-DR):[[0.0, 0.04, 1.13]][[0.51, 0.37, -0.34]][[73.04, 73.04, 73.04]][6.3]
***
MC-based ATE = 0.36
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.88, -1.85, -2.37]][[-2.59, -2.6, -2.75]][[-0.36, -0.36, -0.36]][-0.36]
std:[[0.2, 0.19, 0.18]][[0.06, 0.08, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.89, 1.86, 2.38]][[2.59, 2.6, 2.75]][[0.36, 0.36, 0.36]][0.36]
MSE(-DR):[[0.0, -0.03, 0.49]][[0.7, 0.71, 0.86]][[-1.53, -1.53, -1.53]][-1.53]
*
==============


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:[[-4.99, -5.05, -5.44]][[-4.17, -4.28, -5.33]][[-77.17, -77.17, -77.17]][-10.42]
std:[[0.32, 0.34, 0.22]][[0.12, 0.12, 0.17]][[0.0, 0.0, 0.0]][0.15]
MSE:[[5.0, 5.06, 5.44]][[4.17, 4.28, 5.33]][[77.17, 77.17, 77.17]][10.42]
MSE(-DR):[[0.0, 0.06, 0.44]][[-0.83, -0.72, 0.33]][[72.17, 72.17, 72.17]][5.42]
MC-based ATE = 4.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.54, -6.53, -6.24]][[-7.73, -7.72, -8.19]][[-4.02, -4.02, -4.02]][-4.02]
std:[[0.52, 0.51, 0.19]][[0.1, 0.11, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.56, 6.55, 6.24]][[7.73, 7.72, 8.19]][[4.02, 4.02, 4.02]][4.02]
MSE(-DR):[[0.0, -0.01, -0.32]][[1.17, 1.16, 1.63]][[-2.54, -2.54, -2.54]][-2.54]
**
==============


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:[[-7.49, -7.56, -7.56]][[-7.85, -7.97, -9.23]][[-80.27, -80.27, -80.27]][-13.52]
std:[[0.28, 0.31, 0.22]][[0.16, 0.15, 0.22]][[0.0, 0.0, 0.0]][0.15]
MSE:[[7.5, 7.57, 7.56]][[7.85, 7.97, 9.23]][[80.27, 80.27, 80.27]][13.52]
MSE(-DR):[[0.0, 0.07, 0.06]][[0.35, 0.47, 1.73]][[72.77, 72.77, 72.77]][6.02]
***
```

```
MC-based ATE = 7.12
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.04, -9.04, -8.36]][[-11.41, -11.4, -12.09]][[-7.12, -7.12, -7.12]][-7.12]
std:[[0.56, 0.56, 0.44]][[0.13, 0.12, 0.16]][[0.0, 0.0, 0.0]][0.0]
MSE:[[9.06, 9.06, 8.37]][[11.41, 11.4, 12.09]][[7.12, 7.12, 7.12]][7.12]
MSE(-DR):[[0.0, 0.0, -0.69]][[2.35, 2.34, 3.03]][[-1.94, -1.94, -1.94]][-1.94]
**
==============


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, -9.57]][[-11.18, -11.21, -12.28]][[-78.02, -78.02, -78.02]][-11.27]
std:[[0.2, 0.23, 0.21]][[0.08, 0.08, 0.16]][[0.0, 0.0, 0.0]][0.15]
MSE:[[9.6, 9.6, 9.57]][[11.18, 11.21, 12.28]][[78.02, 78.02, 78.02]][11.27]
MSE(-DR):[[0.0, 0.0, -0.03]][[1.58, 1.61, 2.68]][[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:[[-11.15, -11.08, -10.37]][[-14.74, -14.64, -15.13]][[-4.87, -4.87, -4.87]][-4.87]
std:[[0.38, 0.35, 0.47]][[0.13, 0.15, 0.12]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.16, 11.09, 10.38]][[14.74, 14.64, 15.13]][[4.87, 4.87, 4.87]][4.87]
MSE(-DR):[[0.0, -0.07, -0.78]][[3.58, 3.48, 3.97]][[-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:[[-9.44, -9.44, -9.71]][[-11.53, -11.51, -12.49]][[-75.73, -75.73, -75.73]][-8.99]
std:[[0.22, 0.24, 0.1]][[0.11, 0.12, 0.17]][[0.0, 0.0, 0.0]][0.15]
MSE:[[9.44, 9.44, 9.71]][[11.53, 11.51, 12.49]][[75.73, 75.73, 75.73]][8.99]
MSE(-DR):[[0.0, 0.0, 0.27]][[2.09, 2.07, 3.05]][[66.29, 66.29, 66.29]][-0.45]
***
MC-based ATE = 2.58
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.98, -10.91, -10.51]][[-15.09, -14.94, -15.35]][[-2.58, -2.58, -2.58]][-2.58]
std:[[0.32, 0.3, 0.43]][[0.19, 0.21, 0.16]][[0.0, 0.0, 0.0]][0.0]
MSE:[[10.98, 10.91, 10.52]][[15.09, 14.94, 15.35]][[2.58, 2.58, 2.58]][2.58]
MSE(-DR):[[0.0, -0.07, -0.46]][[4.11, 3.96, 4.37]][[-8.4, -8.4, -8.4]][-8.4]
**
==============


time spent until now: 61.8 mins


--------------------------------------
[pattern_seed, T, sd_R] = [1, 672, 10]

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:66.761
O_threshold = 80
MC for this TARGET:[73.147, 0.087]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.55, 1.49, 0.84]][[3.59, 3.45, 2.88]][[-73.15, -73.15, -73.15]][-6.39]
std:[[0.61, 0.63, 0.45]][[0.16, 0.16, 0.19]][[0.0, 0.0, 0.0]][0.17]
MSE:[[1.67, 1.62, 0.95]][[3.59, 3.45, 2.89]][[73.15, 73.15, 73.15]][6.39]
MSE(-DR):[[0.0, -0.05, -0.72]][[1.92, 1.78, 1.22]][[71.48, 71.48, 71.48]][4.72]
**
==============


O_threshold = 90
MC for this TARGET:[73.511, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.33, -0.37, -1.55]][[1.02, 0.89, 0.12]][[-73.51, -73.51, -73.51]][-6.75]
std:[[0.59, 0.57, 0.41]][[0.06, 0.07, 0.12]][[0.0, 0.0, 0.0]][0.17]
MSE:[[0.68, 0.68, 1.6]][[1.02, 0.89, 0.17]][[73.51, 73.51, 73.51]][6.75]
MSE(-DR):[[0.0, 0.0, 0.92]][[0.34, 0.21, -0.51]][[72.83, 72.83, 72.83]][6.07]
***
MC-based ATE = 0.36
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.89, -1.86, -2.39]][[-2.57, -2.56, -2.76]][[-0.36, -0.36, -0.36]][-0.36]
std:[[0.14, 0.14, 0.06]][[0.1, 0.1, 0.08]][[0.0, 0.0, 0.0]][0.0]
```

```
MSE:[[1.9, 1.87, 2.39]][[2.57, 2.56, 2.76]][[0.36, 0.36, 0.36]][0.36]
MSE(-DR):[[0.0, -0.03, 0.49]][[0.67, 0.66, 0.86]][[-1.54, -1.54, -1.54]][-1.54]
*
==============


O_threshold = 100
MC for this TARGET:[77.163, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.03, -5.09, -5.55]][[-4.08, -4.2, -5.29]][[-77.16, -77.16, -77.16]][-10.4]
std:[[0.49, 0.49, 0.35]][[0.16, 0.17, 0.18]][[0.0, 0.0, 0.0]][0.17]
MSE:[[5.05, 5.11, 5.56]][[4.08, 4.2, 5.29]][[77.16, 77.16, 77.16]][10.4]
MSE(-DR):[[0.0, 0.06, 0.51]][[-0.97, -0.85, 0.24]][[72.11, 72.11, 72.11]][5.35]
MC-based ATE = 4.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.59, -6.58, -6.39]][[-7.67, -7.65, -8.17]][[-4.02, -4.02, -4.02]][-4.02]
std:[[0.69, 0.71, 0.11]][[0.16, 0.18, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.63, 6.62, 6.39]][[7.67, 7.65, 8.17]][[4.02, 4.02, 4.02]][4.02]
MSE(-DR):[[0.0, -0.01, -0.24]][[1.04, 1.02, 1.54]][[-2.61, -2.61, -2.61]][-2.61]
**
==============


O_threshold = 110
MC for this TARGET:[80.264, 0.083]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.53, -7.6, -7.63]][[-7.77, -7.9, -9.18]][[-80.26, -80.26, -80.26]][-13.5]
std:[[0.46, 0.47, 0.28]][[0.18, 0.17, 0.23]][[0.0, 0.0, 0.0]][0.17]
MSE:[[7.54, 7.61, 7.64]][[7.77, 7.9, 9.18]][[80.26, 80.26, 80.26]][13.5]
MSE(-DR):[[0.0, 0.07, 0.1]][[0.23, 0.36, 1.64]][[72.72, 72.72, 72.72]][5.96]
***
MC-based ATE = 7.12
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.08, -9.09, -8.47]][[-11.36, -11.35, -12.06]][[-7.12, -7.12, -7.12]][-7.12]
std:[[0.79, 0.8, 0.43]][[0.16, 0.16, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[9.11, 9.13, 8.48]][[11.36, 11.35, 12.06]][[7.12, 7.12, 7.12]][7.12]
MSE(-DR):[[0.0, 0.02, -0.63]][[2.25, 2.24, 2.95]][[-1.99, -1.99, -1.99]][-1.99]
**
==============


O_threshold = 120
MC for this TARGET:[78.015, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.49, -9.51, -9.63]][[-11.13, -11.15, -12.25]][[-78.02, -78.02, -78.02]][-11.25]
std:[[0.21, 0.21, 0.26]][[0.08, 0.1, 0.15]][[0.0, 0.0, 0.0]][0.17]
MSE:[[9.49, 9.51, 9.63]][[11.13, 11.15, 12.25]][[78.02, 78.02, 78.02]][11.25]
MSE(-DR):[[0.0, 0.02, 0.14]][[1.64, 1.66, 2.76]][[68.53, 68.53, 68.53]][1.76]
***
MC-based ATE = 4.87
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.04, -11.0, -10.47]][[-14.71, -14.6, -15.13]][[-4.87, -4.87, -4.87]][-4.87]
std:[[0.67, 0.66, 0.53]][[0.2, 0.2, 0.2]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.06, 11.02, 10.48]][[14.71, 14.6, 15.13]][[4.87, 4.87, 4.87]][4.87]
MSE(-DR):[[0.0, -0.04, -0.58]][[3.65, 3.54, 4.07]][[-6.19, -6.19, -6.19]][-6.19]
**
==============


O_threshold = 130
MC for this TARGET:[75.726, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.28, -9.27, -9.72]][[-11.46, -11.44, -12.47]][[-75.73, -75.73, -75.73]][-8.97]
std:[[0.11, 0.13, 0.17]][[0.14, 0.15, 0.21]][[0.0, 0.0, 0.0]][0.17]
MSE:[[9.28, 9.27, 9.72]][[11.46, 11.44, 12.47]][[75.73, 75.73, 75.73]][8.97]
MSE(-DR):[[0.0, -0.01, 0.44]][[2.18, 2.16, 3.19]][[66.45, 66.45, 66.45]][-0.31]
***
MC-based ATE = 2.58
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.84, -10.76, -10.56]][[-15.05, -14.88, -15.36]][[-2.58, -2.58, -2.58]][-2.58]
std:[[0.58, 0.58, 0.44]][[0.27, 0.27, 0.28]][[0.0, 0.0, 0.0]][0.0]
MSE:[[10.86, 10.78, 10.57]][[15.05, 14.88, 15.36]][[2.58, 2.58, 2.58]][2.58]
MSE(-DR):[[0.0, -0.08, -0.29]][[4.19, 4.02, 4.5]][[-8.28, -8.28, -8.28]][-8.28]
**
==============


time spent until now: 72.4 mins


--------------------------------------
[pattern_seed, T, sd_R] = [1, 672, 15]

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:66.779
O_threshold = 80
MC for this TARGET:[73.145, 0.118]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.57, 1.5, 0.89]][[3.6, 3.46, 2.85]][[-73.14, -73.14, -73.14]][-6.37]
std:[[0.76, 0.76, 0.48]][[0.2, 0.2, 0.24]][[0.0, 0.0, 0.0]][0.18]
MSE:[[1.74, 1.68, 1.01]][[3.61, 3.47, 2.86]][[73.14, 73.14, 73.14]][6.37]
MSE(-DR):[[0.0, -0.06, -0.73]][[1.87, 1.73, 1.12]][[71.4, 71.4, 71.4]][4.63]
**
==============


O_threshold = 90
MC for this TARGET:[73.51, 0.118]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.31, -0.37, -1.47]][[1.06, 0.94, 0.15]][[-73.51, -73.51, -73.51]][-6.73]
std:[[0.81, 0.8, 0.46]][[0.09, 0.09, 0.16]][[0.0, 0.0, 0.0]][0.18]
MSE:[[0.87, 0.88, 1.54]][[1.06, 0.94, 0.22]][[73.51, 73.51, 73.51]][6.73]
MSE(-DR):[[0.0, 0.01, 0.67]][[0.19, 0.07, -0.65]][[72.64, 72.64, 72.64]][5.86]
***
MC-based ATE = 0.37
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.88, -1.86, -2.36]][[-2.54, -2.52, -2.69]][[-0.37, -0.37, -0.37]][-0.37]
std:[[0.16, 0.16, 0.05]][[0.12, 0.12, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.89, 1.87, 2.36]][[2.54, 2.52, 2.69]][[0.37, 0.37, 0.37]][0.37]
MSE(-DR):[[0.0, -0.02, 0.47]][[0.65, 0.63, 0.8]][[-1.52, -1.52, -1.52]][-1.52]
*
==============


O_threshold = 100
MC for this TARGET:[77.161, 0.117]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.08, -5.13, -5.49]][[-4.0, -4.12, -5.24]][[-77.16, -77.16, -77.16]][-10.38]
std:[[0.64, 0.63, 0.46]][[0.21, 0.22, 0.23]][[0.0, 0.0, 0.0]][0.18]
MSE:[[5.12, 5.17, 5.51]][[4.01, 4.13, 5.25]][[77.16, 77.16, 77.16]][10.38]
MSE(-DR):[[0.0, 0.05, 0.39]][[-1.11, -0.99, 0.13]][[72.04, 72.04, 72.04]][5.26]
MC-based ATE = 4.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.65, -6.63, -6.39]][[-7.6, -7.58, -8.09]][[-4.02, -4.02, -4.02]][-4.02]
std:[[0.91, 0.9, 0.22]][[0.23, 0.24, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.71, 6.69, 6.39]][[7.6, 7.58, 8.09]][[4.02, 4.02, 4.02]][4.02]
MSE(-DR):[[0.0, -0.02, -0.32]][[0.89, 0.87, 1.38]][[-2.69, -2.69, -2.69]][-2.69]
**
==============


O_threshold = 110
MC for this TARGET:[80.262, 0.115]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.55, -7.64, -7.68]][[-7.71, -7.83, -9.2]][[-80.26, -80.26, -80.26]][-13.48]
std:[[0.62, 0.64, 0.32]][[0.18, 0.19, 0.22]][[0.0, 0.0, 0.0]][0.18]
MSE:[[7.58, 7.67, 7.69]][[7.71, 7.83, 9.2]][[80.26, 80.26, 80.26]][13.48]
MSE(-DR):[[0.0, 0.09, 0.11]][[0.13, 0.25, 1.62]][[72.68, 72.68, 72.68]][5.9]
***
MC-based ATE = 7.12
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.12, -9.13, -8.57]][[-11.31, -11.29, -12.05]][[-7.12, -7.12, -7.12]][-7.12]
std:[[1.05, 1.04, 0.48]][[0.18, 0.2, 0.12]][[0.0, 0.0, 0.0]][0.0]
MSE:[[9.18, 9.19, 8.58]][[11.31, 11.29, 12.05]][[7.12, 7.12, 7.12]][7.12]
MSE(-DR):[[0.0, 0.01, -0.6]][[2.13, 2.11, 2.87]][[-2.06, -2.06, -2.06]][-2.06]
**
==============


O_threshold = 120
MC for this TARGET:[78.013, 0.119]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.42, -9.43, -9.63]][[-11.08, -11.1, -12.23]][[-78.01, -78.01, -78.01]][-11.23]
std:[[0.3, 0.31, 0.31]][[0.1, 0.11, 0.19]][[0.0, 0.0, 0.0]][0.18]
MSE:[[9.42, 9.44, 9.63]][[11.08, 11.1, 12.23]][[78.01, 78.01, 78.01]][11.23]
MSE(-DR):[[0.0, 0.02, 0.21]][[1.66, 1.68, 2.81]][[68.59, 68.59, 68.59]][1.81]
***
MC-based ATE = 4.87
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.99, -10.92, -10.53]][[-14.67, -14.56, -15.08]][[-4.87, -4.87, -4.87]][-4.87]
std:[[0.99, 0.98, 0.63]][[0.24, 0.25, 0.23]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.03, 10.96, 10.55]][[14.67, 14.56, 15.08]][[4.87, 4.87, 4.87]][4.87]
MSE(-DR):[[0.0, -0.07, -0.48]][[3.64, 3.53, 4.05]][[-6.16, -6.16, -6.16]][-6.16]
**
==============
O_threshold = 130
MC for this TARGET:[75.724, 0.12]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.1, -9.11, -9.7]][[-11.39, -11.37, -12.4]][[-75.72, -75.72, -75.72]][-8.94]
std:[[0.13, 0.13, 0.15]][[0.16, 0.17, 0.23]][[0.0, 0.0, 0.0]][0.18]
MSE:[[9.1, 9.11, 9.7]][[11.39, 11.37, 12.4]][[75.72, 75.72, 75.72]][8.94]
MSE(-DR):[[0.0, 0.01, 0.6]][[2.29, 2.27, 3.3]][[66.62, 66.62, 66.62]][-0.16]
***
MC-based ATE = 2.58
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.67, -10.6, -10.59]][[-14.98, -14.82, -15.24]][[-2.58, -2.58, -2.58]][-2.58]
std:[[0.88, 0.86, 0.49]][[0.31, 0.33, 0.3]][[0.0, 0.0, 0.0]][0.0]
MSE:[[10.71, 10.63, 10.6]][[14.98, 14.82, 15.24]][[2.58, 2.58, 2.58]][2.58]
MSE(-DR):[[0.0, -0.08, -0.11]][[4.27, 4.11, 4.53]][[-8.13, -8.13, -8.13]][-8.13]
**
==============


time spent until now: 83.0 mins


--------------------------------------
[pattern_seed, T, sd_R] = [2, 672, 0.5]

max(u_O) =  157.3
O_threshold = 80
means of Order:

91.5 98.4 64.9 138.1 69.5

84.1 110.0 77.6 80.5 82.9

111.1 157.3 100.3 79.6 110.8

88.3 99.1 125.8 85.7 99.7

83.5 96.4 104.7 81.6 93.0

target policy:

1 1 0 1 0

1 1 0 1 1

1 1 1 0 1

1 1 1 1 1

1 1 1 1 1

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

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  14
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
O_threshold = 120
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 130
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  2
1 -th target;
```