```
Last login: Tue Mar 31 23:14:09 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-228-10-241.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 disabled due to load higher than 16.0

 * 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: Wed Apr  1 03:14:14 2020 from 107.13.161.147
ubuntu@ip-172-31-14-85:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
00:22, 04/01; num of cores:16

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A,  simple, M_in_R, u_O_u_D, mean_reversion, pisO] = [None, 5, 5, N
one, 0.2, 1, 1, False, True, 0, False, False]


---------------------------------------
[pattern_seed, sd_OD] = [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
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:79.148
O_threshold = 80
MC for this TARGET:[88.794, 0.146]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.98, -2.17, -0.38]][[0.22, 0.08, -0.28]][[-88.79, -88.79, -88.79]][-9.65]
std:[[0.23, 0.23, 0.44]][[0.31, 0.3, 0.23]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.99, 2.18, 0.58]][[0.38, 0.31, 0.36]][[88.79, 88.79, 88.79]][9.65]
MSE(-DR):[[0.0, 0.19, -1.41]][[-1.61, -1.68, -1.63]][[86.8, 86.8, 86.8]][7.66]
==============


O_threshold = 90
MC for this TARGET:[87.319, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.04, -1.25, -0.46]][[1.34, 1.14, 0.67]][[-87.32, -87.32, -87.32]][-8.17]
std:[[0.29, 0.29, 0.35]][[0.52, 0.5, 0.41]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.08, 1.28, 0.58]][[1.44, 1.24, 0.79]][[87.32, 87.32, 87.32]][8.17]
MSE(-DR):[[0.0, 0.2, -0.5]][[0.36, 0.16, -0.29]][[86.24, 86.24, 86.24]][7.09]
**
MC-based ATE = -1.47
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.94, 0.92, -0.09]][[1.11, 1.06, 0.95]][[1.47, 1.47, 1.47]][1.47]
std:[[0.06, 0.06, 0.08]][[0.2, 0.2, 0.19]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.94, 0.92, 0.12]][[1.13, 1.08, 0.97]][[1.47, 1.47, 1.47]][1.47]
MSE(-DR):[[0.0, -0.02, -0.82]][[0.19, 0.14, 0.03]][[0.53, 0.53, 0.53]][0.53]
**
==============
```

```
O_threshold = 100
MC for this TARGET:[91.564, 0.144]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.98, -3.24, -4.38]][[-1.45, -1.72, -2.59]][[-91.56, -91.56, -91.56]][-12.42]
std:[[0.06, 0.09, 0.13]][[0.47, 0.47, 0.41]][[0.0, 0.0, 0.0]][0.06]
MSE:[[2.98, 3.24, 4.38]][[1.52, 1.78, 2.62]][[91.56, 91.56, 91.56]][12.42]
MSE(-DR):[[0.0, 0.26, 1.4]][[-1.46, -1.2, -0.36]][[88.58, 88.58, 88.58]][9.44]
MC-based ATE = 2.77
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.0, -1.07, -4.0]][[-1.68, -1.8, -2.31]][[-2.77, -2.77, -2.77]][-2.77]
std:[[0.17, 0.13, 0.31]][[0.15, 0.17, 0.18]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.01, 1.08, 4.01]][[1.69, 1.81, 2.32]][[2.77, 2.77, 2.77]][2.77]
MSE(-DR):[[0.0, 0.07, 3.0]][[0.68, 0.8, 1.31]][[1.76, 1.76, 1.76]][1.76]
*
==============


O_threshold = 110
MC for this TARGET:[88.696, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.27, -3.5, -3.57]][[-2.9, -3.1, -4.25]][[-88.7, -88.7, -88.7]][-9.55]
std:[[0.04, 0.07, 0.03]][[0.19, 0.17, 0.22]][[0.0, 0.0, 0.0]][0.06]
MSE:[[3.27, 3.5, 3.57]][[2.91, 3.1, 4.26]][[88.7, 88.7, 88.7]][9.55]
MSE(-DR):[[0.0, 0.23, 0.3]][[-0.36, -0.17, 0.99]][[85.43, 85.43, 85.43]][6.28]
MC-based ATE = -0.1
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.3, -1.33, -3.19]][[-3.12, -3.18, -3.97]][[0.1, 0.1, 0.1]][0.1]
std:[[0.26, 0.29, 0.46]][[0.13, 0.12, 0.01]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.33, 1.36, 3.22]][[3.12, 3.18, 3.97]][[0.1, 0.1, 0.1]][0.1]
MSE(-DR):[[0.0, 0.03, 1.89]][[1.79, 1.85, 2.64]][[-1.23, -1.23, -1.23]][-1.23]
*
==============


O_threshold = 120
MC for this TARGET:[90.808, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.68, -8.86, -8.03]][[-7.68, -7.87, -9.02]][[-90.81, -90.81, -90.81]][-11.66]
std:[[0.74, 0.76, 0.46]][[0.31, 0.33, 0.36]][[0.0, 0.0, 0.0]][0.06]
MSE:[[8.71, 8.89, 8.04]][[7.69, 7.88, 9.03]][[90.81, 90.81, 90.81]][11.66]
MSE(-DR):[[0.0, 0.18, -0.67]][[-1.02, -0.83, 0.32]][[82.1, 82.1, 82.1]][2.95]
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.7, -6.69, -7.65]][[-7.9, -7.95, -8.74]][[-2.01, -2.01, -2.01]][-2.01]
std:[[0.97, 0.98, 0.89]][[0.0, 0.03, 0.14]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.77, 6.76, 7.7]][[7.9, 7.95, 8.74]][[2.01, 2.01, 2.01]][2.01]
MSE(-DR):[[0.0, -0.01, 0.93]][[1.13, 1.18, 1.97]][[-4.76, -4.76, -4.76]][-4.76]
*
==============


[[ 1.99  2.18  0.58  0.38  0.31  0.36 88.79 88.79 88.79  9.65]
 [ 1.08  1.28  0.58  1.44  1.24  0.79 87.32 87.32 87.32  8.17]
 [ 2.98  3.24  4.38  1.52  1.78  2.62 91.56 91.56 91.56 12.42]
 [ 3.27  3.5   3.57  2.91  3.1   4.26 88.7  88.7  88.7   9.55]
 [ 8.71  8.89  8.04  7.69  7.88  9.03 90.81 90.81 90.81 11.66]]
time spent until now: 6.0 mins


--------------------------------------
[pattern_seed, sd_OD] = [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
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

--------------------------------------
Value of Behaviour policy:79.148
O_threshold = 80
MC for this TARGET:[88.794, 0.146]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.96, -2.17, -0.36]][[0.25, 0.08, -0.24]][[-88.79, -88.79, -88.79]][-9.65]
std:[[0.22, 0.23, 0.4]][[0.31, 0.3, 0.22]][[0.0, 0.0, 0.0]][0.06]
```

```
MSE:[[1.97, 2.18, 0.54]][[0.4, 0.31, 0.33]][[88.79, 88.79, 88.79]][9.65]
MSE(-DR):[[0.0, 0.21, -1.43]][[-1.57, -1.66, -1.64]][[86.82, 86.82, 86.82]][7.68]
==============


O_threshold = 90
MC for this TARGET:[87.319, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.03, -1.25, -0.51]][[1.3, 1.14, 0.61]][[-87.32, -87.32, -87.32]][-8.17]
std:[[0.3, 0.29, 0.35]][[0.49, 0.5, 0.38]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.07, 1.28, 0.62]][[1.39, 1.24, 0.72]][[87.32, 87.32, 87.32]][8.17]
MSE(-DR):[[0.0, 0.21, -0.45]][[0.32, 0.17, -0.35]][[86.25, 86.25, 86.25]][7.1]
**
MC-based ATE = -1.47
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.93, 0.92, -0.15]][[1.05, 1.06, 0.85]][[1.47, 1.47, 1.47]][1.47]
std:[[0.08, 0.06, 0.06]][[0.18, 0.2, 0.16]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.93, 0.92, 0.16]][[1.07, 1.08, 0.86]][[1.47, 1.47, 1.47]][1.47]
MSE(-DR):[[0.0, -0.01, -0.77]][[0.14, 0.15, -0.07]][[0.54, 0.54, 0.54]][0.54]
**
==============


O_threshold = 100
MC for this TARGET:[91.564, 0.144]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.96, -3.24, -4.36]][[-1.45, -1.72, -2.58]][[-91.56, -91.56, -91.56]][-12.42]
std:[[0.07, 0.09, 0.16]][[0.43, 0.47, 0.35]][[0.0, 0.0, 0.0]][0.06]
MSE:[[2.96, 3.24, 4.36]][[1.51, 1.78, 2.6]][[91.56, 91.56, 91.56]][12.42]
MSE(-DR):[[0.0, 0.28, 1.4]][[-1.45, -1.18, -0.36]][[88.6, 88.6, 88.6]][9.46]
MC-based ATE = 2.77
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.0, -1.07, -4.0]][[-1.7, -1.8, -2.33]][[-2.77, -2.77, -2.77]][-2.77]
std:[[0.15, 0.13, 0.24]][[0.12, 0.17, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.01, 1.08, 4.01]][[1.7, 1.81, 2.33]][[2.77, 2.77, 2.77]][2.77]
MSE(-DR):[[0.0, 0.07, 3.0]][[0.69, 0.8, 1.32]][[1.76, 1.76, 1.76]][1.76]
*
==============


O_threshold = 110
MC for this TARGET:[88.696, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.23, -3.5, -3.53]][[-2.9, -3.1, -4.26]][[-88.7, -88.7, -88.7]][-9.55]
std:[[0.06, 0.07, 0.04]][[0.18, 0.17, 0.23]][[0.0, 0.0, 0.0]][0.06]
MSE:[[3.23, 3.5, 3.53]][[2.91, 3.1, 4.27]][[88.7, 88.7, 88.7]][9.55]
MSE(-DR):[[0.0, 0.27, 0.3]][[-0.32, -0.13, 1.04]][[85.47, 85.47, 85.47]][6.32]
MC-based ATE = -0.1
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.27, -1.33, -3.17]][[-3.16, -3.18, -4.02]][[0.1, 0.1, 0.1]][0.1]
std:[[0.28, 0.29, 0.45]][[0.13, 0.12, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.3, 1.36, 3.2]][[3.16, 3.18, 4.02]][[0.1, 0.1, 0.1]][0.1]
MSE(-DR):[[0.0, 0.06, 1.9]][[1.86, 1.88, 2.72]][[-1.2, -1.2, -1.2]][-1.2]
*
==============


O_threshold = 120
MC for this TARGET:[90.808, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.66, -8.86, -7.99]][[-7.71, -7.87, -9.1]][[-90.81, -90.81, -90.81]][-11.66]
std:[[0.73, 0.76, 0.48]][[0.37, 0.33, 0.44]][[0.0, 0.0, 0.0]][0.06]
MSE:[[8.69, 8.89, 8.0]][[7.72, 7.88, 9.11]][[90.81, 90.81, 90.81]][11.66]
MSE(-DR):[[0.0, 0.2, -0.69]][[-0.97, -0.81, 0.42]][[82.12, 82.12, 82.12]][2.97]
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.7, -6.69, -7.62]][[-7.96, -7.95, -8.86]][[-2.01, -2.01, -2.01]][-2.01]
std:[[0.95, 0.98, 0.89]][[0.06, 0.03, 0.22]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.77, 6.76, 7.67]][[7.96, 7.95, 8.86]][[2.01, 2.01, 2.01]][2.01]
MSE(-DR):[[0.0, -0.01, 0.9]][[1.19, 1.18, 2.09]][[-4.76, -4.76, -4.76]][-4.76]
*
==============


[[ 1.99  2.18  0.58  0.38  0.31  0.36 88.79 88.79 88.79  9.65]
 [ 1.08  1.28  0.58  1.44  1.24  0.79 87.32 87.32 87.32  8.17]
 [ 2.98  3.24  4.38  1.52  1.78  2.62 91.56 91.56 91.56 12.42]
 [ 3.27  3.5   3.57  2.91  3.1   4.26 88.7  88.7  88.7   9.55]
```

```
 [ 8.71  8.89  8.04  7.69  7.88  9.03 90.81 90.81 90.81 11.66]]
[[ 1.97  2.18  0.54  0.4   0.31  0.33 88.79 88.79 88.79  9.65]
 [ 1.07  1.28  0.62  1.39  1.24  0.72 87.32 87.32 87.32  8.17]
 [ 2.96  3.24  4.36  1.51  1.78  2.6  91.56 91.56 91.56 12.42]
 [ 3.23  3.5   3.53  2.91  3.1   4.27 88.7  88.7  88.7   9.55]
 [ 8.69  8.89  8.    7.72  7.88  9.11 90.81 90.81 90.81 11.66]]
time spent until now: 12.0 mins


---------------------------------------
[pattern_seed, sd_OD] = [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
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

---------------------------------------
Value of Behaviour policy:79.148
O_threshold = 80
MC for this TARGET:[88.794, 0.146]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.96, -2.17, -0.34]][[0.25, 0.08, -0.23]][[-88.79, -88.79, -88.79]][-9.65]
std:[[0.24, 0.23, 0.43]][[0.3, 0.3, 0.21]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.97, 2.18, 0.55]][[0.39, 0.31, 0.31]][[88.79, 88.79, 88.79]][9.65]
MSE(-DR):[[0.0, 0.21, -1.42]][[-1.58, -1.66, -1.66]][[86.82, 86.82, 86.82]][7.68]
==============


O_threshold = 90
MC for this TARGET:[87.319, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.07, -1.25, -0.51]][[1.33, 1.14, 0.65]][[-87.32, -87.32, -87.32]][-8.17]
std:[[0.3, 0.29, 0.34]][[0.51, 0.5, 0.38]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.11, 1.28, 0.61]][[1.42, 1.24, 0.75]][[87.32, 87.32, 87.32]][8.17]
MSE(-DR):[[0.0, 0.17, -0.5]][[0.31, 0.13, -0.36]][[86.21, 86.21, 86.21]][7.06]
**
MC-based ATE = -1.47
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.89, 0.92, -0.17]][[1.08, 1.06, 0.88]][[1.47, 1.47, 1.47]][1.47]
std:[[0.06, 0.06, 0.08]][[0.21, 0.2, 0.17]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.89, 0.92, 0.19]][[1.1, 1.08, 0.9]][[1.47, 1.47, 1.47]][1.47]
MSE(-DR):[[0.0, 0.03, -0.7]][[0.21, 0.19, 0.01]][[0.58, 0.58, 0.58]][0.58]
**
==============


O_threshold = 100
MC for this TARGET:[91.564, 0.144]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.96, -3.24, -4.36]][[-1.47, -1.72, -2.61]][[-91.56, -91.56, -91.56]][-12.42]
std:[[0.06, 0.09, 0.17]][[0.48, 0.47, 0.43]][[0.0, 0.0, 0.0]][0.06]
MSE:[[2.96, 3.24, 4.36]][[1.55, 1.78, 2.65]][[91.56, 91.56, 91.56]][12.42]
MSE(-DR):[[0.0, 0.28, 1.4]][[-1.41, -1.18, -0.31]][[88.6, 88.6, 88.6]][9.46]
MC-based ATE = 2.77
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.0, -1.07, -4.02]][[-1.71, -1.8, -2.38]][[-2.77, -2.77, -2.77]][-2.77]
std:[[0.18, 0.13, 0.25]][[0.18, 0.17, 0.22]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.02, 1.08, 4.03]][[1.72, 1.81, 2.39]][[2.77, 2.77, 2.77]][2.77]
MSE(-DR):[[0.0, 0.06, 3.01]][[0.7, 0.79, 1.37]][[1.75, 1.75, 1.75]][1.75]
*
==============


O_threshold = 110
MC for this TARGET:[88.696, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.26, -3.5, -3.53]][[-2.9, -3.1, -4.23]][[-88.7, -88.7, -88.7]][-9.55]
std:[[0.02, 0.07, 0.02]][[0.2, 0.17, 0.23]][[0.0, 0.0, 0.0]][0.06]
MSE:[[3.26, 3.5, 3.53]][[2.91, 3.1, 4.24]][[88.7, 88.7, 88.7]][9.55]
MSE(-DR):[[0.0, 0.24, 0.27]][[-0.35, -0.16, 0.98]][[85.44, 85.44, 85.44]][6.29]
MC-based ATE = -0.1
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.3, -1.33, -3.2]][[-3.14, -3.18, -4.0]][[0.1, 0.1, 0.1]][0.1]
std:[[0.26, 0.29, 0.41]][[0.11, 0.12, 0.02]][[0.0, 0.0, 0.0]][0.0]
```

```
MSE:[[1.33, 1.36, 3.23]][[3.14, 3.18, 4.0]][[0.1, 0.1, 0.1]][0.1]
MSE(-DR):[[0.0, 0.03, 1.9]][[1.81, 1.85, 2.67]][[-1.23, -1.23, -1.23]][-1.23]
*
==============


O_threshold = 120
MC for this TARGET:[90.808, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.67, -8.86, -7.99]][[-7.68, -7.87, -9.06]][[-90.81, -90.81, -90.81]][-11.66]
std:[[0.75, 0.76, 0.45]][[0.3, 0.33, 0.37]][[0.0, 0.0, 0.0]][0.06]
MSE:[[8.7, 8.89, 8.0]][[7.69, 7.88, 9.07]][[90.81, 90.81, 90.81]][11.66]
MSE(-DR):[[0.0, 0.19, -0.7]][[-1.01, -0.82, 0.37]][[82.11, 82.11, 82.11]][2.96]
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.7, -6.69, -7.65]][[-7.93, -7.95, -8.83]][[-2.01, -2.01, -2.01]][-2.01]
std:[[0.99, 0.98, 0.88]][[0.0, 0.03, 0.16]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.77, 6.76, 7.7]][[7.93, 7.95, 8.83]][[2.01, 2.01, 2.01]][2.01]
MSE(-DR):[[0.0, -0.01, 0.93]][[1.16, 1.18, 2.06]][[-4.76, -4.76, -4.76]][-4.76]
*
==============


[[ 1.99  2.18  0.58  0.38  0.31  0.36 88.79 88.79 88.79  9.65]
 [ 1.08  1.28  0.58  1.44  1.24  0.79 87.32 87.32 87.32  8.17]
 [ 2.98  3.24  4.38  1.52  1.78  2.62 91.56 91.56 91.56 12.42]
 [ 3.27  3.5   3.57  2.91  3.1   4.26 88.7  88.7  88.7   9.55]
 [ 8.71  8.89  8.04  7.69  7.88  9.03 90.81 90.81 90.81 11.66]]
[[ 1.97  2.18  0.54  0.4   0.31  0.33 88.79 88.79 88.79  9.65]
 [ 1.07  1.28  0.62  1.39  1.24  0.72 87.32 87.32 87.32  8.17]
 [ 2.96  3.24  4.36  1.51  1.78  2.6  91.56 91.56 91.56 12.42]
 [ 3.23  3.5   3.53  2.91  3.1   4.27 88.7  88.7  88.7   9.55]
 [ 8.69  8.89  8.    7.72  7.88  9.11 90.81 90.81 90.81 11.66]]
[[ 1.97  2.18  0.55  0.39  0.31  0.31 88.79 88.79 88.79  9.65]
 [ 1.11  1.28  0.61  1.42  1.24  0.75 87.32 87.32 87.32  8.17]
 [ 2.96  3.24  4.36  1.55  1.78  2.65 91.56 91.56 91.56 12.42]
 [ 3.26  3.5   3.53  2.91  3.1   4.24 88.7  88.7  88.7   9.55]
 [ 8.7   8.89  8.    7.69  7.88  9.07 90.81 90.81 90.81 11.66]]
time spent until now: 17.9 mins


----------------------------------------
[pattern_seed, sd_OD] = [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
```