```
Last login: Tue Mar 31 17:22:32 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 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: Tue Mar 31 21:22:37 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
18:48, 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, 0, False, [3, 7, 14], [80, 90, 100, 110, 120]]


--------------------------------------
[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
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
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:78.322
O_threshold = 80
MC for this TARGET:[87.605, 0.066]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.11, -0.17, -0.53]][[0.66, 0.59, 0.11]][[-87.6, -87.6, -87.6]][-9.28]
std:[[0.28, 0.27, 0.17]][[0.12, 0.12, 0.05]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.3, 0.32, 0.56]][[0.67, 0.6, 0.12]][[87.6, 87.6, 87.6]][9.28]
MSE(-DR):[[0.0, 0.02, 0.26]][[0.37, 0.3, -0.18]][[87.3, 87.3, 87.3]][8.98]
***
==============


O_threshold = 90
MC for this TARGET:[85.961, 0.061]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.59, 0.53, -0.56]][[1.81, 1.72, 1.07]][[-85.96, -85.96, -85.96]][-7.64]
std:[[0.14, 0.14, 0.22]][[0.15, 0.15, 0.11]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.61, 0.55, 0.6]][[1.82, 1.73, 1.08]][[85.96, 85.96, 85.96]][7.64]
MSE(-DR):[[0.0, -0.06, -0.01]][[1.21, 1.12, 0.47]][[85.35, 85.35, 85.35]][7.03]
**
MC-based ATE = -1.64
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.7, 0.7, -0.03]][[1.15, 1.13, 0.96]][[1.64, 1.64, 1.64]][1.64]
std:[[0.16, 0.17, 0.15]][[0.05, 0.04, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.72, 0.72, 0.15]][[1.15, 1.13, 0.96]][[1.64, 1.64, 1.64]][1.64]
MSE(-DR):[[0.0, 0.0, -0.57]][[0.43, 0.41, 0.24]][[0.92, 0.92, 0.92]][0.92]
**
==============


O_threshold = 100
MC for this TARGET:[89.618, 0.065]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.69, -1.76, -3.74]][[-0.46, -0.61, -1.62]][[-89.62, -89.62, -89.62]][-11.3]
std:[[0.21, 0.22, 0.1]][[0.09, 0.09, 0.11]][[0.0, 0.0, 0.0]][0.07]
MSE:[[1.7, 1.77, 3.74]][[0.47, 0.62, 1.62]][[89.62, 89.62, 89.62]][11.3]
MSE(-DR):[[0.0, 0.07, 2.04]][[-1.23, -1.08, -0.08]][[87.92, 87.92, 87.92]][9.6]
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.58, -1.6, -3.21]][[-1.12, -1.2, -1.73]][[-2.01, -2.01, -2.01]][-2.01]
std:[[0.47, 0.48, 0.21]][[0.03, 0.04, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.65, 1.67, 3.22]][[1.12, 1.2, 1.73]][[2.01, 2.01, 2.01]][2.01]
MSE(-DR):[[0.0, 0.02, 1.57]][[-0.53, -0.45, 0.08]][[0.36, 0.36, 0.36]][0.36]
==============


O_threshold = 110
MC for this TARGET:[87.558, 0.053]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

```
bias:[[-2.44, -2.5, -3.37]][[-2.82, -2.92, -4.21]][[-87.56, -87.56, -87.56]][-9.24]
std:[[0.1, 0.09, 0.2]][[0.06, 0.07, 0.12]][[0.0, 0.0, 0.0]][0.07]
MSE:[[2.44, 2.5, 3.38]][[2.82, 2.92, 4.21]][[87.56, 87.56, 87.56]][9.24]
MSE(-DR):[[0.0, 0.06, 0.94]][[0.38, 0.48, 1.77]][[85.12, 85.12, 85.12]][6.8]
***
MC-based ATE = -0.05
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.34, -2.33, -2.84]][[-3.48, -3.51, -4.32]][[0.05, 0.05, 0.05]][0.05]
std:[[0.22, 0.21, 0.05]][[0.06, 0.05, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.35, 2.34, 2.84]][[3.48, 3.51, 4.32]][[0.05, 0.05, 0.05]][0.05]
MSE(-DR):[[0.0, -0.01, 0.49]][[1.13, 1.16, 1.97]][[-2.3, -2.3, -2.3]][-2.3]
*
==============


O_threshold = 120
MC for this TARGET:[89.165, 0.055]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.82, -6.86, -7.18]][[-7.05, -7.13, -8.48]][[-89.16, -89.16, -89.16]][-10.84]
std:[[0.37, 0.36, 0.14]][[0.08, 0.07, 0.13]][[0.0, 0.0, 0.0]][0.07]
MSE:[[6.83, 6.87, 7.18]][[7.05, 7.13, 8.48]][[89.16, 89.16, 89.16]][10.84]
MSE(-DR):[[0.0, 0.04, 0.35]][[0.22, 0.3, 1.65]][[82.33, 82.33, 82.33]][4.01]
***
MC-based ATE = 1.56
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.71, -6.69, -6.65]][[-7.71, -7.72, -8.59]][[-1.56, -1.56, -1.56]][-1.56]
std:[[0.47, 0.48, 0.05]][[0.06, 0.08, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.73, 6.71, 6.65]][[7.71, 7.72, 8.59]][[1.56, 1.56, 1.56]][1.56]
MSE(-DR):[[0.0, -0.02, -0.08]][[0.98, 0.99, 1.86]][[-5.17, -5.17, -5.17]][-5.17]
**
==============


time spent until now: 8.3 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
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
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:78.338
O_threshold = 80
MC for this TARGET:[87.603, 0.074]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.23, -0.29, -0.59]][[0.64, 0.57, 0.1]][[-87.6, -87.6, -87.6]][-9.26]
std:[[0.33, 0.31, 0.23]][[0.15, 0.13, 0.07]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.4, 0.42, 0.63]][[0.66, 0.58, 0.12]][[87.6, 87.6, 87.6]][9.26]
MSE(-DR):[[0.0, 0.02, 0.23]][[0.26, 0.18, -0.28]][[87.2, 87.2, 87.2]][8.86]
***
==============


O_threshold = 90
MC for this TARGET:[85.96, 0.068]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.57, 0.5, -0.58]][[1.77, 1.7, 1.05]][[-85.96, -85.96, -85.96]][-7.62]
std:[[0.12, 0.11, 0.19]][[0.17, 0.17, 0.16]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.58, 0.51, 0.61]][[1.78, 1.71, 1.06]][[85.96, 85.96, 85.96]][7.62]
MSE(-DR):[[0.0, -0.07, 0.03]][[1.2, 1.13, 0.48]][[85.38, 85.38, 85.38]][7.04]
***
MC-based ATE = -1.64
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.79, 0.78, 0.0]][[1.14, 1.14, 0.96]][[1.64, 1.64, 1.64]][1.64]
std:[[0.21, 0.21, 0.17]][[0.08, 0.09, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.82, 0.81, 0.17]][[1.14, 1.14, 0.97]][[1.64, 1.64, 1.64]][1.64]
MSE(-DR):[[0.0, -0.01, -0.65]][[0.32, 0.32, 0.15]][[0.82, 0.82, 0.82]][0.82]
**
==============


O_threshold = 100
MC for this TARGET:[89.616, 0.074]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.75, -1.83, -3.71]][[-0.45, -0.57, -1.67]][[-89.62, -89.62, -89.62]][-11.28]
std:[[0.21, 0.21, 0.17]][[0.13, 0.1, 0.16]][[0.0, 0.0, 0.0]][0.07]
MSE:[[1.76, 1.84, 3.71]][[0.47, 0.58, 1.68]][[89.62, 89.62, 89.62]][11.28]
MSE(-DR):[[0.0, 0.08, 1.95]][[-1.29, -1.18, -0.08]][[87.86, 87.86, 87.86]][9.52]
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.53, -1.54, -3.12]][[-1.08, -1.13, -1.76]][[-2.01, -2.01, -2.01]][-2.01]
std:[[0.53, 0.53, 0.3]][[0.08, 0.07, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.62, 1.63, 3.13]][[1.08, 1.13, 1.76]][[2.01, 2.01, 2.01]][2.01]
MSE(-DR):[[0.0, 0.01, 1.51]][[-0.54, -0.49, 0.14]][[0.39, 0.39, 0.39]][0.39]
==============


O_threshold = 110
MC for this TARGET:[87.556, 0.06]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.48, -2.55, -3.38]][[-2.79, -2.88, -4.19]][[-87.56, -87.56, -87.56]][-9.22]
std:[[0.2, 0.18, 0.13]][[0.1, 0.07, 0.17]][[0.0, 0.0, 0.0]][0.07]
```

```
MSE:[[2.49, 2.56, 3.38]][[2.79, 2.88, 4.19]][[87.56, 87.56, 87.56]][9.22]
MSE(-DR):[[0.0, 0.07, 0.89]][[0.3, 0.39, 1.7]][[85.07, 85.07, 85.07]][6.73]
***
MC-based ATE = -0.05
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.26, -2.26, -2.79]][[-3.43, -3.45, -4.28]][[0.05, 0.05, 0.05]][0.05]
std:[[0.2, 0.18, 0.13]][[0.07, 0.07, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.27, 2.27, 2.79]][[3.43, 3.45, 4.28]][[0.05, 0.05, 0.05]][0.05]
MSE(-DR):[[0.0, 0.0, 0.52]][[1.16, 1.18, 2.01]][[-2.22, -2.22, -2.22]][-2.22]
*
==============


O_threshold = 120
MC for this TARGET:[89.164, 0.063]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.86, -6.92, -7.17]][[-7.0, -7.07, -8.41]][[-89.16, -89.16, -89.16]][-10.83]
std:[[0.36, 0.34, 0.16]][[0.07, 0.06, 0.15]][[0.0, 0.0, 0.0]][0.07]
MSE:[[6.87, 6.93, 7.17]][[7.0, 7.07, 8.41]][[89.16, 89.16, 89.16]][10.83]
MSE(-DR):[[0.0, 0.06, 0.3]][[0.13, 0.2, 1.54]][[82.29, 82.29, 82.29]][3.96]
***
MC-based ATE = 1.56
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.63, -6.64, -6.58]][[-7.63, -7.64, -8.51]][[-1.56, -1.56, -1.56]][-1.56]
std:[[0.34, 0.35, 0.09]][[0.08, 0.08, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.64, 6.65, 6.58]][[7.63, 7.64, 8.51]][[1.56, 1.56, 1.56]][1.56]
MSE(-DR):[[0.0, 0.01, -0.06]][[0.99, 1.0, 1.87]][[-5.08, -5.08, -5.08]][-5.08]
**
==============


time spent until now: 17.0 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
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
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:78.357
O_threshold = 80
MC for this TARGET:[87.601, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.35, -0.42, -0.62]][[0.61, 0.54, 0.09]][[-87.6, -87.6, -87.6]][-9.24]
std:[[0.4, 0.39, 0.35]][[0.15, 0.14, 0.08]][[0.0, 0.0, 0.0]][0.08]
MSE:[[0.53, 0.57, 0.71]][[0.63, 0.56, 0.12]][[87.6, 87.6, 87.6]][9.24]
MSE(-DR):[[0.0, 0.04, 0.18]][[0.1, 0.03, -0.41]][[87.07, 87.07, 87.07]][8.71]
***
==============


O_threshold = 90
MC for this TARGET:[85.958, 0.09]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.53, 0.46, -0.57]][[1.78, 1.68, 1.09]][[-85.96, -85.96, -85.96]][-7.6]
std:[[0.24, 0.21, 0.22]][[0.22, 0.21, 0.2]][[0.0, 0.0, 0.0]][0.08]
MSE:[[0.58, 0.51, 0.61]][[1.79, 1.69, 1.11]][[85.96, 85.96, 85.96]][7.6]
MSE(-DR):[[0.0, -0.07, 0.03]][[1.21, 1.11, 0.53]][[85.38, 85.38, 85.38]][7.02]
***
MC-based ATE = -1.64
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.88, 0.88, 0.05]][[1.17, 1.14, 1.0]][[1.64, 1.64, 1.64]][1.64]
std:[[0.24, 0.26, 0.22]][[0.13, 0.14, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.91, 0.92, 0.23]][[1.18, 1.15, 1.01]][[1.64, 1.64, 1.64]][1.64]
MSE(-DR):[[0.0, 0.01, -0.68]][[0.27, 0.24, 0.1]][[0.73, 0.73, 0.73]][0.73]
**
==============


O_threshold = 100
MC for this TARGET:[89.615, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.81, -1.9, -3.67]][[-0.4, -0.52, -1.58]][[-89.62, -89.62, -89.62]][-11.26]
std:[[0.2, 0.21, 0.26]][[0.14, 0.13, 0.18]][[0.0, 0.0, 0.0]][0.08]
MSE:[[1.82, 1.91, 3.68]][[0.42, 0.54, 1.59]][[89.62, 89.62, 89.62]][11.26]
MSE(-DR):[[0.0, 0.09, 1.86]][[-1.4, -1.28, -0.23]][[87.8, 87.8, 87.8]][9.44]
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.46, -1.48, -3.05]][[-1.01, -1.06, -1.67]][[-2.01, -2.01, -2.01]][-2.01]
std:[[0.61, 0.6, 0.45]][[0.1, 0.12, 0.12]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.58, 1.6, 3.08]][[1.01, 1.07, 1.67]][[2.01, 2.01, 2.01]][2.01]
MSE(-DR):[[0.0, 0.02, 1.5]][[-0.57, -0.51, 0.09]][[0.43, 0.43, 0.43]][0.43]
==============


O_threshold = 110
MC for this TARGET:[87.554, 0.083]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.53, -2.6, -3.36]][[-2.76, -2.85, -4.13]][[-87.55, -87.55, -87.55]][-9.2]
std:[[0.33, 0.32, 0.15]][[0.1, 0.08, 0.16]][[0.0, 0.0, 0.0]][0.08]
MSE:[[2.55, 2.62, 3.36]][[2.76, 2.85, 4.13]][[87.55, 87.55, 87.55]][9.2]
MSE(-DR):[[0.0, 0.07, 0.81]][[0.21, 0.3, 1.58]][[85.0, 85.0, 85.0]][6.65]
```

```
***
MC-based ATE = -0.05
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.18, -2.18, -2.74]][[-3.36, -3.39, -4.22]][[0.05, 0.05, 0.05]][0.05]
std:[[0.36, 0.35, 0.28]][[0.06, 0.1, 0.09]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.21, 2.21, 2.75]][[3.36, 3.39, 4.22]][[0.05, 0.05, 0.05]][0.05]
MSE(-DR):[[0.0, 0.0, 0.54]][[1.15, 1.18, 2.01]][[-2.16, -2.16, -2.16]][-2.16]
*
==============


O_threshold = 120
MC for this TARGET:[89.162, 0.087]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.96, -6.99, -7.33]][[-6.94, -7.02, -8.37]][[-89.16, -89.16, -89.16]][-10.81]
std:[[0.37, 0.35, 0.15]][[0.09, 0.07, 0.15]][[0.0, 0.0, 0.0]][0.08]
MSE:[[6.97, 7.0, 7.33]][[6.94, 7.02, 8.37]][[89.16, 89.16, 89.16]][10.81]
MSE(-DR):[[0.0, 0.03, 0.36]][[-0.03, 0.05, 1.4]][[82.19, 82.19, 82.19]][3.84]
MC-based ATE = 1.56
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.61, -6.57, -6.71]][[-7.55, -7.55, -8.46]][[-1.56, -1.56, -1.56]][-1.56]
std:[[0.2, 0.21, 0.2]][[0.05, 0.08, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.61, 6.57, 6.71]][[7.55, 7.55, 8.46]][[1.56, 1.56, 1.56]][1.56]
MSE(-DR):[[0.0, -0.04, 0.1]][[0.94, 0.94, 1.85]][[-5.05, -5.05, -5.05]][-5.05]
*
==============


time spent until now: 25.6 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
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
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:78.375
O_threshold = 80
MC for this TARGET:[87.599, 0.125]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.49, -0.55, -0.68]][[0.59, 0.51, 0.11]][[-87.6, -87.6, -87.6]][-9.22]
std:[[0.51, 0.49, 0.47]][[0.17, 0.16, 0.09]][[0.0, 0.0, 0.0]][0.09]
MSE:[[0.71, 0.74, 0.83]][[0.61, 0.53, 0.14]][[87.6, 87.6, 87.6]][9.22]
MSE(-DR):[[0.0, 0.03, 0.12]][[-0.1, -0.18, -0.57]][[86.89, 86.89, 86.89]][8.51]
==============


O_threshold = 90
MC for this TARGET:[85.956, 0.119]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.49, 0.43, -0.63]][[1.76, 1.66, 1.1]][[-85.96, -85.96, -85.96]][-7.58]
std:[[0.36, 0.35, 0.25]][[0.29, 0.26, 0.26]][[0.0, 0.0, 0.0]][0.09]
MSE:[[0.61, 0.55, 0.68]][[1.78, 1.68, 1.13]][[85.96, 85.96, 85.96]][7.58]
MSE(-DR):[[0.0, -0.06, 0.07]][[1.17, 1.07, 0.52]][[85.35, 85.35, 85.35]][6.97]
***
MC-based ATE = -1.64
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.97, 0.98, 0.05]][[1.16, 1.15, 0.99]][[1.64, 1.64, 1.64]][1.64]
std:[[0.28, 0.31, 0.27]][[0.19, 0.19, 0.17]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.01, 1.03, 0.27]][[1.18, 1.17, 1.0]][[1.64, 1.64, 1.64]][1.64]
MSE(-DR):[[0.0, 0.02, -0.74]][[0.17, 0.16, -0.01]][[0.63, 0.63, 0.63]][0.63]
**
==============


O_threshold = 100
MC for this TARGET:[89.613, 0.125]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.9, -1.96, -3.67]][[-0.36, -0.48, -1.54]][[-89.61, -89.61, -89.61]][-11.24]
std:[[0.2, 0.2, 0.32]][[0.2, 0.18, 0.24]][[0.0, 0.0, 0.0]][0.09]
MSE:[[1.91, 1.97, 3.68]][[0.41, 0.51, 1.56]][[89.61, 89.61, 89.61]][11.24]
MSE(-DR):[[0.0, 0.06, 1.77]][[-1.5, -1.4, -0.35]][[87.7, 87.7, 87.7]][9.33]
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.41, -1.41, -2.99]][[-0.96, -0.99, -1.65]][[-2.01, -2.01, -2.01]][-2.01]
std:[[0.7, 0.68, 0.55]][[0.16, 0.16, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.57, 1.57, 3.04]][[0.97, 1.0, 1.66]][[2.01, 2.01, 2.01]][2.01]
MSE(-DR):[[0.0, 0.0, 1.47]][[-0.6, -0.57, 0.09]][[0.44, 0.44, 0.44]][0.44]
==============


O_threshold = 110
MC for this TARGET:[87.552, 0.113]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.59, -2.65, -3.38]][[-2.71, -2.81, -4.09]][[-87.55, -87.55, -87.55]][-9.18]
std:[[0.47, 0.47, 0.07]][[0.13, 0.1, 0.22]][[0.0, 0.0, 0.0]][0.09]
MSE:[[2.63, 2.69, 3.38]][[2.71, 2.81, 4.1]][[87.55, 87.55, 87.55]][9.18]
MSE(-DR):[[0.0, 0.06, 0.75]][[0.08, 0.18, 1.47]][[84.92, 84.92, 84.92]][6.55]
***
MC-based ATE = -0.05
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.1, -2.1, -2.7]][[-3.3, -3.32, -4.2]][[0.05, 0.05, 0.05]][0.05]
```

```
std:[[0.6, 0.57, 0.46]][[0.09, 0.12, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.18, 2.18, 2.74]][[3.3, 3.32, 4.2]][[0.05, 0.05, 0.05]][0.05]
MSE(-DR):[[0.0, 0.0, 0.56]][[1.12, 1.14, 2.02]][[-2.13, -2.13, -2.13]][-2.13]
*
==============


O_threshold = 120
MC for this TARGET:[89.16, 0.117]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.03, -7.06, -7.34]][[-6.88, -6.96, -8.3]][[-89.16, -89.16, -89.16]][-10.78]
std:[[0.4, 0.38, 0.17]][[0.11, 0.08, 0.21]][[0.0, 0.0, 0.0]][0.09]
MSE:[[7.04, 7.07, 7.34]][[6.88, 6.96, 8.3]][[89.16, 89.16, 89.16]][10.78]
MSE(-DR):[[0.0, 0.03, 0.3]][[-0.16, -0.08, 1.26]][[82.12, 82.12, 82.12]][3.74]
MC-based ATE = 1.56
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.54, -6.51, -6.66]][[-7.47, -7.47, -8.41]][[-1.56, -1.56, -1.56]][-1.56]
std:[[0.11, 0.13, 0.3]][[0.07, 0.09, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.54, 6.51, 6.67]][[7.47, 7.47, 8.41]][[1.56, 1.56, 1.56]][1.56]
MSE(-DR):[[0.0, -0.03, 0.13]][[0.93, 0.93, 1.87]][[-4.98, -4.98, -4.98]][-4.98]
*
==============


time spent until now: 34.4 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
1 -th target; 2 -th target;
```