```
Last login: Mon Mar 30 21:14:13 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-18-232-178-254.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

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


*** System restart required ***
Last login: Tue Mar 31 01:14:26 2020 from 107.13.161.147
ubuntu@ip-172-31-8-160:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1
ubuntu@ip-172-31-8-160:~$ python EC2.py
  File "EC2.py", line 20
    sd_D = 1o #1
           ^
SyntaxError: invalid syntax
ubuntu@ip-172-31-8-160:~$ python EC2.py
  File "EC2.py", line 20
    sd_D = 1o #1
           ^
SyntaxError: invalid syntax
ubuntu@ip-172-31-8-160:~$ python EC2.py
  File "EC2.py", line 20
    sd_D = 1o #1
           ^
SyntaxError: invalid syntax
ubuntu@ip-172-31-8-160:~$
ubuntu@ip-172-31-8-160:~$ python EC2.py
22:10, 03/30; 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_ran
ge] = [None, 10, 10, 5, 0.2, 0.5, 1, 0.0001, False, True, 5, False, [3, 7, 14], [80, 90, 100, 110, 120, 130]]


---------------------------------------
[pattern_seed, T, sd_R] = [0, 144, 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
Q_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
Q_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
Q_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
Q_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
Q_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:73.049
O_threshold = 80
MC for this TARGET:[83.963, 0.188]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.09, -0.2, 0.21]][[2.9, 2.74, 2.36]][[-83.96, -83.96, -83.96]][-10.91]
std:[[0.87, 0.87, 0.59]][[0.15, 0.18, 0.16]][[0.0, 0.0, 0.0]][0.26]
MSE:[[0.87, 0.89, 0.63]][[2.9, 2.75, 2.37]][[83.96, 83.96, 83.96]][10.91]
MSE(-DR):[[0.0, 0.02, -0.24]][[2.03, 1.88, 1.5]][[83.09, 83.09, 83.09]][10.04]
**
==============


O_threshold = 90
MC for this TARGET:[81.135, 0.162]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.95, 0.84, 0.6]][[4.36, 4.16, 3.75]][[-81.14, -81.14, -81.14]][-8.09]
std:[[0.95, 0.94, 0.49]][[0.08, 0.12, 0.12]][[0.0, 0.0, 0.0]][0.26]
MSE:[[1.34, 1.26, 0.77]][[4.36, 4.16, 3.75]][[81.14, 81.14, 81.14]][8.09]
MSE(-DR):[[0.0, -0.08, -0.57]][[3.02, 2.82, 2.41]][[79.8, 79.8, 79.8]][6.75]
**
MC-based ATE = -2.83
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.05, 1.04, 0.39]][[1.46, 1.41, 1.39]][[2.83, 2.83, 2.83]][2.83]
std:[[0.08, 0.07, 0.19]][[0.16, 0.17, 0.14]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.05, 1.04, 0.43]][[1.47, 1.42, 1.4]][[2.83, 2.83, 2.83]][2.83]
MSE(-DR):[[0.0, -0.01, -0.62]][[0.42, 0.37, 0.35]][[1.78, 1.78, 1.78]][1.78]
**
==============


O_threshold = 100
MC for this TARGET:[84.535, 0.155]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.57, -0.69, -2.5]][[1.84, 1.6, 0.89]][[-84.54, -84.54, -84.54]][-11.49]
std:[[0.24, 0.23, 0.33]][[0.02, 0.03, 0.24]][[0.0, 0.0, 0.0]][0.26]
MSE:[[0.62, 0.73, 2.52]][[1.84, 1.6, 0.92]][[84.54, 84.54, 84.54]][11.49]
MSE(-DR):[[0.0, 0.11, 1.9]][[1.22, 0.98, 0.3]][[83.92, 83.92, 83.92]][10.87]
***
MC-based ATE = 0.57
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.48, -0.49, -2.71]][[-1.07, -1.14, -1.48]][[-0.57, -0.57, -0.57]][-0.57]
std:[[1.05, 1.07, 0.76]][[0.14, 0.16, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.15, 1.18, 2.81]][[1.08, 1.15, 1.49]][[0.57, 0.57, 0.57]][0.57]
MSE(-DR):[[0.0, 0.03, 1.66]][[-0.07, 0.0, 0.34]][[-0.58, -0.58, -0.58]][-0.58]
==============


O_threshold = 110
MC for this TARGET:[80.437, 0.138]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.07, -1.17, -1.36]][[-1.19, -1.39, -1.87]][[-80.44, -80.44, -80.44]][-7.39]
std:[[0.61, 0.6, 0.34]][[0.41, 0.34, 0.37]][[0.0, 0.0, 0.0]][0.26]
MSE:[[1.23, 1.31, 1.4]][[1.26, 1.43, 1.91]][[80.44, 80.44, 80.44]][7.39]
MSE(-DR):[[0.0, 0.08, 0.17]][[0.03, 0.2, 0.68]][[79.21, 79.21, 79.21]][6.16]
***
MC-based ATE = -3.53
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.98, -0.97, -1.57]][[-4.09, -4.13, -4.23]][[3.53, 3.53, 3.53]][3.53]
std:[[0.91, 0.95, 0.46]][[0.53, 0.5, 0.36]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.34, 1.36, 1.64]][[4.12, 4.16, 4.25]][[3.53, 3.53, 3.53]][3.53]
MSE(-DR):[[0.0, 0.02, 0.3]][[2.78, 2.82, 2.91]][[2.19, 2.19, 2.19]][2.19]
*
==============


O_threshold = 120
MC for this TARGET:[82.265, 0.146]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.88, -6.96, -6.28]][[-7.0, -7.2, -7.64]][[-82.26, -82.26, -82.26]][-9.22]
std:[[1.35, 1.34, 0.52]][[0.46, 0.38, 0.41]][[0.0, 0.0, 0.0]][0.26]
MSE:[[7.01, 7.09, 6.3]][[7.02, 7.21, 7.65]][[82.26, 82.26, 82.26]][9.22]
MSE(-DR):[[0.0, 0.08, -0.71]][[0.01, 0.2, 0.64]][[75.25, 75.25, 75.25]][2.21]
**
MC-based ATE = -1.7
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.79, -6.77, -6.49]][[-9.9, -9.95, -10.01]][[1.7, 1.7, 1.7]][1.7]
std:[[1.91, 1.92, 1.02]][[0.6, 0.55, 0.45]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.05, 7.04, 6.57]][[9.92, 9.97, 10.02]][[1.7, 1.7, 1.7]][1.7]
MSE(-DR):[[0.0, -0.01, -0.48]][[2.87, 2.92, 2.97]][[-5.35, -5.35, -5.35]][-5.35]
**
==============


O_threshold = 130
MC for this TARGET:[86.657, 0.157]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.84, -10.84, -9.95]][[-13.74, -13.87, -14.4]][[-86.66, -86.66, -86.66]][-13.61]
std:[[1.32, 1.3, 0.5]][[0.58, 0.5, 0.5]][[0.0, 0.0, 0.0]][0.26]
MSE:[[10.92, 10.92, 9.96]][[13.75, 13.88, 14.41]][[86.66, 86.66, 86.66]][13.61]
MSE(-DR):[[0.0, 0.0, -0.96]][[2.83, 2.96, 3.49]][[75.74, 75.74, 75.74]][2.69]
**
MC-based ATE = 2.69
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.75, -10.64, -10.16]][[-16.64, -16.61, -16.77]][[-2.69, -2.69, -2.69]][-2.69]
std:[[1.44, 1.46, 1.02]][[0.71, 0.67, 0.51]][[0.0, 0.0, 0.0]][0.0]
MSE:[[10.85, 10.74, 10.21]][[16.66, 16.62, 16.78]][[2.69, 2.69, 2.69]][2.69]
MSE(-DR):[[0.0, -0.11, -0.64]][[5.81, 5.77, 5.93]][[-8.16, -8.16, -8.16]][-8.16]
**
==============


time spent until now: 12.1 mins


---------------------------------------
[pattern_seed, T, sd_R] = [0, 336, 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:72.626
O_threshold = 80
MC for this TARGET:[83.957, 0.113]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.2, 0.13, 0.1]][[2.17, 1.99, 1.58]][[-83.96, -83.96, -83.96]][-11.33]
std:[[0.78, 0.8, 0.1]][[0.18, 0.19, 0.21]][[0.0, 0.0, 0.0]][0.23]
MSE:[[0.81, 0.81, 0.14]][[2.18, 2.0, 1.59]][[83.96, 83.96, 83.96]][11.33]
MSE(-DR):[[0.0, 0.0, -0.67]][[1.37, 1.19, 0.78]][[83.15, 83.15, 83.15]][10.52]
**
==============


O_threshold = 90
MC for this TARGET:[81.146, 0.109]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.89, 0.8, 0.42]][[3.6, 3.4, 2.97]][[-81.15, -81.15, -81.15]][-8.52]
std:[[0.8, 0.81, 0.07]][[0.23, 0.23, 0.23]][[0.0, 0.0, 0.0]][0.23]
MSE:[[1.2, 1.14, 0.43]][[3.61, 3.41, 2.98]][[81.15, 81.15, 81.15]][8.52]
MSE(-DR):[[0.0, -0.06, -0.77]][[2.41, 2.21, 1.78]][[79.95, 79.95, 79.95]][7.32]
**
MC-based ATE = -2.81
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.69, 0.67, 0.32]][[1.43, 1.41, 1.39]][[2.81, 2.81, 2.81]][2.81]
std:[[0.05, 0.05, 0.04]][[0.11, 0.1, 0.05]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.69, 0.67, 0.32]][[1.43, 1.41, 1.39]][[2.81, 2.81, 2.81]][2.81]
MSE(-DR):[[0.0, -0.02, -0.37]][[0.74, 0.72, 0.7]][[2.12, 2.12, 2.12]][2.12]
**
==============


O_threshold = 100
MC for this TARGET:[84.559, 0.107]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.65, -0.76, -2.57]][[1.13, 0.88, 0.24]][[-84.56, -84.56, -84.56]][-11.93]
std:[[0.83, 0.85, 0.14]][[0.27, 0.26, 0.32]][[0.0, 0.0, 0.0]][0.23]
MSE:[[1.05, 1.14, 2.57]][[1.16, 0.92, 0.4]][[84.56, 84.56, 84.56]][11.93]
MSE(-DR):[[0.0, 0.09, 1.52]][[0.11, -0.13, -0.65]][[83.51, 83.51, 83.51]][10.88]
***
MC-based ATE = 0.6
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.85, -0.89, -2.67]][[-1.03, -1.11, -1.34]][[-0.6, -0.6, -0.6]][-0.6]
std:[[0.46, 0.48, 0.12]][[0.18, 0.17, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.97, 1.01, 2.67]][[1.05, 1.12, 1.34]][[0.6, 0.6, 0.6]][0.6]
MSE(-DR):[[0.0, 0.04, 1.7]][[0.08, 0.15, 0.37]][[-0.37, -0.37, -0.37]][-0.37]
*
==============


O_threshold = 110
MC for this TARGET:[80.465, 0.103]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.59, -0.74, -1.46]][[-1.56, -1.76, -2.32]][[-80.46, -80.46, -80.46]][-7.84]
std:[[0.34, 0.35, 0.31]][[0.21, 0.22, 0.27]][[0.0, 0.0, 0.0]][0.23]
MSE:[[0.68, 0.82, 1.49]][[1.57, 1.77, 2.34]][[80.46, 80.46, 80.46]][7.84]
MSE(-DR):[[0.0, 0.14, 0.81]][[0.89, 1.09, 1.66]][[79.78, 79.78, 79.78]][7.16]
***
MC-based ATE = -3.49
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.79, -0.87, -1.56]][[-3.72, -3.76, -3.9]][[3.49, 3.49, 3.49]][3.49]
std:[[1.0, 1.03, 0.25]][[0.17, 0.17, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.27, 1.35, 1.58]][[3.72, 3.76, 3.9]][[3.49, 3.49, 3.49]][3.49]
MSE(-DR):[[0.0, 0.08, 0.31]][[2.45, 2.49, 2.63]][[2.22, 2.22, 2.22]][2.22]
*
==============


O_threshold = 120
MC for this TARGET:[82.262, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.8, -5.88, -6.25]][[-7.44, -7.57, -8.06]][[-82.26, -82.26, -82.26]][-9.64]
std:[[0.63, 0.67, 0.23]][[0.26, 0.23, 0.25]][[0.0, 0.0, 0.0]][0.23]
MSE:[[5.83, 5.92, 6.25]][[7.44, 7.57, 8.06]][[82.26, 82.26, 82.26]][9.64]
MSE(-DR):[[0.0, 0.09, 0.42]][[1.61, 1.74, 2.23]][[76.43, 76.43, 76.43]][3.81]
***
MC-based ATE = -1.69
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.0, -6.01, -6.36]][[-9.61, -9.56, -9.64]][[1.69, 1.69, 1.69]][1.69]
std:[[1.29, 1.35, 0.17]][[0.26, 0.23, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.14, 6.16, 6.36]][[9.61, 9.56, 9.64]][[1.69, 1.69, 1.69]][1.69]
MSE(-DR):[[0.0, 0.02, 0.22]][[3.47, 3.42, 3.5]][[-4.45, -4.45, -4.45]][-4.45]
*
==============


O_threshold = 130
MC for this TARGET:[86.65, 0.095]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.23, -11.3, -10.45]][[-14.11, -14.23, -14.68]][[-86.65, -86.65, -86.65]][-14.02]
std:[[0.45, 0.48, 0.27]][[0.26, 0.25, 0.27]][[0.0, 0.0, 0.0]][0.23]
MSE:[[11.24, 11.31, 10.45]][[14.11, 14.23, 14.68]][[86.65, 86.65, 86.65]][14.02]
MSE(-DR):[[0.0, 0.07, -0.79]][[2.87, 2.99, 3.44]][[75.41, 75.41, 75.41]][2.78]
**
MC-based ATE = 2.69
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.43, -11.43, -10.55]][[-16.28, -16.22, -16.26]][[-2.69, -2.69, -2.69]][-2.69]
std:[[1.19, 1.24, 0.17]][[0.29, 0.28, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.49, 11.5, 10.55]][[16.28, 16.22, 16.26]][[2.69, 2.69, 2.69]][2.69]
MSE(-DR):[[0.0, 0.01, -0.94]][[4.79, 4.73, 4.77]][[-8.8, -8.8, -8.8]][-8.8]
**
```

==========

time spent until now: 24.4 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:72.847
O_threshold = 80
MC for this TARGET:[83.948, 0.075]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.11, 0.99, 0.39]][[2.11, 1.92, 1.71]][[-83.95, -83.95, -83.95]][-11.1]
std:[[0.34, 0.33, 0.26]][[0.08, 0.09, 0.04]][[0.0, 0.0, 0.0]][0.05]
MSE:[[1.16, 1.04, 0.47]][[2.11, 1.92, 1.71]][[83.95, 83.95, 83.95]][11.1]
MSE(-DR):[[0.0, -0.12, -0.69]][[0.95, 0.76, 0.55]][[82.79, 82.79, 82.79]][9.94]
**
==============


O_threshold = 90
MC for this TARGET:[81.134, 0.067]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[2.18, 2.07, 0.67]][[3.61, 3.41, 3.16]][[-81.13, -81.13, -81.13]][-8.29]
std:[[0.2, 0.19, 0.26]][[0.18, 0.15, 0.19]][[0.0, 0.0, 0.0]][0.05]
MSE:[[2.19, 2.08, 0.72]][[3.61, 3.41, 3.17]][[81.13, 81.13, 81.13]][8.29]
MSE(-DR):[[0.0, -0.11, -1.47]][[1.42, 1.22, 0.98]][[78.94, 78.94, 78.94]][6.1]
**
MC-based ATE = -2.81
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.07, 1.08, 0.29]][[1.5, 1.49, 1.45]][[2.81, 2.81, 2.81]][2.81]
std:[[0.19, 0.17, 0.22]][[0.13, 0.1, 0.17]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.09, 1.09, 0.36]][[1.51, 1.49, 1.46]][[2.81, 2.81, 2.81]][2.81]
MSE(-DR):[[0.0, 0.0, -0.73]][[0.42, 0.4, 0.37]][[1.72, 1.72, 1.72]][1.72]
**
==============


O_threshold = 100
MC for this TARGET:[84.549, 0.072]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.03, -0.19, -2.39]][[1.29, 1.04, 0.56]][[-84.55, -84.55, -84.55]][-11.7]
std:[[0.2, 0.19, 0.14]][[0.12, 0.11, 0.15]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.2, 0.27, 2.39]][[1.3, 1.05, 0.58]][[84.55, 84.55, 84.55]][11.7]
MSE(-DR):[[0.0, 0.07, 2.19]][[1.1, 0.85, 0.38]][[84.35, 84.35, 84.35]][11.5]
***
MC-based ATE = 0.6
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.14, -1.18, -2.78]][[-0.82, -0.88, -1.15]][[-0.6, -0.6, -0.6]][-0.6]
std:[[0.54, 0.52, 0.27]][[0.07, 0.06, 0.14]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.26, 1.29, 2.79]][[0.82, 0.88, 1.16]][[0.6, 0.6, 0.6]][0.6]
MSE(-DR):[[0.0, 0.03, 1.53]][[-0.44, -0.38, -0.1]][[-0.66, -0.66, -0.66]][-0.66]
==============
```

```
O_threshold = 110
MC for this TARGET:[80.45, 0.059]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.84, -0.97, -1.19]][[-1.55, -1.76, -2.17]][[-80.45, -80.45, -80.45]][-7.6]
std:[[0.15, 0.12, 0.05]][[0.14, 0.12, 0.18]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.85, 0.98, 1.19]][[1.56, 1.76, 2.18]][[80.45, 80.45, 80.45]][7.6]
MSE(-DR):[[0.0, 0.13, 0.34]][[0.71, 0.91, 1.33]][[79.6, 79.6, 79.6]][6.75]
***
MC-based ATE = -3.5
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.95, -1.95, -1.58]][[-3.67, -3.67, -3.88]][[3.5, 3.5, 3.5]][3.5]
std:[[0.3, 0.28, 0.21]][[0.06, 0.04, 0.17]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.97, 1.97, 1.59]][[3.67, 3.67, 3.88]][[3.5, 3.5, 3.5]][3.5]
MSE(-DR):[[0.0, 0.0, -0.38]][[1.7, 1.7, 1.91]][[1.53, 1.53, 1.53]][1.53]
**
==============


O_threshold = 120
MC for this TARGET:[82.255, 0.058]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.08, -6.13, -5.91]][[-7.27, -7.4, -7.84]][[-82.26, -82.26, -82.26]][-9.41]
std:[[0.34, 0.37, 0.1]][[0.16, 0.14, 0.18]][[0.0, 0.0, 0.0]][0.05]
MSE:[[6.09, 6.14, 5.91]][[7.27, 7.4, 7.84]][[82.26, 82.26, 82.26]][9.41]
MSE(-DR):[[0.0, 0.05, -0.18]][[1.18, 1.31, 1.75]][[76.17, 76.17, 76.17]][3.32]
**
MC-based ATE = -1.69
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.19, -7.12, -6.3]][[-9.39, -9.31, -9.55]][[1.69, 1.69, 1.69]][1.69]
std:[[0.36, 0.39, 0.16]][[0.08, 0.05, 0.17]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.2, 7.13, 6.3]][[9.39, 9.31, 9.55]][[1.69, 1.69, 1.69]][1.69]
MSE(-DR):[[0.0, -0.07, -0.9]][[2.19, 2.11, 2.35]][[-5.51, -5.51, -5.51]][-5.51]
**
==============


O_threshold = 130
MC for this TARGET:[86.646, 0.06]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.4, -11.4, -10.27]][[-13.96, -14.07, -14.48]][[-86.65, -86.65, -86.65]][-13.8]
std:[[0.19, 0.18, 0.07]][[0.11, 0.08, 0.18]][[0.0, 0.0, 0.0]][0.05]
MSE:[[11.4, 11.4, 10.27]][[13.96, 14.07, 14.48]][[86.65, 86.65, 86.65]][13.8]
MSE(-DR):[[0.0, 0.0, -1.13]][[2.56, 2.67, 3.08]][[75.25, 75.25, 75.25]][2.4]
**
MC-based ATE = 2.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-12.5, -12.39, -10.66]][[-16.07, -15.99, -16.19]][[-2.7, -2.7, -2.7]][-2.7]
std:[[0.39, 0.4, 0.24]][[0.03, 0.01, 0.17]][[0.0, 0.0, 0.0]][0.0]
MSE:[[12.51, 12.4, 10.66]][[16.07, 15.99, 16.19]][[2.7, 2.7, 2.7]][2.7]
MSE(-DR):[[0.0, -0.11, -1.85]][[3.56, 3.48, 3.68]][[-9.81, -9.81, -9.81]][-9.81]
**
==============


time spent until now: 37.6 mins


---------------------------------------
[pattern_seed, T, sd_R] = [1, 144, 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
```