```
Last login: Mon Mar 30 11:15:52 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-215-134-165.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 Mon Mar 30 15:30:13 UTC 2020

  System load:  0.19               Processes:           211
  Usage of /:   56.9% of 15.45GB   Users logged in:     0
  Memory usage: 1%                 IP address for ens5: 172.31.9.80
  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

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


*** System restart required ***
Last login: Mon Mar 30 15:15:55 2020 from 107.13.161.147
ubuntu@ip-172-31-9-80:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1
ubuntu@ip-172-31-9-80:~$ python EC2.py
11:30, 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] = [672, 5, 10, 10, 0.2, 1, 1, 1e-05, False, True]


----------------------------------------
[pattern_seed, T, sd_R] = [0, 336, 10]

max(u_O) =  156.6
O_threshold = 100
Traceback (most recent call last):
  File "EC2.py", line 70, in <module>
    file = file, print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 40, in simu
    target_policy = simu_target_policy_pattern(l = l, u_O = u_O, threshold =  O_thre, print_flag = "all")
  File "/home/ubuntu/simu_DGP.py", line 115, in simu_target_policy_pattern
    if noise_ratio is not None:
NameError: name 'noise_ratio' is not defined
ubuntu@ip-172-31-9-80:~$ python EC2.py
11:30, 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] = [672, 5, 10, 10, 0.2, 1, 1, 1e-05, False, True]


----------------------------------------
[pattern_seed, T, sd_R] = [0, 336, 10]

max(u_O) =  156.6
O_threshold = 100
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

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 = 80
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 = 85
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 1 1

number of reward locations:  22
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 = 95
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 = 105
target policy:

1 1 1 1 1

0 1 0 0 1

0 1 1 0 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  16
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
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
```

```
Value of Behaviour policy:88.966
O_threshold = 100
MC for this TARGET:[95.526, 0.151]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.13, 0.08, -0.55]][[0.13, 0.0, -0.18]][[-95.53, -95.53, -95.53]][[-0.6, -6.56]]
std:[[0.59, 0.56, 0.14]][[0.28, 0.25, 0.36]][[0.0, 0.0, 0.0]][[0.11, 0.11]]
MSE:[[0.6, 0.57, 0.57]][[0.31, 0.25, 0.4]][[95.53, 95.53, 95.53]][[0.61, 6.56]]
MSE(-DR):[[0.0, -0.03, -0.03]][[-0.29, -0.35, -0.2]][[94.93, 94.93, 94.93]][[0.01, 5.96]]
==============


O_threshold = 80
MC for this TARGET:[92.478, 0.141]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.19, 0.11, 0.49]][[1.74, 1.71, 1.46]][[-92.48, -92.48, -92.48]][[0.42, -3.51]]
std:[[0.37, 0.35, 0.08]][[0.16, 0.17, 0.26]][[0.0, 0.0, 0.0]][[0.05, 0.11]]
MSE:[[0.42, 0.37, 0.5]][[1.75, 1.72, 1.48]][[92.48, 92.48, 92.48]][[0.42, 3.51]]
MSE(-DR):[[0.0, -0.05, 0.08]][[1.33, 1.3, 1.06]][[92.06, 92.06, 92.06]][[0.0, 3.09]]
******
MC-based ATE = -3.05
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.06, 0.03, 1.04]][[1.61, 1.7, 1.63]][[3.05, 3.05, 3.05]][1.02]
std:[[0.22, 0.21, 0.06]][[0.12, 0.07, 0.1]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.23, 0.21, 1.04]][[1.61, 1.7, 1.63]][[3.05, 3.05, 3.05]][1.02]
MSE(-DR):[[0.0, -0.02, 0.81]][[1.38, 1.47, 1.4]][[2.82, 2.82, 2.82]][0.79]
******
==============


O_threshold = 85
MC for this TARGET:[93.235, 0.143]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.3, 0.27, 0.23]][[1.4, 1.34, 1.1]][[-93.24, -93.24, -93.24]][[0.2, -4.27]]
std:[[0.85, 0.81, 0.27]][[0.14, 0.14, 0.23]][[0.0, 0.0, 0.0]][[0.23, 0.11]]
MSE:[[0.9, 0.85, 0.35]][[1.41, 1.35, 1.12]][[93.24, 93.24, 93.24]][[0.3, 4.27]]
MSE(-DR):[[0.0, -0.05, -0.55]][[0.51, 0.45, 0.22]][[92.34, 92.34, 92.34]][[-0.6, 3.37]]
better than DR_NO_MARL
MC-based ATE = -2.29
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.17, 0.18, 0.79]][[1.27, 1.34, 1.28]][[2.29, 2.29, 2.29]][0.8]
std:[[0.26, 0.25, 0.13]][[0.14, 0.1, 0.13]][[0.0, 0.0, 0.0]][0.12]
MSE:[[0.31, 0.31, 0.8]][[1.28, 1.34, 1.29]][[2.29, 2.29, 2.29]][0.81]
MSE(-DR):[[0.0, 0.0, 0.49]][[0.97, 1.03, 0.98]][[1.98, 1.98, 1.98]][0.5]
******
==============


O_threshold = 90
MC for this TARGET:[93.592, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.07, -0.09, 0.04]][[1.29, 1.19, 0.98]][[-93.59, -93.59, -93.59]][[0.02, -4.63]]
std:[[0.4, 0.41, 0.02]][[0.12, 0.11, 0.22]][[0.0, 0.0, 0.0]][[0.01, 0.11]]
MSE:[[0.41, 0.42, 0.04]][[1.3, 1.2, 1.0]][[93.59, 93.59, 93.59]][[0.02, 4.63]]
MSE(-DR):[[0.0, 0.01, -0.37]][[0.89, 0.79, 0.59]][[93.18, 93.18, 93.18]][[-0.39, 4.22]]
better than DR_NO_MARL
MC-based ATE = -1.93
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.2, -0.17, 0.6]][[1.15, 1.19, 1.16]][[1.93, 1.93, 1.93]][0.62]
std:[[0.19, 0.16, 0.16]][[0.17, 0.14, 0.14]][[0.0, 0.0, 0.0]][0.13]
MSE:[[0.28, 0.23, 0.62]][[1.16, 1.2, 1.17]][[1.93, 1.93, 1.93]][0.63]
MSE(-DR):[[0.0, -0.05, 0.34]][[0.88, 0.92, 0.89]][[1.65, 1.65, 1.65]][0.35]
******
==============


O_threshold = 95
MC for this TARGET:[93.592, 0.145]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.05, -0.09, 0.08]][[1.29, 1.19, 1.0]][[-93.59, -93.59, -93.59]][[0.04, -4.63]]
std:[[0.41, 0.41, 0.0]][[0.12, 0.11, 0.22]][[0.0, 0.0, 0.0]][[0.01, 0.11]]
MSE:[[0.41, 0.42, 0.08]][[1.3, 1.2, 1.02]][[93.59, 93.59, 93.59]][[0.04, 4.63]]
MSE(-DR):[[0.0, 0.01, -0.33]][[0.89, 0.79, 0.61]][[93.18, 93.18, 93.18]][[-0.37, 4.22]]
better than DR_NO_MARL
MC-based ATE = -1.93
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.18, -0.17, 0.64]][[1.16, 1.19, 1.17]][[1.93, 1.93, 1.93]][0.64]
std:[[0.18, 0.16, 0.14]][[0.16, 0.14, 0.14]][[0.0, 0.0, 0.0]][0.12]
MSE:[[0.25, 0.23, 0.66]][[1.17, 1.2, 1.18]][[1.93, 1.93, 1.93]][0.65]
MSE(-DR):[[0.0, -0.02, 0.41]][[0.92, 0.95, 0.93]][[1.68, 1.68, 1.68]][0.4]
******
==============


O_threshold = 105
MC for this TARGET:[95.728, 0.155]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.27, -0.35, -1.08]][[-0.19, -0.34, -0.52]][[-95.73, -95.73, -95.73]][[-1.15, -6.76]]
std:[[0.13, 0.17, 0.09]][[0.29, 0.26, 0.31]][[0.0, 0.0, 0.0]][[0.05, 0.11]]
MSE:[[0.3, 0.39, 1.08]][[0.35, 0.43, 0.61]][[95.73, 95.73, 95.73]][[1.15, 6.76]]
```

MSE(-DR):[[0.0, 0.09, 0.78]][[0.05, 0.13, 0.31]][[95.43, 95.43, 95.43]][[0.85, 6.46]]
******
MC-based ATE = 0.2
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.4, -0.43, -0.52]][[-0.32, -0.34, -0.34]][[-0.2, -0.2, -0.2]][-0.55]
std:[[0.72, 0.73, 0.05]][[0.01, 0.01, 0.05]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.82, 0.85, 0.52]][[0.32, 0.34, 0.34]][[0.2, 0.2, 0.2]][0.55]
MSE(-DR):[[0.0, 0.03, -0.3]][[-0.5, -0.48, -0.48]][[-0.62, -0.62, -0.62]][-0.27]
==============

O_threshold = 110
MC for this TARGET:[95.684, 0.16]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.02, -1.18, -1.59]][[-1.7, -1.85, -2.0]][[-95.68, -95.68, -95.68]][[-1.75, -6.72]]
std:[[0.11, 0.12, 0.22]][[0.23, 0.18, 0.2]][[0.0, 0.0, 0.0]][[0.23, 0.11]]
MSE:[[1.03, 1.19, 1.61]][[1.72, 1.86, 2.01]][[95.68, 95.68, 95.68]][[1.77, 6.72]]
MSE(-DR):[[0.0, 0.16, 0.58]][[0.69, 0.83, 0.98]][[94.65, 94.65, 94.65]][[0.74, 5.69]]
******
MC-based ATE = 0.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.15, -1.26, -1.04]][[-1.84, -1.85, -1.83]][[-0.16, -0.16, -0.16]][-1.15]
std:[[0.48, 0.44, 0.08]][[0.05, 0.07, 0.17]][[0.0, 0.0, 0.0]][0.12]
MSE:[[1.25, 1.33, 1.04]][[1.84, 1.85, 1.84]][[0.16, 0.16, 0.16]][1.16]
MSE(-DR):[[0.0, 0.08, -0.21]][[0.59, 0.6, 0.59]][[-1.09, -1.09, -1.09]][-0.09]
better than DR_NO_MARL
==============

time spent until now: 5.1 mins

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

max(u_O) = 156.6
O_threshold = 100
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

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 = 80
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 = 85
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 1 1

```
number of reward locations:  22
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 = 95
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 = 105
target policy:

1 1 1 1 1

0 1 0 0 1

0 1 1 0 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  16
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
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:89.038
O_threshold = 100
MC for this TARGET:[95.518, 0.103]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.3, -0.41, -0.72]][[-0.1, -0.2, -0.21]][[-95.52, -95.52, -95.52]][[-0.83, -6.48]]
std:[[0.3, 0.32, 0.34]][[0.05, 0.0, 0.01]][[0.0, 0.0, 0.0]][[0.33, 0.02]]
MSE:[[0.42, 0.52, 0.8]][[0.11, 0.2, 0.21]][[95.52, 95.52, 95.52]][[0.89, 6.48]]
MSE(-DR):[[0.0, 0.1, 0.38]][[-0.31, -0.22, -0.21]][[95.1, 95.1, 95.1]][[0.47, 6.06]]
==============


O_threshold = 80
MC for this TARGET:[92.474, 0.1]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.05, 0.99, 0.88]][[1.14, 1.11, 1.12]][[-92.47, -92.47, -92.47]][[0.82, -3.44]]
std:[[0.21, 0.24, 0.38]][[0.08, 0.05, 0.02]][[0.0, 0.0, 0.0]][[0.41, 0.02]]
MSE:[[1.07, 1.02, 0.96]][[1.14, 1.11, 1.12]][[92.47, 92.47, 92.47]][[0.92, 3.44]]
MSE(-DR):[[0.0, -0.05, -0.11]][[0.07, 0.04, 0.05]][[91.4, 91.4, 91.4]][[-0.15, 2.37]]
better than DR_NO_MARL
MC-based ATE = -3.04
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.35, 1.4, 1.6]][[1.24, 1.31, 1.33]][[3.04, 3.04, 3.04]][1.65]
std:[[0.51, 0.56, 0.04]][[0.03, 0.06, 0.03]][[0.0, 0.0, 0.0]][0.08]
MSE:[[1.44, 1.51, 1.6]][[1.24, 1.31, 1.33]][[3.04, 3.04, 3.04]][1.65]
MSE(-DR):[[0.0, 0.07, 0.16]][[-0.2, -0.13, -0.11]][[1.6, 1.6, 1.6]][0.21]
==============


O_threshold = 85
MC for this TARGET:[93.228, 0.101]
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.51, 0.43, 0.29]][[0.77, 0.71, 0.79]][[-93.23, -93.23, -93.23]][[0.2, -4.19]]
std:[[0.01, 0.0, 0.29]][[0.22, 0.18, 0.07]][[0.0, 0.0, 0.0]][[0.28, 0.02]]
MSE:[[0.51, 0.43, 0.41]][[0.8, 0.73, 0.79]][[93.23, 93.23, 93.23]][[0.34, 4.19]]
MSE(-DR):[[0.0, -0.08, -0.1]][[0.29, 0.22, 0.28]][[92.72, 92.72, 92.72]][[-0.17, 3.68]]
better than DR_NO_MARL
MC-based ATE = -2.29
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.81, 0.84, 1.01]][[0.87, 0.91, 1.0]][[2.29, 2.29, 2.29]][1.04]
std:[[0.32, 0.32, 0.05]][[0.17, 0.19, 0.06]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.87, 0.9, 1.01]][[0.89, 0.93, 1.0]][[2.29, 2.29, 2.29]][1.04]
MSE(-DR):[[0.0, 0.03, 0.14]][[0.02, 0.06, 0.13]][[1.42, 1.42, 1.42]][0.17]
******
==============


O_threshold = 90
MC for this TARGET:[93.585, 0.101]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.6, 0.49, 0.27]][[0.67, 0.61, 0.66]][[-93.58, -93.58, -93.58]][[0.15, -4.55]]
std:[[0.36, 0.34, 0.48]][[0.1, 0.07, 0.03]][[0.0, 0.0, 0.0]][[0.47, 0.02]]
MSE:[[0.7, 0.6, 0.55]][[0.68, 0.61, 0.66]][[93.58, 93.58, 93.58]][[0.49, 4.55]]
MSE(-DR):[[0.0, -0.1, -0.15]][[-0.02, -0.09, -0.04]][[92.88, 92.88, 92.88]][[-0.21, 3.85]]
MC-based ATE = -1.93
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.9, 0.9, 0.99]][[0.77, 0.81, 0.87]][[1.93, 1.93, 1.93]][0.98]
std:[[0.66, 0.66, 0.14]][[0.06, 0.07, 0.04]][[0.0, 0.0, 0.0]][0.13]
MSE:[[1.12, 1.12, 1.0]][[0.77, 0.81, 0.87]][[1.93, 1.93, 1.93]][0.99]
MSE(-DR):[[0.0, 0.0, -0.12]][[-0.35, -0.31, -0.25]][[0.81, 0.81, 0.81]][-0.13]
==============


O_threshold = 95
MC for this TARGET:[93.585, 0.101]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.61, 0.49, 0.27]][[0.67, 0.61, 0.66]][[-93.58, -93.58, -93.58]][[0.15, -4.55]]
std:[[0.36, 0.34, 0.45]][[0.11, 0.07, 0.02]][[0.0, 0.0, 0.0]][[0.43, 0.02]]
MSE:[[0.71, 0.6, 0.52]][[0.68, 0.61, 0.66]][[93.58, 93.58, 93.58]][[0.46, 4.55]]
MSE(-DR):[[0.0, -0.11, -0.19]][[-0.03, -0.1, -0.05]][[92.87, 92.87, 92.87]][[-0.25, 3.84]]
MC-based ATE = -1.93
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.91, 0.9, 0.99]][[0.77, 0.81, 0.87]][[1.93, 1.93, 1.93]][0.99]
std:[[0.66, 0.66, 0.1]][[0.07, 0.07, 0.03]][[0.0, 0.0, 0.0]][0.1]
MSE:[[1.12, 1.12, 1.0]][[0.77, 0.81, 0.87]][[1.93, 1.93, 1.93]][1.0]
MSE(-DR):[[0.0, 0.0, -0.12]][[-0.35, -0.31, -0.25]][[0.81, 0.81, 0.81]][-0.12]
==============


O_threshold = 105
MC for this TARGET:[95.718, 0.1]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.77, -0.86, -1.22]][[-0.49, -0.62, -0.63]][[-95.72, -95.72, -95.72]][[-1.3, -6.68]]
std:[[0.15, 0.11, 0.23]][[0.12, 0.06, 0.02]][[0.0, 0.0, 0.0]][[0.26, 0.02]]
MSE:[[0.78, 0.87, 1.24]][[0.5, 0.62, 0.63]][[95.72, 95.72, 95.72]][[1.33, 6.68]]
MSE(-DR):[[0.0, 0.09, 0.46]][[-0.28, -0.16, -0.15]][[94.94, 94.94, 94.94]][[0.55, 5.9]]
MC-based ATE = 0.2
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.47, -0.44, -0.5]][[-0.39, -0.42, -0.42]][[-0.2, -0.2, -0.2]][-0.47]
std:[[0.16, 0.21, 0.12]][[0.08, 0.06, 0.01]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.5, 0.49, 0.51]][[0.4, 0.42, 0.42]][[0.2, 0.2, 0.2]][0.48]
MSE(-DR):[[0.0, -0.01, 0.01]][[-0.1, -0.08, -0.08]][[-0.3, -0.3, -0.3]][-0.02]
==============


O_threshold = 110
MC for this TARGET:[95.66, 0.094]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.46, -1.53, -1.77]][[-1.91, -2.02, -2.14]][[-95.66, -95.66, -95.66]][[-1.84, -6.62]]
std:[[0.33, 0.31, 0.03]][[0.16, 0.11, 0.06]][[0.0, 0.0, 0.0]][[0.05, 0.02]]
MSE:[[1.5, 1.56, 1.77]][[1.92, 2.02, 2.14]][[95.66, 95.66, 95.66]][[1.84, 6.62]]
MSE(-DR):[[0.0, 0.06, 0.27]][[0.42, 0.52, 0.64]][[94.16, 94.16, 94.16]][[0.34, 5.12]]
******
MC-based ATE = 0.14
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.16, -1.11, -1.05]][[-1.8, -1.82, -1.93]][[-0.14, -0.14, -0.14]][-1.01]
std:[[0.02, 0.0, 0.31]][[0.11, 0.11, 0.05]][[0.0, 0.0, 0.0]][0.28]
MSE:[[1.16, 1.11, 1.09]][[1.8, 1.82, 1.93]][[0.14, 0.14, 0.14]][1.05]
MSE(-DR):[[0.0, -0.05, -0.07]][[0.64, 0.66, 0.77]][[-1.02, -1.02, -1.02]][-0.11]
better than DR_NO_MARL
==============


time spent until now: 10.9 mins

ubuntu@ip-172-31-9-80:~$ O_threshold = 85
O_threshold: command not found
ubuntu@ip-172-31-9-80:~$ MC for this TARGET:[93.228, 0.101]
MC: command not found
ubuntu@ip-172-31-9-80:~$    [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
```

```
-bash: [DR/QV/IS]: No such file or directory
-bash: [DR/QV/IS]_NO_MARL: No such file or directory
-bash: [DR/QV/IS]_NO_MF: No such file or directory
[DR2,: command not found
ubuntu@ip-172-31-9-80:~$ bias:[[0.51, 0.43^C
ubuntu@ip-172-31-9-80:~$ MC for this TARGET:[93.228, 0.101]
MC: command not found
ubuntu@ip-172-31-9-80:~$ python EC2.py
11:56, 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] = [672, 5, 10, 10, 0.2, 1, 1, 1e-05, False, True]


----------------------------------------
[pattern_seed, T, sd_R] = [0, 336, 10]

max(u_O) =  156.6
O_threshold = 100
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

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 = 80
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 = 85
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 1 1

number of reward locations:  22
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 = 95
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 = 105
target policy:

1 1 1 1 1

0 1 0 0 1

0 1 1 0 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  16
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
1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:74.824
O_threshold = 100
MC for this TARGET:[85.644, 0.131]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.78, -0.94, -2.59]][[0.87, 0.47, -0.02]][[-85.64, -85.64, -85.64]][[-2.75, -10.82]]
std:[[0.4, 0.41, 0.22]][[0.4, 0.39, 0.43]][[0.0, 0.0, 0.0]][[0.22, 0.13]]
MSE:[[0.88, 1.03, 2.6]][[0.96, 0.61, 0.43]][[85.64, 85.64, 85.64]][[2.76, 10.82]]
MSE(-DR):[[0.0, 0.15, 1.72]][[0.08, -0.27, -0.45]][[84.76, 84.76, 84.76]][[1.88, 9.94]]
******
==============


O_threshold = 80
MC for this TARGET:[83.932, 0.137]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.03, -0.11, 0.56]][[3.74, 3.39, 3.12]][[-83.93, -83.93, -83.93]][[0.49, -9.11]]
std:[[0.66, 0.57, 0.15]][[0.34, 0.35, 0.37]][[0.0, 0.0, 0.0]][[0.06, 0.13]]
MSE:[[0.66, 0.58, 0.58]][[3.76, 3.41, 3.14]][[83.93, 83.93, 83.93]][[0.49, 9.11]]
MSE(-DR):[[0.0, -0.08, -0.08]][[3.1, 2.75, 2.48]][[83.27, 83.27, 83.27]][[-0.17, 8.45]]
better than DR_NO_MARL
MC-based ATE = -1.71
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.75, 0.83, 3.16]][[2.87, 2.92, 3.14]][[1.71, 1.71, 1.71]][3.24]
std:[[0.25, 0.16, 0.07]][[0.06, 0.04, 0.06]][[0.0, 0.0, 0.0]][0.16]
MSE:[[0.79, 0.85, 3.16]][[2.87, 2.92, 3.14]][[1.71, 1.71, 1.71]][3.24]
MSE(-DR):[[0.0, 0.06, 2.37]][[2.08, 2.13, 2.35]][[0.92, 0.92, 0.92]][2.45]
******
==============


O_threshold = 85
MC for this TARGET:[82.792, 0.136]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.11, 0.05, 0.23]][[3.91, 3.55, 3.24]][[-82.79, -82.79, -82.79]][[0.18, -7.97]]
std:[[0.87, 0.77, 0.36]][[0.35, 0.36, 0.37]][[0.0, 0.0, 0.0]][[0.27, 0.13]]
MSE:[[0.88, 0.77, 0.43]][[3.93, 3.57, 3.26]][[82.79, 82.79, 82.79]][[0.32, 7.97]]
MSE(-DR):[[0.0, -0.11, -0.45]][[3.05, 2.69, 2.38]][[81.91, 81.91, 81.91]][[-0.56, 7.09]]
better than DR_NO_MARL
MC-based ATE = -2.85
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.89, 0.99, 2.82]][[3.04, 3.09, 3.26]][[2.85, 2.85, 2.85]][2.93]
std:[[0.46, 0.36, 0.14]][[0.05, 0.03, 0.07]][[0.0, 0.0, 0.0]][0.04]
MSE:[[1.0, 1.05, 2.82]][[3.04, 3.09, 3.26]][[2.85, 2.85, 2.85]][2.93]
MSE(-DR):[[0.0, 0.05, 1.82]][[2.04, 2.09, 2.26]][[1.85, 1.85, 1.85]][1.93]
******
==============


O_threshold = 90
```

```
MC for this TARGET:[82.098, 0.136]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.06, -0.08, 0.21]][[4.05, 3.7, 3.35]][[-82.1, -82.1, -82.1]][[0.18, -7.27]]
std:[[0.42, 0.35, 0.17]][[0.38, 0.39, 0.33]][[0.0, 0.0, 0.0]][[0.1, 0.13]]
MSE:[[0.42, 0.36, 0.27]][[4.07, 3.72, 3.37]][[82.1, 82.1, 82.1]][[0.21, 7.27]]
MSE(-DR):[[0.0, -0.06, -0.15]][[3.65, 3.3, 2.95]][[81.68, 81.68, 81.68]][[-0.21, 6.85]]
better than DR_NO_MARL
MC-based ATE = -3.55
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.72, 0.86, 2.8]][[3.18, 3.23, 3.37]][[3.55, 3.55, 3.55]][2.93]
std:[[0.02, 0.06, 0.05]][[0.02, 0.01, 0.1]][[0.0, 0.0, 0.0]][0.12]
MSE:[[0.72, 0.86, 2.8]][[3.18, 3.23, 3.37]][[3.55, 3.55, 3.55]][2.93]
MSE(-DR):[[0.0, 0.14, 2.08]][[2.46, 2.51, 2.65]][[2.83, 2.83, 2.83]][2.21]
******
==============


O_threshold = 95
MC for this TARGET:[82.098, 0.136]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.04, -0.08, 0.26]][[4.04, 3.7, 3.34]][[-82.1, -82.1, -82.1]][[0.21, -7.27]]
std:[[0.41, 0.35, 0.17]][[0.38, 0.39, 0.34]][[0.0, 0.0, 0.0]][[0.11, 0.13]]
MSE:[[0.41, 0.36, 0.31]][[4.06, 3.72, 3.36]][[82.1, 82.1, 82.1]][[0.24, 7.27]]
MSE(-DR):[[0.0, -0.05, -0.1]][[3.65, 3.31, 2.95]][[81.69, 81.69, 81.69]][[-0.17, 6.86]]
better than DR_NO_MARL
MC-based ATE = -3.55
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.74, 0.86, 2.85]][[3.17, 3.23, 3.36]][[3.55, 3.55, 3.55]][2.96]
std:[[0.0, 0.06, 0.05]][[0.02, 0.01, 0.09]][[0.0, 0.0, 0.0]][0.11]
MSE:[[0.74, 0.86, 2.85]][[3.17, 3.23, 3.36]][[3.55, 3.55, 3.55]][2.96]
MSE(-DR):[[0.0, 0.12, 2.11]][[2.43, 2.49, 2.62]][[2.81, 2.81, 2.81]][2.22]
******
==============


O_threshold = 105
MC for this TARGET:[85.872, 0.133]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.84, -2.03, -3.79]][[-0.64, -1.05, -1.55]][[-85.87, -85.87, -85.87]][[-3.97, -11.05]]
std:[[0.08, 0.05, 0.16]][[0.32, 0.34, 0.38]][[0.0, 0.0, 0.0]][[0.19, 0.13]]
MSE:[[1.84, 2.03, 3.79]][[0.72, 1.1, 1.6]][[85.87, 85.87, 85.87]][[3.97, 11.05]]
MSE(-DR):[[0.0, 0.19, 1.95]][[-1.12, -0.74, -0.24]][[84.03, 84.03, 84.03]][[2.13, 9.21]]
MC-based ATE = 0.23
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.06, -1.09, -1.2]][[-1.51, -1.52, -1.54]][[-0.23, -0.23, -0.23]][-1.22]
std:[[0.49, 0.46, 0.06]][[0.08, 0.06, 0.05]][[0.0, 0.0, 0.0]][0.03]
MSE:[[1.17, 1.18, 1.2]][[1.51, 1.52, 1.54]][[0.23, 0.23, 0.23]][1.22]
MSE(-DR):[[0.0, 0.01, 0.03]][[0.34, 0.35, 0.37]][[-0.94, -0.94, -0.94]][0.05]
******
==============


O_threshold = 110
MC for this TARGET:[83.161, 0.135]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.72, -2.89, -3.44]][[-2.9, -3.21, -3.6]][[-83.16, -83.16, -83.16]][[-3.61, -8.34]]
std:[[0.23, 0.2, 0.53]][[0.31, 0.31, 0.32]][[0.0, 0.0, 0.0]][[0.5, 0.13]]
MSE:[[2.73, 2.9, 3.48]][[2.92, 3.22, 3.61]][[83.16, 83.16, 83.16]][[3.64, 8.34]]
MSE(-DR):[[0.0, 0.17, 0.75]][[0.19, 0.49, 0.88]][[80.43, 80.43, 80.43]][[0.91, 5.61]]
******
MC-based ATE = -2.48
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.94, -1.96, -0.85]][[-3.77, -3.68, -3.58]][[2.48, 2.48, 2.48]][-0.86]
std:[[0.17, 0.21, 0.31]][[0.09, 0.08, 0.11]][[0.0, 0.0, 0.0]][0.27]
MSE:[[1.95, 1.97, 0.9]][[3.77, 3.68, 3.58]][[2.48, 2.48, 2.48]][0.9]
MSE(-DR):[[0.0, 0.02, -1.05]][[1.82, 1.73, 1.63]][[0.53, 0.53, 0.53]][-1.05]
better than DR_NO_MARL
==============


time spent until now: 4.8 mins


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

max(u_O) =  156.6
O_threshold = 100
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

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 = 80
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 = 85
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 1 1

number of reward locations:  22
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 = 95
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 = 105
target policy:

1 1 1 1 1

0 1 0 0 1

0 1 1 0 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  16
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
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:74.886
O_threshold = 100
MC for this TARGET:[85.629, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.13, -1.33, -2.65]][[0.81, 0.4, 0.07]][[-85.63, -85.63, -85.63]][[-2.86, -10.74]]
std:[[0.17, 0.16, 0.2]][[0.11, 0.05, 0.03]][[0.0, 0.0, 0.0]][[0.21, 0.07]]
MSE:[[1.14, 1.34, 2.66]][[0.82, 0.4, 0.08]][[85.63, 85.63, 85.63]][[2.87, 10.74]]
MSE(-DR):[[0.0, 0.2, 1.52]][[-0.32, -0.74, -1.06]][[84.49, 84.49, 84.49]][[1.73, 9.6]]
==============


O_threshold = 80
MC for this TARGET:[83.925, 0.091]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.62, 1.49, 1.1]][[3.31, 2.94, 2.96]][[-83.92, -83.92, -83.92]][[0.96, -9.04]]
std:[[0.2, 0.24, 0.23]][[0.18, 0.14, 0.0]][[0.0, 0.0, 0.0]][[0.2, 0.07]]
MSE:[[1.63, 1.51, 1.12]][[3.31, 2.94, 2.96]][[83.92, 83.92, 83.92]][[0.98, 9.04]]
MSE(-DR):[[0.0, -0.12, -0.51]][[1.68, 1.31, 1.33]][[82.29, 82.29, 82.29]][[-0.65, 7.41]]
better than DR_NO_MARL
MC-based ATE = -1.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.76, 2.82, 3.75]][[2.5, 2.54, 2.89]][[1.7, 1.7, 1.7]][3.82]
std:[[0.03, 0.07, 0.04]][[0.08, 0.09, 0.03]][[0.0, 0.0, 0.0]][0.01]
MSE:[[2.76, 2.82, 3.75]][[2.5, 2.54, 2.89]][[1.7, 1.7, 1.7]][3.82]
MSE(-DR):[[0.0, 0.06, 0.99]][[-0.26, -0.22, 0.13]][[-1.06, -1.06, -1.06]][1.06]
==============


O_threshold = 85
MC for this TARGET:[82.783, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.46, 1.29, 0.57]][[3.52, 3.15, 3.11]][[-82.78, -82.78, -82.78]][[0.41, -7.9]]
std:[[0.43, 0.49, 0.13]][[0.15, 0.12, 0.02]][[0.0, 0.0, 0.0]][[0.07, 0.07]]
MSE:[[1.52, 1.38, 0.58]][[3.52, 3.15, 3.11]][[82.78, 82.78, 82.78]][[0.42, 7.9]]
MSE(-DR):[[0.0, -0.14, -0.94]][[2.0, 1.63, 1.59]][[81.26, 81.26, 81.26]][[-1.1, 6.38]]
better than DR_NO_MARL
MC-based ATE = -2.85
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.59, 2.63, 3.23]][[2.71, 2.75, 3.04]][[2.85, 2.85, 2.85]][3.26]
std:[[0.26, 0.32, 0.07]][[0.04, 0.06, 0.04]][[0.0, 0.0, 0.0]][0.14]
MSE:[[2.6, 2.65, 3.23]][[2.71, 2.75, 3.04]][[2.85, 2.85, 2.85]][3.26]
MSE(-DR):[[0.0, 0.05, 0.63]][[0.11, 0.15, 0.44]][[0.25, 0.25, 0.25]][0.66]
******
==============


O_threshold = 90
MC for this TARGET:[82.087, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.75, 1.58, 0.69]][[3.65, 3.28, 3.24]][[-82.09, -82.09, -82.09]][[0.52, -7.2]]
std:[[0.1, 0.15, 0.26]][[0.08, 0.06, 0.08]][[0.0, 0.0, 0.0]][[0.21, 0.07]]
MSE:[[1.75, 1.59, 0.74]][[3.65, 3.28, 3.24]][[82.09, 82.09, 82.09]][[0.56, 7.2]]
MSE(-DR):[[0.0, -0.16, -1.01]][[1.9, 1.53, 1.49]][[80.34, 80.34, 80.34]][[-1.19, 5.45]]
better than DR_NO_MARL
MC-based ATE = -3.54
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.88, 2.92, 3.35]][[2.85, 2.88, 3.17]][[3.54, 3.54, 3.54]][3.38]
std:[[0.07, 0.01, 0.07]][[0.03, 0.01, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.88, 2.92, 3.35]][[2.85, 2.88, 3.17]][[3.54, 3.54, 3.54]][3.38]
MSE(-DR):[[0.0, 0.04, 0.47]][[-0.03, 0.0, 0.29]][[0.66, 0.66, 0.66]][0.5]
==============


O_threshold = 95
MC for this TARGET:[82.087, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.76, 1.58, 0.7]][[3.65, 3.28, 3.26]][[-82.09, -82.09, -82.09]][[0.53, -7.2]]
std:[[0.13, 0.15, 0.23]][[0.08, 0.06, 0.08]][[0.0, 0.0, 0.0]][[0.21, 0.07]]
MSE:[[1.76, 1.59, 0.74]][[3.65, 3.28, 3.26]][[82.09, 82.09, 82.09]][[0.57, 7.2]]
MSE(-DR):[[0.0, -0.17, -1.02]][[1.89, 1.52, 1.5]][[80.33, 80.33, 80.33]][[-1.19, 5.44]]
better than DR_NO_MARL
MC-based ATE = -3.54
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.89, 2.92, 3.36]][[2.85, 2.88, 3.19]][[3.54, 3.54, 3.54]][3.38]
std:[[0.04, 0.01, 0.04]][[0.03, 0.01, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.89, 2.92, 3.36]][[2.85, 2.88, 3.19]][[3.54, 3.54, 3.54]][3.38]
MSE(-DR):[[0.0, 0.03, 0.47]][[-0.04, -0.01, 0.3]][[0.65, 0.65, 0.65]][0.49]
==============
```

```
O_threshold = 105
MC for this TARGET:[85.861, 0.084]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.95, -2.13, -3.75]][[-0.83, -1.23, -1.58]][[-85.86, -85.86, -85.86]][[-3.93, -10.97]]
std:[[0.16, 0.12, 0.23]][[0.19, 0.11, 0.05]][[0.0, 0.0, 0.0]][[0.27, 0.07]]
MSE:[[1.96, 2.13, 3.76]][[0.85, 1.23, 1.58]][[85.86, 85.86, 85.86]][[3.94, 10.97]]
MSE(-DR):[[0.0, 0.17, 1.8]][[-1.11, -0.73, -0.38]][[83.9, 83.9, 83.9]][[1.98, 9.01]]
MC-based ATE = 0.23
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.82, -0.8, -1.09]][[-1.63, -1.63, -1.65]][[-0.23, -0.23, -0.23]][-1.08]
std:[[0.01, 0.04, 0.04]][[0.08, 0.06, 0.02]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.82, 0.8, 1.09]][[1.63, 1.63, 1.65]][[0.23, 0.23, 0.23]][1.08]
MSE(-DR):[[0.0, -0.02, 0.27]][[0.81, 0.81, 0.83]][[-0.59, -0.59, -0.59]][0.26]
******
==============


O_threshold = 110
MC for this TARGET:[83.145, 0.082]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.36, -2.46, -3.36]][[-2.96, -3.25, -3.68]][[-83.14, -83.14, -83.14]][[-3.47, -8.26]]
std:[[0.45, 0.4, 0.1]][[0.11, 0.03, 0.05]][[0.0, 0.0, 0.0]][[0.15, 0.07]]
MSE:[[2.4, 2.49, 3.36]][[2.96, 3.25, 3.68]][[83.14, 83.14, 83.14]][[3.47, 8.26]]
MSE(-DR):[[0.0, 0.09, 0.96]][[0.56, 0.85, 1.28]][[80.74, 80.74, 80.74]][[1.07, 5.86]]
******
MC-based ATE = -2.48
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.22, -1.13, -0.71]][[-3.77, -3.65, -3.75]][[2.48, 2.48, 2.48]][-0.62]
std:[[0.28, 0.24, 0.1]][[0.0, 0.03, 0.02]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.25, 1.16, 0.72]][[3.77, 3.65, 3.75]][[2.48, 2.48, 2.48]][0.62]
MSE(-DR):[[0.0, -0.09, -0.53]][[2.52, 2.4, 2.5]][[1.23, 1.23, 1.23]][-0.63]
better than DR_NO_MARL
==============


time spent until now: 10.2 mins

ubuntu@ip-172-31-9-80:~$ python EC2.py
12:10, 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] = [672, 5, 10, 10, 0.2, 1, 1, 1e-05, False, True]


----------------------------------------
[pattern_seed, T, sd_R] = [0, 336, 10]

max(u_O) =  156.6
O_threshold = 100
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

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 = 80
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 = 85
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 1 1

number of reward locations:  22
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 = 95
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 = 105
target policy:

1 1 1 1 1

0 1 0 0 1

0 1 1 0 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  16
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
1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:70.046
O_threshold = 100
MC for this TARGET:[78.869, 0.119]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.18, -0.36, -1.6]][[2.21, 1.74, 1.42]][[-78.87, -78.87, -78.87]][[-1.78, -8.82]]
std:[[0.29, 0.27, 0.07]][[0.47, 0.42, 0.46]][[0.0, 0.0, 0.0]][[0.05, 0.1]]
MSE:[[0.34, 0.45, 1.6]][[2.26, 1.79, 1.49]][[78.87, 78.87, 78.87]][[1.78, 8.82]]
MSE(-DR):[[0.0, 0.11, 1.26]][[1.92, 1.45, 1.15]][[78.53, 78.53, 78.53]][[1.44, 8.48]]
******
==============


O_threshold = 80
MC for this TARGET:[77.828, 0.129]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.51, 1.43, 2.16]][[6.59, 6.1, 5.86]][[-77.83, -77.83, -77.83]][[2.09, -7.78]]
std:[[0.44, 0.4, 0.22]][[0.5, 0.46, 0.47]][[0.0, 0.0, 0.0]][[0.17, 0.1]]
```

```
MSE:[[1.57, 1.48, 2.17]][[6.61, 6.12, 5.88]][[77.83, 77.83, 77.83]][[2.1, 7.78]]
MSE(-DR):[[0.0, -0.09, 0.6]][[5.04, 4.55, 4.31]][[76.26, 76.26, 76.26]][[0.53, 6.21]]
******
MC-based ATE = -1.04
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.69, 1.79, 3.76]][[4.37, 4.36, 4.44]][[1.04, 1.04, 1.04]][3.87]
std:[[0.15, 0.13, 0.14]][[0.03, 0.03, 0.0]][[0.0, 0.0, 0.0]][0.12]
MSE:[[1.7, 1.79, 3.76]][[4.37, 4.36, 4.44]][[1.04, 1.04, 1.04]][3.87]
MSE(-DR):[[0.0, 0.09, 2.06]][[2.67, 2.66, 2.74]][[-0.66, -0.66, -0.66]][2.17]
******
==============


O_threshold = 85
MC for this TARGET:[76.71, 0.126]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.13, 1.04, 1.57]][[6.13, 5.66, 5.41]][[-76.71, -76.71, -76.71]][[1.47, -6.66]]
std:[[0.77, 0.66, 0.34]][[0.47, 0.44, 0.42]][[0.0, 0.0, 0.0]][[0.22, 0.1]]
MSE:[[1.37, 1.23, 1.61]][[6.15, 5.68, 5.43]][[76.71, 76.71, 76.71]][[1.49, 6.66]]
MSE(-DR):[[0.0, -0.14, 0.24]][[4.78, 4.31, 4.06]][[75.34, 75.34, 75.34]][[0.12, 5.29]]
******
MC-based ATE = -2.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.32, 1.39, 3.17]][[3.91, 3.92, 3.99]][[2.16, 2.16, 2.16]][3.25]
std:[[0.48, 0.39, 0.26]][[0.01, 0.02, 0.04]][[0.0, 0.0, 0.0]][0.17]
MSE:[[1.4, 1.44, 3.18]][[3.91, 3.92, 3.99]][[2.16, 2.16, 2.16]][3.25]
MSE(-DR):[[0.0, 0.04, 1.78]][[2.51, 2.52, 2.59]][[0.76, 0.76, 0.76]][1.85]
******
==============


O_threshold = 90
MC for this TARGET:[76.212, 0.126]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.57, 0.53, 1.01]][[5.73, 5.29, 5.05]][[-76.21, -76.21, -76.21]][[0.97, -6.17]]
std:[[0.35, 0.28, 0.14]][[0.49, 0.44, 0.42]][[0.0, 0.0, 0.0]][[0.07, 0.1]]
MSE:[[0.67, 0.6, 1.02]][[5.75, 5.31, 5.07]][[76.21, 76.21, 76.21]][[0.97, 6.17]]
MSE(-DR):[[0.0, -0.07, 0.35]][[5.08, 4.64, 4.4]][[75.54, 75.54, 75.54]][[0.3, 5.5]]
******
MC-based ATE = -2.66
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.75, 0.89, 2.61]][[3.52, 3.55, 3.63]][[2.66, 2.66, 2.66]][2.75]
std:[[0.06, 0.01, 0.07]][[0.02, 0.02, 0.05]][[0.0, 0.0, 0.0]][0.02]
MSE:[[0.75, 0.89, 2.61]][[3.52, 3.55, 3.63]][[2.66, 2.66, 2.66]][2.75]
MSE(-DR):[[0.0, 0.14, 1.86]][[2.77, 2.8, 2.88]][[1.91, 1.91, 1.91]][2.0]
******
==============


O_threshold = 95
MC for this TARGET:[76.212, 0.126]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.56, 0.53, 1.03]][[5.74, 5.29, 5.05]][[-76.21, -76.21, -76.21]][[1.0, -6.17]]
std:[[0.36, 0.28, 0.18]][[0.48, 0.44, 0.41]][[0.0, 0.0, 0.0]][[0.1, 0.1]]
MSE:[[0.67, 0.6, 1.05]][[5.76, 5.31, 5.07]][[76.21, 76.21, 76.21]][[1.0, 6.17]]
MSE(-DR):[[0.0, -0.07, 0.38]][[5.09, 4.64, 4.4]][[75.54, 75.54, 75.54]][[0.33, 5.5]]
******
MC-based ATE = -2.66
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.75, 0.89, 2.64]][[3.53, 3.55, 3.63]][[2.66, 2.66, 2.66]][2.78]
std:[[0.07, 0.01, 0.11]][[0.01, 0.02, 0.05]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.75, 0.89, 2.64]][[3.53, 3.55, 3.63]][[2.66, 2.66, 2.66]][2.78]
MSE(-DR):[[0.0, 0.14, 1.89]][[2.78, 2.8, 2.88]][[1.91, 1.91, 1.91]][2.03]
******
==============


O_threshold = 105
MC for this TARGET:[78.833, 0.12]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.87, -1.07, -2.47]][[0.43, -0.05, -0.34]][[-78.83, -78.83, -78.83]][[-2.67, -8.79]]
std:[[0.04, 0.02, 0.09]][[0.37, 0.33, 0.42]][[0.0, 0.0, 0.0]][[0.11, 0.1]]
MSE:[[0.87, 1.07, 2.47]][[0.57, 0.33, 0.54]][[78.83, 78.83, 78.83]][[2.67, 8.79]]
MSE(-DR):[[0.0, 0.2, 1.6]][[-0.3, -0.54, -0.33]][[77.96, 77.96, 77.96]][[1.8, 7.92]]
MC-based ATE = -0.04
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.69, -0.71, -0.87]][[-1.79, -1.78, -1.76]][[0.04, 0.04, 0.04]][-0.89]
std:[[0.33, 0.28, 0.01]][[0.1, 0.09, 0.04]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.76, 0.76, 0.87]][[1.79, 1.78, 1.76]][[0.04, 0.04, 0.04]][0.89]
MSE(-DR):[[0.0, 0.0, 0.11]][[1.03, 1.02, 1.0]][[-0.72, -0.72, -0.72]][0.13]
******
==============


O_threshold = 110
MC for this TARGET:[76.708, 0.126]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.87, -3.0, -3.38]][[-3.1, -3.39, -3.55]][[-76.71, -76.71, -76.71]][[-3.51, -6.66]]
std:[[0.34, 0.28, 0.48]][[0.35, 0.29, 0.31]][[0.0, 0.0, 0.0]][[0.43, 0.1]]
```

```
MSE:[[2.89, 3.01, 3.41]][[3.12, 3.4, 3.56]][[76.71, 76.71, 76.71]][[3.54, 6.66]]
MSE(-DR):[[0.0, 0.12, 0.52]][[0.23, 0.51, 0.67]][[73.82, 73.82, 73.82]][[0.65, 3.77]]
******
MC-based ATE = -2.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.69, -2.64, -1.78]][[-5.31, -5.13, -4.98]][[2.16, 2.16, 2.16]][-1.73]
std:[[0.04, 0.01, 0.41]][[0.12, 0.13, 0.15]][[0.0, 0.0, 0.0]][0.38]
MSE:[[2.69, 2.64, 1.83]][[5.31, 5.13, 4.98]][[2.16, 2.16, 2.16]][1.77]
MSE(-DR):[[0.0, -0.05, -0.86]][[2.62, 2.44, 2.29]][[-0.53, -0.53, -0.53]][-0.92]
better than DR_NO_MARL
==============


time spent until now: 4.8 mins


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

max(u_O) =  156.6
O_threshold = 100
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

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 = 80
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 = 85
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 1 1

number of reward locations:  22
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 = 95
```

```
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 = 105
target policy:

1 1 1 1 1

0 1 0 0 1

0 1 1 0 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  16
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
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:70.092
O_threshold = 100
MC for this TARGET:[78.856, 0.082]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.09, -0.3, -1.6]][[2.09, 1.63, 1.41]][[-78.86, -78.86, -78.86]][[-1.81, -8.76]]
std:[[0.18, 0.13, 0.17]][[0.05, 0.01, 0.0]][[0.0, 0.0, 0.0]][[0.22, 0.03]]
MSE:[[0.2, 0.33, 1.61]][[2.09, 1.63, 1.41]][[78.86, 78.86, 78.86]][[1.82, 8.76]]
MSE(-DR):[[0.0, 0.13, 1.41]][[1.89, 1.43, 1.21]][[78.66, 78.66, 78.66]][[1.62, 8.56]]
******
==============


O_threshold = 80
MC for this TARGET:[77.821, 0.082]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[3.41, 3.27, 2.59]][[6.08, 5.58, 5.65]][[-77.82, -77.82, -77.82]][[2.45, -7.73]]
std:[[0.23, 0.28, 0.3]][[0.16, 0.13, 0.02]][[0.0, 0.0, 0.0]][[0.25, 0.03]]
MSE:[[3.42, 3.28, 2.61]][[6.08, 5.58, 5.65]][[77.82, 77.82, 77.82]][[2.46, 7.73]]
MSE(-DR):[[0.0, -0.14, -0.81]][[2.66, 2.16, 2.23]][[74.4, 74.4, 74.4]][[-0.96, 4.31]]
better than DR_NO_MARL
MC-based ATE = -1.03
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[3.5, 3.57, 4.2]][[3.99, 3.95, 4.24]][[1.03, 1.03, 1.03]][4.26]
std:[[0.05, 0.14, 0.13]][[0.1, 0.11, 0.02]][[0.0, 0.0, 0.0]][0.03]
MSE:[[3.5, 3.57, 4.2]][[3.99, 3.95, 4.24]][[1.03, 1.03, 1.03]][4.26]
MSE(-DR):[[0.0, 0.07, 0.7]][[0.49, 0.45, 0.74]][[-2.47, -2.47, -2.47]][0.76]
******
==============


O_threshold = 85
MC for this TARGET:[76.706, 0.08]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[2.8, 2.67, 1.67]][[5.67, 5.18, 5.23]][[-76.71, -76.71, -76.71]][[1.55, -6.61]]
std:[[0.44, 0.49, 0.16]][[0.14, 0.1, 0.01]][[0.0, 0.0, 0.0]][[0.12, 0.03]]
MSE:[[2.83, 2.71, 1.68]][[5.67, 5.18, 5.23]][[76.71, 76.71, 76.71]][[1.55, 6.61]]
MSE(-DR):[[0.0, -0.12, -1.15]][[2.84, 2.35, 2.4]][[73.88, 73.88, 73.88]][[-1.28, 3.78]]
better than DR_NO_MARL
MC-based ATE = -2.15
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.89, 2.97, 3.27]][[3.57, 3.55, 3.82]][[2.15, 2.15, 2.15]][3.35]
std:[[0.26, 0.35, 0.01]][[0.09, 0.08, 0.01]][[0.0, 0.0, 0.0]][0.1]
MSE:[[2.9, 2.99, 3.27]][[3.57, 3.55, 3.82]][[2.15, 2.15, 2.15]][3.35]
MSE(-DR):[[0.0, 0.09, 0.37]][[0.67, 0.65, 0.92]][[-0.75, -0.75, -0.75]][0.45]
******
==============
```

```
O_threshold = 90
MC for this TARGET:[76.208, 0.079]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[2.68, 2.55, 1.42]][[5.3, 4.84, 4.9]][[-76.21, -76.21, -76.21]][[1.29, -6.12]]
std:[[0.16, 0.18, 0.25]][[0.05, 0.03, 0.07]][[0.0, 0.0, 0.0]][[0.22, 0.03]]
MSE:[[2.68, 2.56, 1.44]][[5.3, 4.84, 4.9]][[76.21, 76.21, 76.21]][[1.31, 6.12]]
MSE(-DR):[[0.0, -0.12, -1.24]][[2.62, 2.16, 2.22]][[73.53, 73.53, 73.53]][[-1.37, 3.44]]
better than DR_NO_MARL
MC-based ATE = -2.65
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.78, 2.84, 3.02]][[3.2, 3.22, 3.49]][[2.65, 2.65, 2.65]][3.09]
std:[[0.02, 0.04, 0.07]][[0.01, 0.01, 0.07]][[0.0, 0.0, 0.0]][0.01]
MSE:[[2.78, 2.84, 3.02]][[3.2, 3.22, 3.49]][[2.65, 2.65, 2.65]][3.09]
MSE(-DR):[[0.0, 0.06, 0.24]][[0.42, 0.44, 0.71]][[-0.13, -0.13, -0.13]][0.31]
******
==============


O_threshold = 95
MC for this TARGET:[76.208, 0.079]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[2.69, 2.55, 1.36]][[5.3, 4.84, 4.89]][[-76.21, -76.21, -76.21]][[1.22, -6.12]]
std:[[0.19, 0.18, 0.23]][[0.05, 0.03, 0.06]][[0.0, 0.0, 0.0]][[0.24, 0.03]]
MSE:[[2.7, 2.56, 1.38]][[5.3, 4.84, 4.89]][[76.21, 76.21, 76.21]][[1.24, 6.12]]
MSE(-DR):[[0.0, -0.14, -1.32]][[2.6, 2.14, 2.19]][[73.51, 73.51, 73.51]][[-1.46, 3.42]]
better than DR_NO_MARL
MC-based ATE = -2.65
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.78, 2.84, 2.97]][[3.21, 3.22, 3.48]][[2.65, 2.65, 2.65]][3.03]
std:[[0.01, 0.04, 0.06]][[0.0, 0.01, 0.06]][[0.0, 0.0, 0.0]][0.02]
MSE:[[2.78, 2.84, 2.97]][[3.21, 3.22, 3.48]][[2.65, 2.65, 2.65]][3.03]
MSE(-DR):[[0.0, 0.06, 0.19]][[0.43, 0.44, 0.7]][[-0.13, -0.13, -0.13]][0.25]
******
==============


O_threshold = 105
MC for this TARGET:[78.822, 0.08]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.82, -1.01, -2.44]][[0.18, -0.25, -0.46]][[-78.82, -78.82, -78.82]][[-2.63, -8.73]]
std:[[0.25, 0.19, 0.13]][[0.13, 0.06, 0.03]][[0.0, 0.0, 0.0]][[0.19, 0.03]]
MSE:[[0.86, 1.03, 2.44]][[0.22, 0.26, 0.46]][[78.82, 78.82, 78.82]][[2.64, 8.73]]
MSE(-DR):[[0.0, 0.17, 1.58]][[-0.64, -0.6, -0.4]][[77.96, 77.96, 77.96]][[1.78, 7.87]]
MC-based ATE = -0.03
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.73, -0.71, -0.83]][[-1.91, -1.87, -1.87]][[0.03, 0.03, 0.03]][-0.82]
std:[[0.07, 0.06, 0.04]][[0.07, 0.05, 0.03]][[0.0, 0.0, 0.0]][0.03]
MSE:[[0.73, 0.71, 0.83]][[1.91, 1.87, 1.87]][[0.03, 0.03, 0.03]][0.82]
MSE(-DR):[[0.0, -0.02, 0.1]][[1.18, 1.14, 1.14]][[-0.7, -0.7, -0.7]][0.09]
******
==============


O_threshold = 110
MC for this TARGET:[76.689, 0.079]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.71, -2.77, -3.4]][[-3.16, -3.41, -3.64]][[-76.69, -76.69, -76.69]][[-3.46, -6.6]]
std:[[0.38, 0.34, 0.09]][[0.04, 0.03, 0.02]][[0.0, 0.0, 0.0]][[0.12, 0.03]]
MSE:[[2.74, 2.79, 3.4]][[3.16, 3.41, 3.64]][[76.69, 76.69, 76.69]][[3.46, 6.6]]
MSE(-DR):[[0.0, 0.05, 0.66]][[0.42, 0.67, 0.9]][[73.95, 73.95, 73.95]][[0.72, 3.86]]
******
MC-based ATE = -2.17
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.62, -2.48, -1.8]][[-5.25, -5.04, -5.05]][[2.17, 2.17, 2.17]][-1.66]
std:[[0.2, 0.21, 0.09]][[0.01, 0.05, 0.03]][[0.0, 0.0, 0.0]][0.09]
MSE:[[2.63, 2.49, 1.8]][[5.25, 5.04, 5.05]][[2.17, 2.17, 2.17]][1.66]
MSE(-DR):[[0.0, -0.14, -0.83]][[2.62, 2.41, 2.42]][[-0.46, -0.46, -0.46]][-0.97]
better than DR_NO_MARL
==============


time spent until now: 10.2 mins

ubuntu@ip-172-31-9-80:~$
```