```
Last login: Mon Mar 30 09:40:18 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 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: Mon Mar 30 13:41:37 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
10:21, 03/30; num of cores:16

Basic setting:[sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, lam, simple, M_in_R] = [5, 5, 5, 0.2, 1, 1, 1e-05, True, True]


--------------------------------------
[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 = 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 = 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 = 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 w7
--------------------------------------
Value of Behaviour policy:90.868
O_threshold = 80
MC for this TARGET:[94.173, 0.105]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.7, 0.62, 0.86]][[1.83, 1.8, 1.66]][[-94.17, -94.17, -94.17]][[0.77, -3.31]]
std:[[0.15, 0.15, 0.08]][[0.07, 0.07, 0.09]][[0.0, 0.0, 0.0]][[0.07, 0.05]]
MSE:[[0.72, 0.64, 0.86]][[1.83, 1.8, 1.66]][[94.17, 94.17, 94.17]][[0.77, 3.31]]
MSE(-DR):[[0.0, -0.08, 0.14]][[1.11, 1.08, 0.94]][[93.45, 93.45, 93.45]][[0.05, 2.59]]
******
==============


O_threshold = 85
MC for this TARGET:[95.152, 0.106]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.78, 0.72, 0.71]][[1.28, 1.22, 1.13]][[-95.15, -95.15, -95.15]][[0.65, -4.28]]
std:[[0.25, 0.27, 0.01]][[0.06, 0.06, 0.11]][[0.0, 0.0, 0.0]][[0.02, 0.05]]
MSE:[[0.82, 0.77, 0.71]][[1.28, 1.22, 1.14]][[95.15, 95.15, 95.15]][[0.65, 4.28]]
MSE(-DR):[[0.0, -0.05, -0.11]][[0.46, 0.4, 0.32]][[94.33, 94.33, 94.33]][[-0.17, 3.46]]
better than DR_NO_MARL
MC-based ATE = 0.98
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.08, 0.11, -0.15]][[-0.55, -0.58, -0.53]][[-0.98, -0.98, -0.98]][-0.12]
std:[[0.1, 0.12, 0.07]][[0.01, 0.02, 0.02]][[0.0, 0.0, 0.0]][0.09]
```

```
MSE:[[0.13, 0.16, 0.17]][[0.55, 0.58, 0.53]][[0.98, 0.98, 0.98]][0.15]
MSE(-DR):[[0.0, 0.03, 0.04]][[0.42, 0.45, 0.4]][[0.85, 0.85, 0.85]][0.02]
******
==============


O_threshold = 90
MC for this TARGET:[95.681, 0.108]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.76, 0.69, 0.62]][[1.04, 0.95, 0.89]][[-95.68, -95.68, -95.68]][[0.55, -4.81]]
std:[[0.0, 0.02, 0.04]][[0.06, 0.05, 0.1]][[0.0, 0.0, 0.0]][[0.02, 0.05]]
MSE:[[0.76, 0.69, 0.62]][[1.04, 0.95, 0.9]][[95.68, 95.68, 95.68]][[0.55, 4.81]]
MSE(-DR):[[0.0, -0.07, -0.14]][[0.28, 0.19, 0.14]][[94.92, 94.92, 94.92]][[-0.21, 4.05]]
better than DR_NO_MARL
MC-based ATE = 1.51
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.06, 0.07, -0.23]][[-0.79, -0.85, -0.78]][[-1.51, -1.51, -1.51]][-0.22]
std:[[0.14, 0.13, 0.04]][[0.01, 0.03, 0.0]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.15, 0.15, 0.23]][[0.79, 0.85, 0.78]][[1.51, 1.51, 1.51]][0.23]
MSE(-DR):[[0.0, 0.0, 0.08]][[0.64, 0.7, 0.63]][[1.36, 1.36, 1.36]][0.08]
******
==============


O_threshold = 95
MC for this TARGET:[95.681, 0.108]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.75, 0.69, 0.62]][[1.03, 0.95, 0.87]][[-95.68, -95.68, -95.68]][[0.56, -4.81]]
std:[[0.01, 0.02, 0.07]][[0.06, 0.05, 0.1]][[0.0, 0.0, 0.0]][[0.03, 0.05]]
MSE:[[0.75, 0.69, 0.62]][[1.03, 0.95, 0.88]][[95.68, 95.68, 95.68]][[0.56, 4.81]]
MSE(-DR):[[0.0, -0.06, -0.13]][[0.28, 0.2, 0.13]][[94.93, 94.93, 94.93]][[-0.19, 4.06]]
better than DR_NO_MARL
MC-based ATE = 1.51
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.05, 0.07, -0.24]][[-0.8, -0.85, -0.79]][[-1.51, -1.51, -1.51]][-0.22]
std:[[0.15, 0.13, 0.01]][[0.02, 0.03, 0.01]][[0.0, 0.0, 0.0]][0.04]
MSE:[[0.16, 0.15, 0.24]][[0.8, 0.85, 0.79]][[1.51, 1.51, 1.51]][0.22]
MSE(-DR):[[0.0, -0.01, 0.08]][[0.64, 0.69, 0.63]][[1.35, 1.35, 1.35]][0.06]
******
==============


O_threshold = 100
MC for this TARGET:[96.88, 0.109]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.13, -0.16, -0.29]][[0.31, 0.19, 0.15]][[-96.88, -96.88, -96.88]][[-0.32, -6.01]]
std:[[0.35, 0.22, 0.25]][[0.13, 0.11, 0.19]][[0.0, 0.0, 0.0]][[0.13, 0.05]]
MSE:[[0.37, 0.27, 0.38]][[0.34, 0.22, 0.24]][[96.88, 96.88, 96.88]][[0.35, 6.01]]
MSE(-DR):[[0.0, -0.1, 0.01]][[-0.03, -0.15, -0.13]][[96.51, 96.51, 96.51]][[-0.02, 5.64]]
MC-based ATE = 2.71
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.84, -0.78, -1.15]][[-1.51, -1.62, -1.51]][[-2.71, -2.71, -2.71]][-1.09]
std:[[0.49, 0.38, 0.17]][[0.06, 0.04, 0.1]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.97, 0.87, 1.16]][[1.51, 1.62, 1.51]][[2.71, 2.71, 2.71]][1.09]
MSE(-DR):[[0.0, -0.1, 0.19]][[0.54, 0.65, 0.54]][[1.74, 1.74, 1.74]][0.12]
******
==============


O_threshold = 105
MC for this TARGET:[97.2, 0.113]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.69, -0.78, -0.74]][[-0.06, -0.19, -0.23]][[-97.2, -97.2, -97.2]][[-0.83, -6.33]]
std:[[0.54, 0.45, 0.21]][[0.11, 0.09, 0.16]][[0.0, 0.0, 0.0]][[0.11, 0.05]]
MSE:[[0.88, 0.9, 0.77]][[0.13, 0.21, 0.28]][[97.2, 97.2, 97.2]][[0.84, 6.33]]
MSE(-DR):[[0.0, 0.02, -0.11]][[-0.75, -0.67, -0.6]][[96.32, 96.32, 96.32]][[-0.04, 5.45]]
MC-based ATE = 3.03
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.39, -1.4, -1.6]][[-1.89, -1.99, -1.9]][[-3.03, -3.03, -3.03]][-1.6]
std:[[0.69, 0.6, 0.13]][[0.04, 0.02, 0.07]][[0.0, 0.0, 0.0]][0.04]
MSE:[[1.55, 1.52, 1.61]][[1.89, 1.99, 1.9]][[3.03, 3.03, 3.03]][1.6]
MSE(-DR):[[0.0, -0.03, 0.06]][[0.34, 0.44, 0.35]][[1.48, 1.48, 1.48]][0.05]
******
==============


O_threshold = 110
MC for this TARGET:[98.115, 0.116]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.34, -1.41, -2.08]][[-2.19, -2.35, -2.33]][[-98.12, -98.12, -98.12]][[-2.15, -7.25]]
std:[[0.21, 0.15, 0.0]][[0.05, 0.02, 0.04]][[0.0, 0.0, 0.0]][[0.06, 0.05]]
MSE:[[1.36, 1.42, 2.08]][[2.19, 2.35, 2.33]][[98.12, 98.12, 98.12]][[2.15, 7.25]]
MSE(-DR):[[0.0, 0.06, 0.72]][[0.83, 0.99, 0.97]][[96.76, 96.76, 96.76]][[0.79, 5.89]]
******
MC-based ATE = 3.94
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.04, -2.03, -2.94]][[-4.02, -4.15, -4.0]][[-3.94, -3.94, -3.94]][-2.92]
std:[[0.35, 0.3, 0.08]][[0.02, 0.06, 0.05]][[0.0, 0.0, 0.0]][0.13]
MSE:[[2.07, 2.05, 2.94]][[4.02, 4.15, 4.0]][[3.94, 3.94, 3.94]][2.92]
```

```
MSE(-DR):[[0.0, -0.02, 0.87]][[1.95, 2.08, 1.93]][[1.87, 1.87, 1.87]][0.85]
******
==============


time spent until now: 5.0 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 = 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 = 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 = 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:90.884
O_threshold = 80
MC for this TARGET:[94.179, 0.076]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.1, 1.07, 0.79]][[1.44, 1.4, 1.44]][[-94.18, -94.18, -94.18]][[0.76, -3.3]]
std:[[0.32, 0.32, 0.13]][[0.09, 0.05, 0.06]][[0.0, 0.0, 0.0]][[0.13, 0.02]]
MSE:[[1.15, 1.12, 0.8]][[1.44, 1.4, 1.44]][[94.18, 94.18, 94.18]][[0.77, 3.3]]
MSE(-DR):[[0.0, -0.03, -0.35]][[0.29, 0.25, 0.29]][[93.03, 93.03, 93.03]][[-0.38, 2.15]]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[95.157, 0.079]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.23, 1.18, 0.8]][[0.92, 0.85, 0.93]][[-95.16, -95.16, -95.16]][[0.75, -4.27]]
std:[[0.15, 0.18, 0.06]][[0.22, 0.17, 0.15]][[0.0, 0.0, 0.0]][[0.09, 0.02]]
MSE:[[1.24, 1.19, 0.8]][[0.95, 0.87, 0.94]][[95.16, 95.16, 95.16]][[0.76, 4.27]]
MSE(-DR):[[0.0, -0.05, -0.44]][[-0.29, -0.37, -0.3]][[93.92, 93.92, 93.92]][[-0.48, 3.03]]
MC-based ATE = 0.98
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.14, 0.11, 0.01]][[-0.52, -0.55, -0.52]][[-0.98, -0.98, -0.98]][-0.01]
std:[[0.16, 0.14, 0.07]][[0.13, 0.12, 0.09]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.21, 0.18, 0.07]][[0.54, 0.56, 0.53]][[0.98, 0.98, 0.98]][0.05]
MSE(-DR):[[0.0, -0.03, -0.14]][[0.33, 0.35, 0.32]][[0.77, 0.77, 0.77]][-0.16]
better than DR_NO_MARL
==============


O_threshold = 90
MC for this TARGET:[95.687, 0.08]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.1, 1.02, 0.81]][[0.69, 0.61, 0.67]][[-95.69, -95.69, -95.69]][[0.73, -4.8]]
std:[[0.25, 0.26, 0.21]][[0.12, 0.07, 0.08]][[0.0, 0.0, 0.0]][[0.21, 0.02]]
MSE:[[1.13, 1.05, 0.84]][[0.7, 0.61, 0.67]][[95.69, 95.69, 95.69]][[0.76, 4.8]]
MSE(-DR):[[0.0, -0.08, -0.29]][[-0.43, -0.52, -0.46]][[94.56, 94.56, 94.56]][[-0.37, 3.67]]
MC-based ATE = 1.51
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.0, -0.06, 0.02]][[-0.75, -0.79, -0.78]][[-1.51, -1.51, -1.51]][-0.04]
std:[[0.06, 0.06, 0.08]][[0.03, 0.02, 0.02]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.06, 0.08, 0.08]][[0.75, 0.79, 0.78]][[1.51, 1.51, 1.51]][0.08]
MSE(-DR):[[0.0, 0.02, 0.02]][[0.69, 0.73, 0.72]][[1.45, 1.45, 1.45]][0.02]
******
==============


O_threshold = 95
MC for this TARGET:[95.687, 0.08]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.11, 1.02, 0.79]][[0.7, 0.61, 0.66]][[-95.69, -95.69, -95.69]][[0.7, -4.8]]
std:[[0.27, 0.26, 0.19]][[0.12, 0.07, 0.09]][[0.0, 0.0, 0.0]][[0.18, 0.02]]
MSE:[[1.14, 1.05, 0.81]][[0.71, 0.61, 0.67]][[95.69, 95.69, 95.69]][[0.72, 4.8]]
MSE(-DR):[[0.0, -0.09, -0.33]][[-0.43, -0.53, -0.47]][[94.55, 94.55, 94.55]][[-0.42, 3.66]]
MC-based ATE = 1.51
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.01, -0.06, 0.0]][[-0.75, -0.79, -0.78]][[-1.51, -1.51, -1.51]][-0.06]
std:[[0.05, 0.06, 0.05]][[0.03, 0.02, 0.03]][[0.0, 0.0, 0.0]][0.04]
MSE:[[0.05, 0.08, 0.05]][[0.75, 0.79, 0.78]][[1.51, 1.51, 1.51]][0.07]
MSE(-DR):[[0.0, 0.03, 0.0]][[0.7, 0.74, 0.73]][[1.46, 1.46, 1.46]][0.02]
******
==============


O_threshold = 100
MC for this TARGET:[96.882, 0.081]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.16, 0.05, -0.26]][[0.17, 0.05, 0.09]][[-96.88, -96.88, -96.88]][[-0.37, -6.0]]
std:[[0.1, 0.12, 0.05]][[0.09, 0.05, 0.1]][[0.0, 0.0, 0.0]][[0.07, 0.02]]
MSE:[[0.19, 0.13, 0.26]][[0.19, 0.07, 0.13]][[96.88, 96.88, 96.88]][[0.38, 6.0]]
MSE(-DR):[[0.0, -0.06, 0.07]][[0.0, -0.12, -0.06]][[96.69, 96.69, 96.69]][[0.19, 5.81]]
******
MC-based ATE = 2.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.94, -1.02, -1.05]][[-1.28, -1.35, -1.35]][[-2.7, -2.7, -2.7]][-1.13]
std:[[0.21, 0.2, 0.08]][[0.0, 0.0, 0.04]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.96, 1.04, 1.05]][[1.28, 1.35, 1.35]][[2.7, 2.7, 2.7]][1.13]
MSE(-DR):[[0.0, 0.08, 0.09]][[0.32, 0.39, 0.39]][[1.74, 1.74, 1.74]][0.17]
******
==============


O_threshold = 105
MC for this TARGET:[97.197, 0.079]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.31, -0.41, -0.74]][[-0.26, -0.4, -0.37]][[-97.2, -97.2, -97.2]][[-0.84, -6.31]]
std:[[0.22, 0.21, 0.14]][[0.1, 0.06, 0.08]][[0.0, 0.0, 0.0]][[0.12, 0.02]]
MSE:[[0.38, 0.46, 0.75]][[0.28, 0.4, 0.38]][[97.2, 97.2, 97.2]][[0.85, 6.31]]
MSE(-DR):[[0.0, 0.08, 0.37]][[-0.1, 0.02, 0.0]][[96.82, 96.82, 96.82]][[0.47, 5.93]]
MC-based ATE = 3.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.41, -1.49, -1.52]][[-1.71, -1.8, -1.82]][[-3.02, -3.02, -3.02]][-1.6]
std:[[0.09, 0.11, 0.0]][[0.01, 0.01, 0.02]][[0.0, 0.0, 0.0]][0.01]
MSE:[[1.41, 1.49, 1.52]][[1.71, 1.8, 1.82]][[3.02, 3.02, 3.02]][1.6]
MSE(-DR):[[0.0, 0.08, 0.11]][[0.3, 0.39, 0.41]][[1.61, 1.61, 1.61]][0.19]
******
==============


O_threshold = 110
MC for this TARGET:[98.101, 0.073]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.73, -1.8, -2.12]][[-2.31, -2.46, -2.49]][[-98.1, -98.1, -98.1]][[-2.19, -7.22]]
std:[[0.19, 0.2, 0.02]][[0.12, 0.07, 0.12]][[0.0, 0.0, 0.0]][[0.02, 0.02]]
MSE:[[1.74, 1.81, 2.12]][[2.31, 2.46, 2.49]][[98.1, 98.1, 98.1]][[2.19, 7.22]]
MSE(-DR):[[0.0, 0.07, 0.38]][[0.57, 0.72, 0.75]][[96.36, 96.36, 96.36]][[0.45, 5.48]]
******
MC-based ATE = 3.92
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.82, -2.87, -2.9]][[-3.75, -3.86, -3.94]][[-3.92, -3.92, -3.92]][-2.95]
std:[[0.51, 0.52, 0.11]][[0.03, 0.02, 0.06]][[0.0, 0.0, 0.0]][0.12]
MSE:[[2.87, 2.92, 2.9]][[3.75, 3.86, 3.94]][[3.92, 3.92, 3.92]][2.95]
MSE(-DR):[[0.0, 0.05, 0.03]][[0.88, 0.99, 1.07]][[1.05, 1.05, 1.05]][0.08]
******
==============


time spent until now: 10.6 mins

ubuntu@ip-172-31-9-80:~$ python EC2.py
10:33, 03/30; num of cores:16

Basic setting:[sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, lam, simple, M_in_R] = [5, 5, 5, 0.2, 1, 1, 1e-05, True, True]


--------------------------------------
[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 74.4 100.4 95.8 135.2 133.5

102.6 107.3 83.3 66.9 92.8 102.6
```

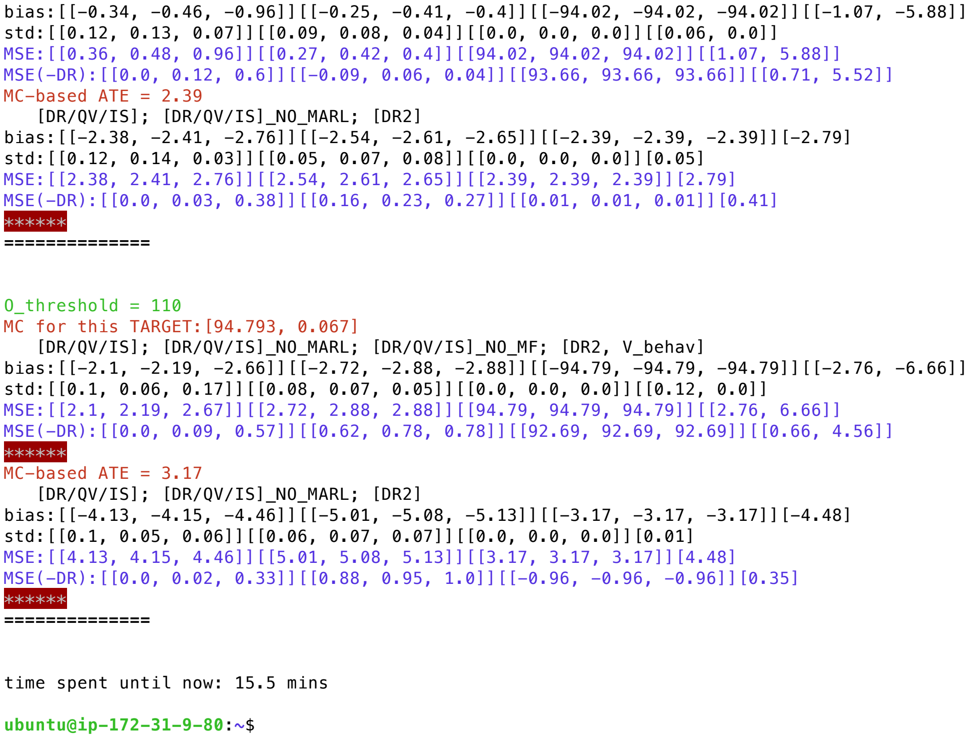
```
target policy:

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 0 1 1 1 1

1 1 1 0 1 1

number of reward locations:  33
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 0 1 1 1 1

1 1 0 0 1 1

number of reward locations:  30
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 0 1 1 1 1

1 1 0 0 1 1

number of reward locations:  29
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 0 1 1 1 1

1 1 0 0 0 1

number of reward locations:  28
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 0 1 0 1 1

1 1 0 0 0 1

number of reward locations:  24
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 0 0 0 1 1

0 1 0 0 0 0

number of reward locations:  19
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 0 0 0 1 1

0 0 0 0 0 0

number of reward locations:  13
1 2 3 4 5 6 7 1 2 3 4 5 6                       7
--------------------------------------
Value of Behaviour policy:88.004
O_threshold = 80
MC for this TARGET:[91.621, 0.093]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.87, 1.82, 1.28]][[1.97, 1.91, 1.78]][[-91.62, -91.62, -91.62]][[1.23, -3.62]]
std:[[0.04, 0.01, 0.1]][[0.03, 0.01, 0.01]][[0.0, 0.0, 0.0]][[0.08, 0.0]]
MSE:[[1.87, 1.82, 1.28]][[1.97, 1.91, 1.78]][[91.62, 91.62, 91.62]][[1.23, 3.62]]
MSE(-DR):[[0.0, -0.05, -0.59]][[0.1, 0.04, -0.09]][[89.75, 89.75, 89.75]][[-0.64, 1.75]]
==============


O_threshold = 85
MC for this TARGET:[92.408, 0.094]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.71, 1.63, 0.9]][[1.71, 1.61, 1.49]][[-92.41, -92.41, -92.41]][[0.82, -4.4]]
std:[[0.12, 0.09, 0.02]][[0.03, 0.02, 0.01]][[0.0, 0.0, 0.0]][[0.0, 0.0]]
MSE:[[1.71, 1.63, 0.9]][[1.71, 1.61, 1.49]][[92.41, 92.41, 92.41]][[0.82, 4.4]]
MSE(-DR):[[0.0, -0.08, -0.81]][[0.0, -0.1, -0.22]][[90.7, 90.7, 90.7]][[-0.89, 2.69]]
MC-based ATE = 0.79
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.17, -0.19, -0.38]][[-0.26, -0.3, -0.29]][[-0.79, -0.79, -0.79]][-0.41]
std:[[0.09, 0.08, 0.07]][[0.0, 0.01, 0.0]][[0.0, 0.0, 0.0]][0.08]
MSE:[[0.19, 0.21, 0.39]][[0.26, 0.3, 0.29]][[0.79, 0.79, 0.79]][0.42]
MSE(-DR):[[0.0, 0.02, 0.2]][[0.07, 0.11, 0.1]][[0.6, 0.6, 0.6]][0.23]
******
==============


O_threshold = 90
MC for this TARGET:[92.785, 0.092]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.45, 1.34, 0.59]][[1.39, 1.28, 1.18]][[-92.78, -92.78, -92.78]][[0.48, -4.78]]
std:[[0.15, 0.12, 0.08]][[0.01, 0.01, 0.01]][[0.0, 0.0, 0.0]][[0.05, 0.0]]
MSE:[[1.46, 1.35, 0.6]][[1.39, 1.28, 1.18]][[92.78, 92.78, 92.78]][[0.48, 4.78]]
MSE(-DR):[[0.0, -0.11, -0.86]][[-0.07, -0.18, -0.28]][[91.32, 91.32, 91.32]][[-0.98, 3.32]]
MC-based ATE = 1.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.42, -0.47, -0.69]][[-0.58, -0.63, -0.6]][[-1.16, -1.16, -1.16]][-0.74]
std:[[0.11, 0.1, 0.02]][[0.04, 0.02, 0.01]][[0.0, 0.0, 0.0]][0.03]
MSE:[[0.43, 0.48, 0.69]][[0.58, 0.63, 0.6]][[1.16, 1.16, 1.16]][0.74]
MSE(-DR):[[0.0, 0.05, 0.26]][[0.15, 0.2, 0.17]][[0.73, 0.73, 0.73]][0.31]
******
==============


O_threshold = 95
MC for this TARGET:[93.017, 0.093]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.42, 1.27, 0.52]][[1.3, 1.17, 1.09]][[-93.02, -93.02, -93.02]][[0.37, -5.01]]
std:[[0.15, 0.15, 0.08]][[0.01, 0.01, 0.03]][[0.0, 0.0, 0.0]][[0.08, 0.0]]
MSE:[[1.43, 1.28, 0.53]][[1.3, 1.17, 1.09]][[93.02, 93.02, 93.02]][[0.38, 5.01]]
MSE(-DR):[[0.0, -0.15, -0.9]][[-0.13, -0.26, -0.34]][[91.59, 91.59, 91.59]][[-1.05, 3.58]]
MC-based ATE = 1.4
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.45, -0.55, -0.76]][[-0.67, -0.74, -0.69]][[-1.4, -1.4, -1.4]][-0.86]
std:[[0.11, 0.13, 0.02]][[0.02, 0.01, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.46, 0.57, 0.76]][[0.67, 0.74, 0.69]][[1.4, 1.4, 1.4]][0.86]
MSE(-DR):[[0.0, 0.11, 0.3]][[0.21, 0.28, 0.23]][[0.94, 0.94, 0.94]][0.4]
******
==============


O_threshold = 100
MC for this TARGET:[93.844, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
```

```
bias:[[0.32, 0.13, -0.48]][[0.69, 0.53, 0.44]][[-93.84, -93.84, -93.84]][[-0.66, -5.84]]
std:[[0.13, 0.13, 0.11]][[0.04, 0.03, 0.07]][[0.0, 0.0, 0.0]][[0.12, 0.0]]
MSE:[[0.35, 0.18, 0.49]][[0.69, 0.53, 0.45]][[93.84, 93.84, 93.84]][[0.67, 5.84]]
MSE(-DR):[[0.0, -0.17, 0.14]][[0.34, 0.18, 0.1]][[93.49, 93.49, 93.49]][[0.32, 5.49]]
******
MC-based ATE = 2.22
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.56, -1.69, -1.76]][[-1.28, -1.38, -1.34]][[-2.22, -2.22, -2.22]][-1.89]
std:[[0.16, 0.15, 0.21]][[0.0, 0.02, 0.06]][[0.0, 0.0, 0.0]][0.19]
MSE:[[1.57, 1.7, 1.77]][[1.28, 1.38, 1.34]][[2.22, 2.22, 2.22]][1.9]
MSE(-DR):[[0.0, 0.13, 0.2]][[-0.29, -0.19, -0.23]][[0.65, 0.65, 0.65]][0.33]
==============


O_threshold = 105
MC for this TARGET:[94.022, 0.085]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.65, -0.78, -1.5]][[-0.29, -0.46, -0.5]][[-94.02, -94.02, -94.02]][[-1.63, -6.02]]
std:[[0.11, 0.16, 0.12]][[0.0, 0.0, 0.06]][[0.0, 0.0, 0.0]][[0.17, 0.0]]
MSE:[[0.66, 0.8, 1.5]][[0.29, 0.46, 0.5]][[94.02, 94.02, 94.02]][[1.64, 6.02]]
MSE(-DR):[[0.0, 0.14, 0.84]][[-0.37, -0.2, -0.16]][[93.36, 93.36, 93.36]][[0.98, 5.36]]
MC-based ATE = 2.4
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.53, -2.6, -2.78]][[-2.26, -2.37, -2.28]][[-2.4, -2.4, -2.4]][-2.86]
std:[[0.14, 0.17, 0.22]][[0.03, 0.01, 0.05]][[0.0, 0.0, 0.0]][0.24]
MSE:[[2.53, 2.61, 2.79]][[2.26, 2.37, 2.28]][[2.4, 2.4, 2.4]][2.87]
MSE(-DR):[[0.0, 0.08, 0.26]][[-0.27, -0.16, -0.25]][[-0.13, -0.13, -0.13]][0.34]
==============


O_threshold = 110
MC for this TARGET:[94.795, 0.085]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.92, -3.08, -2.78]][[-2.61, -2.79, -2.85]][[-94.8, -94.8, -94.8]][[-2.94, -6.79]]
std:[[0.45, 0.44, 0.25]][[0.05, 0.04, 0.08]][[0.0, 0.0, 0.0]][[0.24, 0.0]]
MSE:[[2.95, 3.11, 2.79]][[2.61, 2.79, 2.85]][[94.8, 94.8, 94.8]][[2.95, 6.79]]
MSE(-DR):[[0.0, 0.16, -0.16]][[-0.34, -0.16, -0.1]][[91.85, 91.85, 91.85]][[0.0, 3.84]]
MC-based ATE = 3.17
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-4.8, -4.89, -4.07]][[-4.58, -4.7, -4.63]][[-3.17, -3.17, -3.17]][-4.16]
std:[[0.48, 0.45, 0.34]][[0.02, 0.03, 0.07]][[0.0, 0.0, 0.0]][0.31]
MSE:[[4.82, 4.91, 4.08]][[4.58, 4.7, 4.63]][[3.17, 3.17, 3.17]][4.17]
MSE(-DR):[[0.0, 0.09, -0.74]][[-0.24, -0.12, -0.19]][[-1.65, -1.65, -1.65]][-0.65]
==============


time spent until now: 7.4 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 74.4 100.4 95.8 135.2 133.5

102.6 107.3 83.3 66.9 92.8 102.6

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 0 1 1 1 1

1 1 1 0 1 1

number of reward locations:  33
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 0 1 1 1 1

1 1 0 0 1 1

number of reward locations:  30
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 0 1 1 1 1

1 1 0 0 1 1

number of reward locations:  29
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 0 1 1 1 1

1 1 0 0 0 1

number of reward locations:  28
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 0 1 0 1 1

1 1 0 0 0 1

number of reward locations:  24
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 0 0 0 1 1

0 1 0 0 0 0

number of reward locations:  19
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 0 0 0 1 1

0 0 0 0 0 0
```

```
number of reward locations:  13
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:88.137
O_threshold = 80
MC for this TARGET:[91.626, 0.066]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[2.03, 1.95, 1.8]][[2.29, 2.19, 2.25]][[-91.63, -91.63, -91.63]][[1.72, -3.49]]
std:[[0.0, 0.01, 0.1]][[0.14, 0.14, 0.12]][[0.0, 0.0, 0.0]][[0.11, 0.0]]
MSE:[[2.03, 1.95, 1.8]][[2.29, 2.19, 2.25]][[91.63, 91.63, 91.63]][[1.72, 3.49]]
MSE(-DR):[[0.0, -0.08, -0.23]][[0.26, 0.16, 0.22]][[89.6, 89.6, 89.6]][[-0.31, 1.46]]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[92.414, 0.064]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.62, 1.53, 1.3]][[1.85, 1.74, 1.8]][[-92.41, -92.41, -92.41]][[1.21, -4.28]]
std:[[0.14, 0.17, 0.17]][[0.15, 0.13, 0.1]][[0.0, 0.0, 0.0]][[0.2, 0.0]]
MSE:[[1.63, 1.54, 1.31]][[1.86, 1.74, 1.8]][[92.41, 92.41, 92.41]][[1.23, 4.28]]
MSE(-DR):[[0.0, -0.09, -0.32]][[0.23, 0.11, 0.17]][[90.78, 90.78, 90.78]][[-0.4, 2.65]]
better than DR_NO_MARL
MC-based ATE = 0.79
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.42, -0.43, -0.5]][[-0.44, -0.45, -0.45]][[-0.79, -0.79, -0.79]][-0.51]
std:[[0.14, 0.16, 0.07]][[0.0, 0.01, 0.02]][[0.0, 0.0, 0.0]][0.09]
MSE:[[0.44, 0.46, 0.5]][[0.44, 0.45, 0.45]][[0.79, 0.79, 0.79]][0.52]
MSE(-DR):[[0.0, 0.02, 0.06]][[0.0, 0.01, 0.01]][[0.35, 0.35, 0.35]][0.08]
******
==============


O_threshold = 90
MC for this TARGET:[92.79, 0.064]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.37, 1.29, 1.0]][[1.6, 1.48, 1.53]][[-92.79, -92.79, -92.79]][[0.91, -4.65]]
std:[[0.13, 0.16, 0.13]][[0.15, 0.12, 0.11]][[0.0, 0.0, 0.0]][[0.16, 0.0]]
MSE:[[1.38, 1.3, 1.01]][[1.61, 1.48, 1.53]][[92.79, 92.79, 92.79]][[0.92, 4.65]]
MSE(-DR):[[0.0, -0.08, -0.37]][[0.23, 0.1, 0.15]][[91.41, 91.41, 91.41]][[-0.46, 3.27]]
better than DR_NO_MARL
MC-based ATE = 1.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.66, -0.67, -0.8]][[-0.68, -0.71, -0.72]][[-1.16, -1.16, -1.16]][-0.81]
std:[[0.13, 0.15, 0.03]][[0.01, 0.02, 0.01]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.67, 0.69, 0.8]][[0.68, 0.71, 0.72]][[1.16, 1.16, 1.16]][0.81]
MSE(-DR):[[0.0, 0.02, 0.13]][[0.01, 0.04, 0.05]][[0.49, 0.49, 0.49]][0.14]
******
==============


O_threshold = 95
MC for this TARGET:[93.02, 0.064]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.24, 1.14, 0.87]][[1.44, 1.3, 1.35]][[-93.02, -93.02, -93.02]][[0.78, -4.88]]
std:[[0.17, 0.19, 0.13]][[0.15, 0.12, 0.12]][[0.0, 0.0, 0.0]][[0.15, 0.0]]
MSE:[[1.25, 1.16, 0.88]][[1.45, 1.31, 1.36]][[93.02, 93.02, 93.02]][[0.79, 4.88]]
MSE(-DR):[[0.0, -0.09, -0.37]][[0.2, 0.06, 0.11]][[91.77, 91.77, 91.77]][[-0.46, 3.63]]
better than DR_NO_MARL
MC-based ATE = 1.39
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.8, -0.81, -0.93]][[-0.85, -0.89, -0.9]][[-1.39, -1.39, -1.39]][-0.94]
std:[[0.17, 0.18, 0.03]][[0.01, 0.02, 0.01]][[0.0, 0.0, 0.0]][0.04]
MSE:[[0.82, 0.83, 0.93]][[0.85, 0.89, 0.9]][[1.39, 1.39, 1.39]][0.94]
MSE(-DR):[[0.0, 0.01, 0.11]][[0.03, 0.07, 0.08]][[0.57, 0.57, 0.57]][0.12]
******
==============


O_threshold = 100
MC for this TARGET:[93.841, 0.065]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.31, 0.19, -0.18]][[0.65, 0.51, 0.54]][[-93.84, -93.84, -93.84]][[-0.3, -5.7]]
std:[[0.13, 0.13, 0.13]][[0.15, 0.13, 0.08]][[0.0, 0.0, 0.0]][[0.13, 0.0]]
MSE:[[0.34, 0.23, 0.22]][[0.67, 0.53, 0.55]][[93.84, 93.84, 93.84]][[0.33, 5.7]]
MSE(-DR):[[0.0, -0.11, -0.12]][[0.33, 0.19, 0.21]][[93.5, 93.5, 93.5]][[-0.01, 5.36]]
better than DR_NO_MARL
MC-based ATE = 2.21
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.73, -1.77, -1.98]][[-1.63, -1.69, -1.71]][[-2.21, -2.21, -2.21]][-2.02]
std:[[0.13, 0.12, 0.03]][[0.0, 0.02, 0.04]][[0.0, 0.0, 0.0]][0.02]
MSE:[[1.73, 1.77, 1.98]][[1.63, 1.69, 1.71]][[2.21, 2.21, 2.21]][2.02]
MSE(-DR):[[0.0, 0.04, 0.25]][[-0.1, -0.04, -0.02]][[0.48, 0.48, 0.48]][0.29]
==============


O_threshold = 105
MC for this TARGET:[94.021, 0.068]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
```



```
bias:[[-0.34, -0.46, -0.96]][[-0.25, -0.41, -0.4]][[-94.02, -94.02, -94.02]][[-1.07, -5.88]]
std:[[0.12, 0.13, 0.07]][[0.09, 0.08, 0.04]][[0.0, 0.0, 0.0]][[0.06, 0.0]]
MSE:[[0.36, 0.48, 0.96]][[0.27, 0.42, 0.4]][[94.02, 94.02, 94.02]][[1.07, 5.88]]
MSE(-DR):[[0.0, 0.12, 0.6]][[-0.09, 0.06, 0.04]][[93.66, 93.66, 93.66]][[0.71, 5.52]]
MC-based ATE = 2.39
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.38, -2.41, -2.76]][[-2.54, -2.61, -2.65]][[-2.39, -2.39, -2.39]][-2.79]
std:[[0.12, 0.14, 0.03]][[0.05, 0.07, 0.08]][[0.0, 0.0, 0.0]][0.05]
MSE:[[2.38, 2.41, 2.76]][[2.54, 2.61, 2.65]][[2.39, 2.39, 2.39]][2.79]
MSE(-DR):[[0.0, 0.03, 0.38]][[0.16, 0.23, 0.27]][[0.01, 0.01, 0.01]][0.41]
******
==============


O_threshold = 110
MC for this TARGET:[94.793, 0.067]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.1, -2.19, -2.66]][[-2.72, -2.88, -2.88]][[-94.79, -94.79, -94.79]][[-2.76, -6.66]]
std:[[0.1, 0.06, 0.17]][[0.08, 0.07, 0.05]][[0.0, 0.0, 0.0]][[0.12, 0.0]]
MSE:[[2.1, 2.19, 2.67]][[2.72, 2.88, 2.88]][[94.79, 94.79, 94.79]][[2.76, 6.66]]
MSE(-DR):[[0.0, 0.09, 0.57]][[0.62, 0.78, 0.78]][[92.69, 92.69, 92.69]][[0.66, 4.56]]
******
MC-based ATE = 3.17
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-4.13, -4.15, -4.46]][[-5.01, -5.08, -5.13]][[-3.17, -3.17, -3.17]][-4.48]
std:[[0.1, 0.05, 0.06]][[0.06, 0.07, 0.07]][[0.0, 0.0, 0.0]][0.01]
MSE:[[4.13, 4.15, 4.46]][[5.01, 5.08, 5.13]][[3.17, 3.17, 3.17]][4.48]
MSE(-DR):[[0.0, 0.02, 0.33]][[0.88, 0.95, 1.0]][[-0.96, -0.96, -0.96]][0.35]
******
==============


time spent until now: 15.5 mins

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