```
Last login: Sun Mar 29 23:48:06 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
Warning: Permanently added the ED25519 host key for IP address '3.215.134.165' to the list of known hosts.
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 13:41:36 UTC 2020

  System load:  0.83               Processes:           223
  Usage of /:   55.5% 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

93 packages can be updated.
43 updates are security updates.


Last login: Thu Mar  5 21:23:34 2020 from 107.13.161.147
ubuntu@ip-172-31-9-80:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1ubuntu@ip-172-31-9-80:~$ python EC2.py
Traceback (most recent call last):
  File "EC2.py", line 5, in <module>
    from simu_funs import *
  File "/home/ubuntu/simu_funs.py", line 6, in <module>
    from simu_DGP import *
  File "/home/ubuntu/simu_DGP.py", line 40
    O = rpoisson(u_O, (T, N)).T
    ^
IndentationError: unexpected indent
ubuntu@ip-172-31-9-80:~$ python EC2.py
09:44, 03/30; num of cores:16

Basic setting:[sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, lam] = [5, 10, 10, 0.2, 1, 1, 0.0001]


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

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 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:89.038
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:[[0.49, 0.39, 0.9]][[1.12, 1.09, 1.12]][[-92.47, -92.47, -92.47]][[0.8, -3.44]]
std:[[0.32, 0.32, 0.4]][[0.05, 0.02, 0.02]][[0.0, 0.0, 0.0]][[0.4, 0.02]]
MSE:[[0.59, 0.5, 0.98]][[1.12, 1.09, 1.12]][[92.47, 92.47, 92.47]][[0.89, 3.44]]
MSE(-DR):[[0.0, -0.09, 0.39]][[0.53, 0.5, 0.53]][[91.88, 91.88, 91.88]][[0.3, 2.85]]
******
==============
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:[[0.17, 0.03, 0.31]][[0.77, 0.71, 0.78]][[-93.23, -93.23, -93.23]][[0.16, -4.19]]
std:[[0.14, 0.13, 0.26]][[0.18, 0.14, 0.08]][[0.0, 0.0, 0.0]][[0.25, 0.02]]
MSE:[[0.22, 0.13, 0.4]][[0.79, 0.72, 0.78]][[93.23, 93.23, 93.23]][[0.3, 4.19]]
MSE(-DR):[[0.0, -0.09, 0.18]][[0.57, 0.5, 0.56]][[93.01, 93.01, 93.01]][[0.08, 3.97]]
******
MC-based ATE = 0.75
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.32, -0.37, -0.59]][[-0.35, -0.38, -0.34]][[-0.75, -0.75, -0.75]][-0.64]
std:[[0.18, 0.19, 0.14]][[0.13, 0.12, 0.09]][[0.0, 0.0, 0.0]][0.15]
MSE:[[0.37, 0.42, 0.61]][[0.37, 0.4, 0.35]][[0.75, 0.75, 0.75]][0.66]
MSE(-DR):[[0.0, 0.05, 0.24]][[0.0, 0.03, -0.02]][[0.38, 0.38, 0.38]][0.29]
******
==============


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:[[0.24, 0.1, 0.24]][[0.67, 0.6, 0.66]][[-93.58, -93.58, -93.58]][[0.1, -4.55]]
std:[[0.52, 0.48, 0.48]][[0.07, 0.03, 0.02]][[0.0, 0.0, 0.0]][[0.45, 0.02]]
MSE:[[0.57, 0.49, 0.54]][[0.67, 0.6, 0.66]][[93.58, 93.58, 93.58]][[0.46, 4.55]]
MSE(-DR):[[0.0, -0.08, -0.03]][[0.1, 0.03, 0.09]][[93.01, 93.01, 93.01]][[-0.11, 3.98]]
better than DR_NO_MARL
MC-based ATE = 1.11
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.26, -0.29, -0.66]][[-0.45, -0.49, -0.46]][[-1.11, -1.11, -1.11]][-0.69]
std:[[0.2, 0.16, 0.08]][[0.02, 0.01, 0.0]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.33, 0.33, 0.66]][[0.45, 0.49, 0.46]][[1.11, 1.11, 1.11]][0.69]
MSE(-DR):[[0.0, 0.0, 0.33]][[0.12, 0.16, 0.13]][[0.78, 0.78, 0.78]][0.36]
******
==============


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:[[0.3, 0.1, 0.31]][[0.67, 0.6, 0.67]][[-93.58, -93.58, -93.58]][[0.11, -4.55]]
std:[[0.51, 0.48, 0.45]][[0.07, 0.03, 0.02]][[0.0, 0.0, 0.0]][[0.43, 0.02]]
MSE:[[0.59, 0.49, 0.55]][[0.67, 0.6, 0.67]][[93.58, 93.58, 93.58]][[0.44, 4.55]]
MSE(-DR):[[0.0, -0.1, -0.04]][[0.08, 0.01, 0.08]][[92.99, 92.99, 92.99]][[-0.15, 3.96]]
better than DR_NO_MARL
MC-based ATE = 1.11
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.19, -0.29, -0.59]][[-0.45, -0.49, -0.46]][[-1.11, -1.11, -1.11]][-0.69]
std:[[0.19, 0.16, 0.05]][[0.01, 0.01, 0.0]][[0.0, 0.0, 0.0]][0.03]
MSE:[[0.27, 0.33, 0.59]][[0.45, 0.49, 0.46]][[1.11, 1.11, 1.11]][0.69]
MSE(-DR):[[0.0, 0.06, 0.32]][[0.18, 0.22, 0.19]][[0.84, 0.84, 0.84]][0.42]
******
==============


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:[[-0.49, -0.67, -0.71]][[-0.1, -0.21, -0.21]][[-95.52, -95.52, -95.52]][[-0.88, -6.48]]
std:[[0.03, 0.03, 0.34]][[0.0, 0.06, 0.01]][[0.0, 0.0, 0.0]][[0.27, 0.02]]
MSE:[[0.49, 0.67, 0.79]][[0.1, 0.22, 0.21]][[95.52, 95.52, 95.52]][[0.92, 6.48]]
MSE(-DR):[[0.0, 0.18, 0.3]][[-0.39, -0.27, -0.28]][[95.03, 95.03, 95.03]][[0.43, 5.99]]
MC-based ATE = 3.04
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.99, -1.06, -1.61]][[-1.22, -1.3, -1.33]][[-3.04, -3.04, -3.04]][-1.68]
std:[[0.28, 0.35, 0.06]][[0.06, 0.08, 0.03]][[0.0, 0.0, 0.0]][0.13]
MSE:[[1.03, 1.12, 1.61]][[1.22, 1.3, 1.33]][[3.04, 3.04, 3.04]][1.69]
MSE(-DR):[[0.0, 0.09, 0.58]][[0.19, 0.27, 0.3]][[2.01, 2.01, 2.01]][0.66]
******
==============


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:[[-0.78, -0.91, -1.2]][[-0.5, -0.63, -0.64]][[-95.72, -95.72, -95.72]][[-1.33, -6.68]]
std:[[0.15, 0.13, 0.22]][[0.05, 0.0, 0.0]][[0.0, 0.0, 0.0]][[0.24, 0.02]]
MSE:[[0.79, 0.92, 1.22]][[0.5, 0.63, 0.64]][[95.72, 95.72, 95.72]][[1.35, 6.68]]
MSE(-DR):[[0.0, 0.13, 0.43]][[-0.29, -0.16, -0.15]][[94.93, 94.93, 94.93]][[0.56, 5.89]]
MC-based ATE = 3.24
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.28, -1.3, -2.1]][[-1.62, -1.72, -1.77]][[-3.24, -3.24, -3.24]][-2.12]
std:[[0.46, 0.45, 0.18]][[0.0, 0.02, 0.02]][[0.0, 0.0, 0.0]][0.16]
MSE:[[1.36, 1.38, 2.11]][[1.62, 1.72, 1.77]][[3.24, 3.24, 3.24]][2.13]
MSE(-DR):[[0.0, 0.02, 0.75]][[0.26, 0.36, 0.41]][[1.88, 1.88, 1.88]][0.77]
******
==============
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:[[-1.71, -1.79, -1.8]][[-1.9, -2.03, -2.13]][[-95.66, -95.66, -95.66]][[-1.88, -6.62]]
std:[[0.2, 0.19, 0.01]][[0.12, 0.05, 0.07]][[0.0, 0.0, 0.0]][[0.02, 0.02]]
MSE:[[1.72, 1.8, 1.8]][[1.9, 2.03, 2.13]][[95.66, 95.66, 95.66]][[1.88, 6.62]]
MSE(-DR):[[0.0, 0.08, 0.08]][[0.18, 0.31, 0.41]][[93.94, 93.94, 93.94]][[0.16, 4.9]]
******
MC-based ATE = 3.19
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.2, -2.18, -2.7]][[-3.02, -3.13, -3.26]][[-3.19, -3.19, -3.19]][-2.68]
std:[[0.52, 0.51, 0.39]][[0.06, 0.04, 0.09]][[0.0, 0.0, 0.0]][0.38]
MSE:[[2.26, 2.24, 2.73]][[3.02, 3.13, 3.26]][[3.19, 3.19, 3.19]][2.71]
MSE(-DR):[[0.0, -0.02, 0.47]][[0.76, 0.87, 1.0]][[0.93, 0.93, 0.93]][0.45]
******
==============


time spent until now: 5.9 mins


--------------------------------------
[pattern_seed, T, sd_R] = [1, 672, 10]

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 = 85
target policy:

1 1 1 0 1

0 1 1 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

number of reward locations:  19
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 = 95
target policy:

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1
```

```
0 1 0 1 1

0 1 1 1 1

number of reward locations:  13
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 = 105
target policy:

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1

number of reward locations:  11
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
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:79.85
O_threshold = 80
MC for this TARGET:[83.057, 0.087]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.86, 0.7, 1.1]][[1.68, 1.61, 1.55]][[-83.06, -83.06, -83.06]][[0.94, -3.21]]
std:[[0.64, 0.59, 0.52]][[0.06, 0.07, 0.12]][[0.0, 0.0, 0.0]][[0.46, 0.02]]
MSE:[[1.07, 0.92, 1.22]][[1.68, 1.61, 1.55]][[83.06, 83.06, 83.06]][[1.05, 3.21]]
MSE(-DR):[[0.0, -0.15, 0.15]][[0.61, 0.54, 0.48]][[81.99, 81.99, 81.99]][[-0.02, 2.14]]
******
==============


O_threshold = 85
MC for this TARGET:[84.118, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.15, -0.01, 0.16]][[1.12, 1.03, 0.94]][[-84.12, -84.12, -84.12]][[-0.01, -4.27]]
std:[[0.6, 0.55, 0.48]][[0.11, 0.12, 0.13]][[0.0, 0.0, 0.0]][[0.43, 0.02]]
MSE:[[0.62, 0.55, 0.51]][[1.13, 1.04, 0.95]][[84.12, 84.12, 84.12]][[0.43, 4.27]]
MSE(-DR):[[0.0, -0.07, -0.11]][[0.51, 0.42, 0.33]][[83.5, 83.5, 83.5]][[-0.19, 3.65]]
better than DR_NO_MARL
MC-based ATE = 1.06
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.71, -0.72, -0.95]][[-0.56, -0.58, -0.61]][[-1.06, -1.06, -1.06]][-0.95]
std:[[0.04, 0.04, 0.04]][[0.05, 0.05, 0.02]][[0.0, 0.0, 0.0]][0.04]
MSE:[[0.71, 0.72, 0.95]][[0.56, 0.58, 0.61]][[1.06, 1.06, 1.06]][0.95]
MSE(-DR):[[0.0, 0.01, 0.24]][[-0.15, -0.13, -0.1]][[0.35, 0.35, 0.35]][0.24]
==============


O_threshold = 90
MC for this TARGET:[85.079, 0.09]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.17, -0.32, -0.56]][[0.3, 0.18, 0.11]][[-85.08, -85.08, -85.08]][[-0.71, -5.23]]
std:[[0.39, 0.39, 0.33]][[0.08, 0.09, 0.1]][[0.0, 0.0, 0.0]][[0.33, 0.02]]
MSE:[[0.43, 0.5, 0.65]][[0.31, 0.2, 0.15]][[85.08, 85.08, 85.08]][[0.78, 5.23]]
MSE(-DR):[[0.0, 0.07, 0.22]][[-0.12, -0.23, -0.28]][[84.65, 84.65, 84.65]][[0.35, 4.8]]
MC-based ATE = 2.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.03, -1.02, -1.67]][[-1.39, -1.43, -1.44]][[-2.02, -2.02, -2.02]][-1.65]
std:[[0.25, 0.2, 0.19]][[0.03, 0.02, 0.02]][[0.0, 0.0, 0.0]][0.13]
```

```
MSE:[[1.06, 1.04, 1.68]][[1.39, 1.43, 1.44]][[2.02, 2.02, 2.02]][1.66]
MSE(-DR):[[0.0, -0.02, 0.62]][[0.33, 0.37, 0.38]][[0.96, 0.96, 0.96]][0.6]
******
==============


O_threshold = 95
MC for this TARGET:[86.658, 0.09]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.59, -1.73, -2.1]][[-1.29, -1.45, -1.57]][[-86.66, -86.66, -86.66]][[-2.24, -6.81]]
std:[[0.26, 0.32, 0.16]][[0.03, 0.01, 0.01]][[0.0, 0.0, 0.0]][[0.21, 0.02]]
MSE:[[1.61, 1.76, 2.11]][[1.29, 1.45, 1.57]][[86.66, 86.66, 86.66]][[2.25, 6.81]]
MSE(-DR):[[0.0, 0.15, 0.5]][[-0.32, -0.16, -0.04]][[85.05, 85.05, 85.05]][[0.64, 5.2]]
MC-based ATE = 3.6
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.45, -2.44, -3.2]][[-2.98, -3.06, -3.12]][[-3.6, -3.6, -3.6]][-3.18]
std:[[0.38, 0.27, 0.36]][[0.08, 0.08, 0.1]][[0.0, 0.0, 0.0]][0.25]
MSE:[[2.48, 2.45, 3.22]][[2.98, 3.06, 3.12]][[3.6, 3.6, 3.6]][3.19]
MSE(-DR):[[0.0, -0.03, 0.74]][[0.5, 0.58, 0.64]][[1.12, 1.12, 1.12]][0.71]
******
==============


O_threshold = 100
MC for this TARGET:[87.607, 0.092]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.34, -2.42, -2.47]][[-2.08, -2.26, -2.36]][[-87.61, -87.61, -87.61]][[-2.56, -7.76]]
std:[[0.2, 0.28, 0.14]][[0.04, 0.02, 0.01]][[0.0, 0.0, 0.0]][[0.22, 0.02]]
MSE:[[2.35, 2.44, 2.47]][[2.08, 2.26, 2.36]][[87.61, 87.61, 87.61]][[2.57, 7.76]]
MSE(-DR):[[0.0, 0.09, 0.12]][[-0.27, -0.09, 0.01]][[85.26, 85.26, 85.26]][[0.22, 5.41]]
MC-based ATE = 4.55
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.2, -3.13, -3.57]][[-3.76, -3.87, -3.91]][[-4.55, -4.55, -4.55]][-3.5]
std:[[0.44, 0.3, 0.38]][[0.09, 0.09, 0.11]][[0.0, 0.0, 0.0]][0.24]
MSE:[[3.23, 3.14, 3.59]][[3.76, 3.87, 3.91]][[4.55, 4.55, 4.55]][3.51]
MSE(-DR):[[0.0, -0.09, 0.36]][[0.53, 0.64, 0.68]][[1.32, 1.32, 1.32]][0.28]
******
==============


O_threshold = 105
MC for this TARGET:[88.386, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.64, -2.78, -2.86]][[-2.77, -2.97, -3.07]][[-88.39, -88.39, -88.39]][[-3.0, -8.54]]
std:[[0.28, 0.35, 0.06]][[0.14, 0.12, 0.07]][[0.0, 0.0, 0.0]][[0.13, 0.02]]
MSE:[[2.65, 2.8, 2.86]][[2.77, 2.97, 3.07]][[88.39, 88.39, 88.39]][[3.0, 8.54]]
MSE(-DR):[[0.0, 0.15, 0.21]][[0.12, 0.32, 0.42]][[85.74, 85.74, 85.74]][[0.35, 5.89]]
******
MC-based ATE = 5.33
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.5, -3.48, -3.96]][[-4.45, -4.58, -4.62]][[-5.33, -5.33, -5.33]][-3.94]
std:[[0.36, 0.23, 0.46]][[0.19, 0.19, 0.19]][[0.0, 0.0, 0.0]][0.33]
MSE:[[3.52, 3.49, 3.99]][[4.45, 4.58, 4.62]][[5.33, 5.33, 5.33]][3.95]
MSE(-DR):[[0.0, -0.03, 0.47]][[0.93, 1.06, 1.1]][[1.81, 1.81, 1.81]][0.43]
******
==============


O_threshold = 110
MC for this TARGET:[89.303, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.2, -3.33, -3.32]][[-3.71, -3.92, -4.03]][[-89.3, -89.3, -89.3]][[-3.45, -9.45]]
std:[[0.11, 0.16, 0.04]][[0.04, 0.03, 0.02]][[0.0, 0.0, 0.0]][[0.09, 0.02]]
MSE:[[3.2, 3.33, 3.32]][[3.71, 3.92, 4.03]][[89.3, 89.3, 89.3]][[3.45, 9.45]]
MSE(-DR):[[0.0, 0.13, 0.12]][[0.51, 0.72, 0.83]][[86.1, 86.1, 86.1]][[0.25, 6.25]]
******
MC-based ATE = 6.25
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-4.07, -4.04, -4.43]][[-5.39, -5.53, -5.58]][[-6.25, -6.25, -6.25]][-4.4]
std:[[0.53, 0.42, 0.48]][[0.09, 0.1, 0.14]][[0.0, 0.0, 0.0]][0.37]
MSE:[[4.1, 4.06, 4.46]][[5.39, 5.53, 5.58]][[6.25, 6.25, 6.25]][4.42]
MSE(-DR):[[0.0, -0.04, 0.36]][[1.29, 1.43, 1.48]][[2.15, 2.15, 2.15]][0.32]
******
==============


time spent until now: 11.6 mins


--------------------------------------
[pattern_seed, T, sd_R] = [2, 672, 10]

max(u_O) =  157.3
O_threshold = 80
means of Order:

91.5 98.4 64.9 138.1 69.5

84.1 110.0 77.6 80.5 82.9
```

```
111.1 157.3 100.3 79.6 110.8

88.3 99.1 125.8 85.7 99.7

83.5 96.4 104.7 81.6 93.0

target policy:

1 1 0 1 0

1 1 0 1 1

1 1 1 0 1

1 1 1 1 1

1 1 1 1 1

number of reward locations:  21
O_threshold = 85
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 90
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  14
O_threshold = 95
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 100
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations:  8
O_threshold = 105
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0
```

```
number of reward locations:  6
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:78.431
O_threshold = 80
MC for this TARGET:[82.566, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.59, 0.51, 0.24]][[1.42, 1.32, 1.3]][[-82.57, -82.57, -82.57]][[0.16, -4.13]]
std:[[0.2, 0.16, 0.34]][[0.05, 0.02, 0.01]][[0.0, 0.0, 0.0]][[0.3, 0.01]]
MSE:[[0.62, 0.53, 0.42]][[1.42, 1.32, 1.3]][[82.57, 82.57, 82.57]][[0.34, 4.13]]
MSE(-DR):[[0.0, -0.09, -0.2]][[0.8, 0.7, 0.68]][[81.95, 81.95, 81.95]][[-0.28, 3.51]]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[83.788, 0.098]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.5, -0.62, -1.07]][[-0.03, -0.14, -0.16]][[-83.79, -83.79, -83.79]][[-1.19, -5.36]]
std:[[0.13, 0.17, 0.21]][[0.03, 0.05, 0.07]][[0.0, 0.0, 0.0]][[0.25, 0.01]]
MSE:[[0.52, 0.64, 1.09]][[0.04, 0.15, 0.17]][[83.79, 83.79, 83.79]][[1.22, 5.36]]
MSE(-DR):[[0.0, 0.12, 0.57]][[-0.48, -0.37, -0.35]][[83.27, 83.27, 83.27]][[0.7, 4.84]]
MC-based ATE = 1.22
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.09, -1.13, -1.31]][[-1.45, -1.46, -1.46]][[-1.22, -1.22, -1.22]][-1.35]
std:[[0.07, 0.01, 0.13]][[0.08, 0.07, 0.08]][[0.0, 0.0, 0.0]][0.05]
MSE:[[1.09, 1.13, 1.32]][[1.45, 1.46, 1.46]][[1.22, 1.22, 1.22]][1.35]
MSE(-DR):[[0.0, 0.04, 0.23]][[0.36, 0.37, 0.37]][[0.13, 0.13, 0.13]][0.26]
******
==============


O_threshold = 90
MC for this TARGET:[84.575, 0.1]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.41, -0.51, -0.96]][[-0.65, -0.78, -0.8]][[-84.58, -84.58, -84.58]][[-1.07, -6.14]]
std:[[0.0, 0.03, 0.01]][[0.09, 0.11, 0.1]][[0.0, 0.0, 0.0]][[0.01, 0.01]]
MSE:[[0.41, 0.51, 0.96]][[0.66, 0.79, 0.81]][[84.58, 84.58, 84.58]][[1.07, 6.14]]
MSE(-DR):[[0.0, 0.1, 0.55]][[0.25, 0.38, 0.4]][[84.17, 84.17, 84.17]][[0.66, 5.73]]
******
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.0, -1.02, -1.2]][[-2.07, -2.1, -2.09]][[-2.01, -2.01, -2.01]][-1.23]
std:[[0.2, 0.18, 0.33]][[0.14, 0.13, 0.11]][[0.0, 0.0, 0.0]][0.31]
MSE:[[1.02, 1.04, 1.24]][[2.07, 2.1, 2.09]][[2.01, 2.01, 2.01]][1.27]
MSE(-DR):[[0.0, 0.02, 0.22]][[1.05, 1.08, 1.07]][[0.99, 0.99, 0.99]][0.25]
******
==============


O_threshold = 95
MC for this TARGET:[83.632, 0.099]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.64, -0.68, -0.9]][[-0.87, -0.97, -1.01]][[-83.63, -83.63, -83.63]][[-0.94, -5.2]]
std:[[0.05, 0.04, 0.03]][[0.05, 0.06, 0.05]][[0.0, 0.0, 0.0]][[0.05, 0.01]]
MSE:[[0.64, 0.68, 0.9]][[0.87, 0.97, 1.01]][[83.63, 83.63, 83.63]][[0.94, 5.2]]
MSE(-DR):[[0.0, 0.04, 0.26]][[0.23, 0.33, 0.37]][[82.99, 82.99, 82.99]][[0.3, 4.56]]
******
MC-based ATE = 1.07
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.23, -1.19, -1.14]][[-2.29, -2.29, -2.3]][[-1.07, -1.07, -1.07]][-1.1]
std:[[0.15, 0.12, 0.37]][[0.1, 0.08, 0.06]][[0.0, 0.0, 0.0]][0.35]
MSE:[[1.24, 1.2, 1.2]][[2.29, 2.29, 2.3]][[1.07, 1.07, 1.07]][1.15]
MSE(-DR):[[0.0, -0.04, -0.04]][[1.05, 1.05, 1.06]][[-0.17, -0.17, -0.17]][-0.09]
better than DR_NO_MARL
==============


O_threshold = 100
MC for this TARGET:[83.863, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.29, -1.39, -1.48]][[-2.59, -2.68, -2.79]][[-83.86, -83.86, -83.86]][[-1.57, -5.43]]
std:[[0.18, 0.16, 0.14]][[0.08, 0.09, 0.01]][[0.0, 0.0, 0.0]][[0.13, 0.01]]
MSE:[[1.3, 1.4, 1.49]][[2.59, 2.68, 2.79]][[83.86, 83.86, 83.86]][[1.58, 5.43]]
MSE(-DR):[[0.0, 0.1, 0.19]][[1.29, 1.38, 1.49]][[82.56, 82.56, 82.56]][[0.28, 4.13]]
```

```
******
MC-based ATE = 1.3
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.88, -1.9, -1.72]][[-4.01, -4.0, -4.08]][[-1.3, -1.3, -1.3]][-1.73]
std:[[0.02, 0.0, 0.2]][[0.13, 0.11, 0.02]][[0.0, 0.0, 0.0]][0.17]
MSE:[[1.88, 1.9, 1.73]][[4.01, 4.0, 4.08]][[1.3, 1.3, 1.3]][1.74]
MSE(-DR):[[0.0, 0.02, -0.15]][[2.13, 2.12, 2.2]][[-0.58, -0.58, -0.58]][-0.14]
better than DR_NO_MARL
==============


O_threshold = 105
MC for this TARGET:[83.95, 0.093]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.78, -2.84, -2.7]][[-3.79, -3.86, -4.02]][[-83.95, -83.95, -83.95]][[-2.76, -5.52]]
std:[[0.27, 0.25, 0.25]][[0.05, 0.07, 0.04]][[0.0, 0.0, 0.0]][[0.23, 0.01]]
MSE:[[2.79, 2.85, 2.71]][[3.79, 3.86, 4.02]][[83.95, 83.95, 83.95]][[2.77, 5.52]]
MSE(-DR):[[0.0, 0.06, -0.08]][[1.0, 1.07, 1.23]][[81.16, 81.16, 81.16]][[-0.02, 2.73]]
better than DR_NO_MARL
MC-based ATE = 1.38
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.37, -3.35, -2.94]][[-5.21, -5.18, -5.32]][[-1.38, -1.38, -1.38]][-2.92]
std:[[0.07, 0.09, 0.09]][[0.1, 0.09, 0.03]][[0.0, 0.0, 0.0]][0.07]
MSE:[[3.37, 3.35, 2.94]][[5.21, 5.18, 5.32]][[1.38, 1.38, 1.38]][2.92]
MSE(-DR):[[0.0, -0.02, -0.43]][[1.84, 1.81, 1.95]][[-1.99, -1.99, -1.99]][-0.45]
better than DR_NO_MARL
==============


O_threshold = 110
MC for this TARGET:[83.95, 0.093]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.77, -2.84, -2.68]][[-3.78, -3.86, -4.02]][[-83.95, -83.95, -83.95]][[-2.75, -5.52]]
std:[[0.27, 0.25, 0.27]][[0.05, 0.07, 0.04]][[0.0, 0.0, 0.0]][[0.25, 0.01]]
MSE:[[2.78, 2.85, 2.69]][[3.78, 3.86, 4.02]][[83.95, 83.95, 83.95]][[2.76, 5.52]]
MSE(-DR):[[0.0, 0.07, -0.09]][[1.0, 1.08, 1.24]][[81.17, 81.17, 81.17]][[-0.02, 2.74]]
better than DR_NO_MARL
MC-based ATE = 1.38
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.36, -3.35, -2.92]][[-5.2, -5.18, -5.31]][[-1.38, -1.38, -1.38]][-2.91]
std:[[0.07, 0.09, 0.07]][[0.11, 0.09, 0.03]][[0.0, 0.0, 0.0]][0.05]
MSE:[[3.36, 3.35, 2.92]][[5.2, 5.18, 5.31]][[1.38, 1.38, 1.38]][2.91]
MSE(-DR):[[0.0, -0.01, -0.44]][[1.84, 1.82, 1.95]][[-1.98, -1.98, -1.98]][-0.45]
better than DR_NO_MARL
==============


time spent until now: 17.3 mins


-------------------------------------
[pattern_seed, T, sd_R] = [3, 672, 10]

max(u_O) =  142.3
O_threshold = 80
means of Order:

142.3 108.6 101.4 68.5 94.1

92.7 97.9 87.8 98.6 90.4

76.5 118.7 118.7 140.0 100.5

91.7 89.2 73.0 121.1 79.8

78.5 95.5 133.9 104.3 81.1

target policy:

1 1 1 0 1

1 1 1 1 1

0 1 1 1 1

1 1 0 1 0

0 1 1 1 1

number of reward locations:  20
O_threshold = 85
target policy:

1 1 1 0 1

1 1 1 1 1

0 1 1 1 1
```

```
1 1 0 1 0

0 1 1 1 0

number of reward locations:  19
O_threshold = 90
target policy:

1 1 1 0 1

1 1 0 1 1

0 1 1 1 1

1 0 0 1 0

0 1 1 1 0

number of reward locations:  17
O_threshold = 95
target policy:

1 1 1 0 0

0 1 0 1 0

0 1 1 1 1

0 0 0 1 0

0 1 1 1 0

number of reward locations:  13
O_threshold = 100
target policy:

1 1 1 0 0

0 0 0 0 0

0 1 1 1 1

0 0 0 1 0

0 0 1 1 0

number of reward locations:  10
O_threshold = 105
target policy:

1 1 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 0 0

number of reward locations:  7
O_threshold = 110
target policy:

1 0 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 0 0

number of reward locations:  6
1 2 3 4
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:[[0.24, 0.1, 0.24]][[0.67, 0.6, 0.66]][[-93.58, -93.58, -93.58]][[0.1, -4.55]]
std:[[0.52, 0.48, 0.48]][[0.07, 0.03, 0.02]][[0.0, 0.0, 0.0]][[0.45, 0.02]]
MSE:[[0.57, 0.49, 0.54]][[0.67, 0.6, 0.66]][[93.58, 93.58, 93.58]][[0.46, 4.55]]
MSE(-DR):[[0.0, -0.08, -0.03]][[0.1, 0.03, 0.09]][[93.01, 93.01, 93.01]][[-0.11, 3.98]]
better than DR_NO_MARL
MC-based ATE = 1.11
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.26, -0.29, -0.66]][[-0.45, -0.49, -0.46]][[-1.11, -1.11, -1.11]][-0.69]
std:[[0.2, 0.16, 0.08]][[0.02, 0.01, 0.0]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.33, 0.33, 0.66]][[0.45, 0.49, 0.46]][[1.11, 1.11, 1.11]][0.69]
```

```
MSE(-DR):[[0.0, 0.0, 0.33]][[0.12, 0.16, 0.13]][[0.78, 0.78, 0.78]][0.36]
******
==============


5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:80.78
O_threshold = 80
MC for this TARGET:[85.399, 0.092]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.42, 0.37, 0.0]][[0.94, 0.86, 0.84]][[-85.4, -85.4, -85.4]][[-0.05, -4.62]]
std:[[0.12, 0.12, 0.1]][[0.1, 0.1, 0.03]][[0.0, 0.0, 0.0]][[0.11, 0.0]]
MSE:[[0.44, 0.39, 0.1]][[0.95, 0.87, 0.84]][[85.4, 85.4, 85.4]][[0.12, 4.62]]
MSE(-DR):[[0.0, -0.05, -0.34]][[0.51, 0.43, 0.4]][[84.96, 84.96, 84.96]][[-0.32, 4.18]]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[85.358, 0.092]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.35, 0.25, -0.22]][[0.68, 0.62, 0.58]][[-85.36, -85.36, -85.36]][[-0.32, -4.58]]
std:[[0.11, 0.15, 0.13]][[0.11, 0.11, 0.01]][[0.0, 0.0, 0.0]][[0.08, 0.0]]
MSE:[[0.37, 0.29, 0.26]][[0.69, 0.63, 0.58]][[85.36, 85.36, 85.36]][[0.33, 4.58]]
MSE(-DR):[[0.0, -0.08, -0.11]][[0.32, 0.26, 0.21]][[84.99, 84.99, 84.99]][[-0.04, 4.21]]
better than DR_NO_MARL
MC-based ATE = -0.04
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.07, -0.11, -0.22]][[-0.26, -0.24, -0.26]][[0.04, 0.04, 0.04]][-0.27]
std:[[0.01, 0.03, 0.02]][[0.01, 0.01, 0.01]][[0.0, 0.0, 0.0]][0.02]
MSE:[[0.07, 0.11, 0.22]][[0.26, 0.24, 0.26]][[0.04, 0.04, 0.04]][0.27]
MSE(-DR):[[0.0, 0.04, 0.15]][[0.19, 0.17, 0.19]][[-0.03, -0.03, -0.03]][0.2]
******
==============


O_threshold = 90
MC for this TARGET:[86.391, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.51, 0.36, -0.3]][[0.21, 0.1, 0.09]][[-86.39, -86.39, -86.39]][[-0.46, -5.61]]
std:[[0.26, 0.26, 0.16]][[0.0, 0.02, 0.11]][[0.0, 0.0, 0.0]][[0.15, 0.0]]
MSE:[[0.57, 0.44, 0.34]][[0.21, 0.1, 0.14]][[86.39, 86.39, 86.39]][[0.48, 5.61]]
MSE(-DR):[[0.0, -0.13, -0.23]][[-0.36, -0.47, -0.43]][[85.82, 85.82, 85.82]][[-0.09, 5.04]]
MC-based ATE = 0.99
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.09, -0.01, -0.3]][[-0.73, -0.76, -0.75]][[-0.99, -0.99, -0.99]][-0.41]
std:[[0.14, 0.14, 0.05]][[0.1, 0.08, 0.09]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.17, 0.14, 0.3]][[0.74, 0.76, 0.76]][[0.99, 0.99, 0.99]][0.41]
MSE(-DR):[[0.0, -0.03, 0.13]][[0.57, 0.59, 0.59]][[0.82, 0.82, 0.82]][0.24]
******
==============


O_threshold = 95
MC for this TARGET:[85.875, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.54, -0.65, -1.11]][[-0.55, -0.67, -0.64]][[-85.88, -85.88, -85.88]][[-1.22, -5.1]]
std:[[0.04, 0.07, 0.15]][[0.01, 0.01, 0.09]][[0.0, 0.0, 0.0]][[0.18, 0.0]]
MSE:[[0.54, 0.65, 1.12]][[0.55, 0.67, 0.65]][[85.88, 85.88, 85.88]][[1.23, 5.1]]
MSE(-DR):[[0.0, 0.11, 0.58]][[0.01, 0.13, 0.11]][[85.34, 85.34, 85.34]][[0.69, 4.56]]
******
MC-based ATE = 0.48
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.96, -1.02, -1.12]][[-1.49, -1.53, -1.48]][[-0.48, -0.48, -0.48]][-1.17]
std:[[0.16, 0.19, 0.04]][[0.11, 0.1, 0.07]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.97, 1.04, 1.12]][[1.49, 1.53, 1.48]][[0.48, 0.48, 0.48]][1.17]
MSE(-DR):[[0.0, 0.07, 0.15]][[0.52, 0.56, 0.51]][[-0.49, -0.49, -0.49]][0.2]
******
==============


O_threshold = 100
MC for this TARGET:[87.41, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.14, -1.23, -1.95]][[-2.14, -2.28, -2.29]][[-87.41, -87.41, -87.41]][[-2.04, -6.63]]
std:[[0.53, 0.57, 0.03]][[0.03, 0.05, 0.03]][[0.0, 0.0, 0.0]][[0.08, 0.0]]
MSE:[[1.26, 1.36, 1.95]][[2.14, 2.28, 2.29]][[87.41, 87.41, 87.41]][[2.04, 6.63]]
MSE(-DR):[[0.0, 0.1, 0.69]][[0.88, 1.02, 1.03]][[86.15, 86.15, 86.15]][[0.78, 5.37]]
******
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.56, -1.6, -1.95]][[-3.08, -3.15, -3.13]][[-2.01, -2.01, -2.01]][-1.99]
std:[[0.65, 0.69, 0.07]][[0.07, 0.05, 0.0]][[0.0, 0.0, 0.0]][0.03]
MSE:[[1.69, 1.74, 1.95]][[3.08, 3.15, 3.13]][[2.01, 2.01, 2.01]][1.99]
MSE(-DR):[[0.0, 0.05, 0.26]][[1.39, 1.46, 1.44]][[0.32, 0.32, 0.32]][0.3]
******
==============
O_threshold = 105
MC for this TARGET:[86.35, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.23, -2.26, -2.08]][[-3.2, -3.33, -3.3]][[-86.35, -86.35, -86.35]][[-2.11, -5.57]]
std:[[0.67, 0.68, 0.02]][[0.07, 0.06, 0.11]][[0.0, 0.0, 0.0]][[0.03, 0.0]]
MSE:[[2.33, 2.36, 2.08]][[3.2, 3.33, 3.3]][[86.35, 86.35, 86.35]][[2.11, 5.57]]
MSE(-DR):[[0.0, 0.03, -0.25]][[0.87, 1.0, 0.97]][[84.02, 84.02, 84.02]][[-0.22, 3.24]]
better than DR_NO_MARL
MC-based ATE = 0.95
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.65, -2.62, -2.08]][[-4.14, -4.19, -4.14]][[-0.95, -0.95, -0.95]][-2.06]
std:[[0.79, 0.8, 0.08]][[0.17, 0.16, 0.08]][[0.0, 0.0, 0.0]][0.08]
MSE:[[2.77, 2.74, 2.08]][[4.14, 4.19, 4.14]][[0.95, 0.95, 0.95]][2.06]
MSE(-DR):[[0.0, -0.03, -0.69]][[1.37, 1.42, 1.37]][[-1.82, -1.82, -1.82]][-0.71]
better than DR_NO_MARL
==============


O_threshold = 110
MC for this TARGET:[85.917, 0.097]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.25, -2.28, -2.11]][[-3.6, -3.71, -3.66]][[-85.92, -85.92, -85.92]][[-2.14, -5.14]]
std:[[0.65, 0.66, 0.1]][[0.05, 0.03, 0.08]][[0.0, 0.0, 0.0]][[0.1, 0.0]]
MSE:[[2.34, 2.37, 2.11]][[3.6, 3.71, 3.66]][[85.92, 85.92, 85.92]][[2.14, 5.14]]
MSE(-DR):[[0.0, 0.03, -0.23]][[1.26, 1.37, 1.32]][[83.58, 83.58, 83.58]][[-0.2, 2.8]]
better than DR_NO_MARL
MC-based ATE = 0.52
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.68, -2.65, -2.12]][[-4.54, -4.57, -4.5]][[-0.52, -0.52, -0.52]][-2.09]
std:[[0.77, 0.78, 0.01]][[0.14, 0.14, 0.05]][[0.0, 0.0, 0.0]][0.01]
MSE:[[2.79, 2.76, 2.12]][[4.54, 4.57, 4.5]][[0.52, 0.52, 0.52]][2.09]
MSE(-DR):[[0.0, -0.03, -0.67]][[1.75, 1.78, 1.71]][[-2.27, -2.27, -2.27]][-0.7]
better than DR_NO_MARL
==============


time spent until now: 22.9 mins


--------------------------------------
[pattern_seed, T, sd_R] = [4, 672, 10]

max(u_O) =  155.2
O_threshold = 80
means of Order:

100.5 109.9 81.5 114.3 91.5

72.5 87.4 112.1 106.3 79.1

112.6 97.7 108.3 106.3 78.9

106.7 88.1 135.6 115.0 100.4

81.7 100.6 102.7 78.1 155.2

target policy:

1 1 1 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

1 1 1 0 1

number of reward locations:  21
O_threshold = 85
target policy:

1 1 0 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

0 1 1 0 1

number of reward locations:  19
O_threshold = 90
target policy:

1 1 0 1 1
```

```
0 0 1 1 0

1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  17
O_threshold = 95
target policy:

1 1 0 1 0

0 0 1 1 0

1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 100
target policy:

1 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 105
target policy:

0 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 0

0 0 0 0 1

number of reward locations:  11
O_threshold = 110
target policy:

0 0 0 1 0

0 0 1 0 0

1 0 0 0 0

0 0 1 1 0

0 0 0 0 1

number of reward locations:  6
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:82.764
O_threshold = 80
MC for this TARGET:[87.276, 0.091]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.61, 0.44, 0.32]][[1.04, 0.96, 1.04]][[-87.28, -87.28, -87.28]][[0.15, -4.51]]
std:[[0.33, 0.47, 0.16]][[0.14, 0.14, 0.1]][[0.0, 0.0, 0.0]][[0.02, 0.01]]
MSE:[[0.69, 0.64, 0.36]][[1.05, 0.97, 1.04]][[87.28, 87.28, 87.28]][[0.15, 4.51]]
MSE(-DR):[[0.0, -0.05, -0.33]][[0.36, 0.28, 0.35]][[86.59, 86.59, 86.59]][[-0.54, 3.82]]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[87.292, 0.091]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.47, 0.3, 0.02]][[0.84, 0.77, 0.86]][[-87.29, -87.29, -87.29]][[-0.16, -4.53]]
std:[[0.19, 0.29, 0.13]][[0.12, 0.12, 0.07]][[0.0, 0.0, 0.0]][[0.03, 0.01]]
MSE:[[0.51, 0.42, 0.13]][[0.85, 0.78, 0.86]][[87.29, 87.29, 87.29]][[0.16, 4.53]]
MSE(-DR):[[0.0, -0.09, -0.38]][[0.34, 0.27, 0.35]][[86.78, 86.78, 86.78]][[-0.35, 4.02]]
better than DR_NO_MARL
MC-based ATE = 0.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
```

```
bias:[[-0.14, -0.14, -0.31]][[-0.2, -0.19, -0.18]][[-0.02, -0.02, -0.02]][-0.31]
std:[[0.14, 0.18, 0.03]][[0.02, 0.03, 0.03]][[0.0, 0.0, 0.0]][0.01]
MSE:[[0.2, 0.23, 0.31]][[0.2, 0.19, 0.18]][[0.02, 0.02, 0.02]][0.31]
MSE(-DR):[[0.0, 0.03, 0.11]][[0.0, -0.01, -0.02]][[-0.18, -0.18, -0.18]][0.11]
******
==============


O_threshold = 90
MC for this TARGET:[88.766, 0.095]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.31, -0.47, -0.67]][[0.17, 0.03, 0.05]][[-88.77, -88.77, -88.77]][[-0.84, -6.0]]
std:[[0.14, 0.09, 0.13]][[0.15, 0.16, 0.08]][[0.0, 0.0, 0.0]][[0.07, 0.01]]
MSE:[[0.34, 0.48, 0.68]][[0.23, 0.16, 0.09]][[88.77, 88.77, 88.77]][[0.84, 6.0]]
MSE(-DR):[[0.0, 0.14, 0.34]][[-0.11, -0.18, -0.25]][[88.43, 88.43, 88.43]][[0.5, 5.66]]
MC-based ATE = 1.49
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.92, -0.91, -1.0]][[-0.87, -0.92, -0.99]][[-1.49, -1.49, -1.49]][-0.98]
std:[[0.48, 0.55, 0.03]][[0.01, 0.01, 0.03]][[0.0, 0.0, 0.0]][0.05]
MSE:[[1.04, 1.06, 1.0]][[0.87, 0.92, 0.99]][[1.49, 1.49, 1.49]][0.98]
MSE(-DR):[[0.0, 0.02, -0.04]][[-0.17, -0.12, -0.05]][[0.45, 0.45, 0.45]][-0.06]
==============


O_threshold = 95
MC for this TARGET:[88.247, 0.095]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.44, -0.57, -0.69]][[0.16, 0.03, 0.06]][[-88.25, -88.25, -88.25]][[-0.81, -5.48]]
std:[[0.16, 0.11, 0.15]][[0.17, 0.17, 0.11]][[0.0, 0.0, 0.0]][[0.11, 0.01]]
MSE:[[0.47, 0.58, 0.71]][[0.23, 0.17, 0.13]][[88.25, 88.25, 88.25]][[0.82, 5.48]]
MSE(-DR):[[0.0, 0.11, 0.24]][[-0.24, -0.3, -0.34]][[87.78, 87.78, 87.78]][[0.35, 5.01]]
MC-based ATE = 0.97
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.06, -1.01, -1.01]][[-0.88, -0.93, -0.98]][[-0.97, -0.97, -0.97]][-0.96]
std:[[0.49, 0.58, 0.0]][[0.02, 0.03, 0.0]][[0.0, 0.0, 0.0]][0.09]
MSE:[[1.17, 1.16, 1.01]][[0.88, 0.93, 0.98]][[0.97, 0.97, 0.97]][0.96]
MSE(-DR):[[0.0, -0.01, -0.16]][[-0.29, -0.24, -0.19]][[-0.2, -0.2, -0.2]][-0.21]
==============


O_threshold = 100
MC for this TARGET:[88.597, 0.095]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.19, -0.35, -0.67]][[-0.15, -0.29, -0.26]][[-88.6, -88.6, -88.6]][[-0.83, -5.83]]
std:[[0.17, 0.11, 0.03]][[0.12, 0.12, 0.08]][[0.0, 0.0, 0.0]][[0.09, 0.01]]
MSE:[[0.25, 0.37, 0.67]][[0.19, 0.31, 0.27]][[88.6, 88.6, 88.6]][[0.83, 5.83]]
MSE(-DR):[[0.0, 0.12, 0.42]][[-0.06, 0.06, 0.02]][[88.35, 88.35, 88.35]][[0.58, 5.58]]
MC-based ATE = 1.32
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.81, -0.79, -1.0]][[-1.19, -1.25, -1.3]][[-1.32, -1.32, -1.32]][-0.98]
std:[[0.5, 0.58, 0.19]][[0.02, 0.02, 0.03]][[0.0, 0.0, 0.0]][0.11]
MSE:[[0.95, 0.98, 1.02]][[1.19, 1.25, 1.3]][[1.32, 1.32, 1.32]][0.99]
MSE(-DR):[[0.0, 0.03, 0.07]][[0.24, 0.3, 0.35]][[0.37, 0.37, 0.37]][0.04]
******
==============


O_threshold = 105
MC for this TARGET:[86.663, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.09, -0.18, -0.42]][[-0.51, -0.63, -0.63]][[-86.66, -86.66, -86.66]][[-0.51, -3.9]]
std:[[0.5, 0.49, 0.19]][[0.19, 0.18, 0.1]][[0.0, 0.0, 0.0]][[0.17, 0.01]]
MSE:[[0.51, 0.52, 0.46]][[0.54, 0.66, 0.64]][[86.66, 86.66, 86.66]][[0.54, 3.9]]
MSE(-DR):[[0.0, 0.01, -0.05]][[0.03, 0.15, 0.13]][[86.15, 86.15, 86.15]][[0.03, 3.39]]
better than DR_NO_MARL
MC-based ATE = -0.61
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.7, -0.62, -0.74]][[-1.55, -1.59, -1.67]][[0.61, 0.61, 0.61]][-0.66]
std:[[0.84, 0.96, 0.03]][[0.05, 0.04, 0.0]][[0.0, 0.0, 0.0]][0.15]
MSE:[[1.09, 1.14, 0.74]][[1.55, 1.59, 1.67]][[0.61, 0.61, 0.61]][0.68]
MSE(-DR):[[0.0, 0.05, -0.35]][[0.46, 0.5, 0.58]][[-0.48, -0.48, -0.48]][-0.41]
better than DR_NO_MARL
==============


O_threshold = 110
MC for this TARGET:[86.876, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.76, -2.77, -2.51]][[-3.15, -3.25, -3.32]][[-86.88, -86.88, -86.88]][[-2.52, -4.11]]
std:[[0.68, 0.68, 0.27]][[0.09, 0.09, 0.05]][[0.0, 0.0, 0.0]][[0.27, 0.01]]
MSE:[[2.84, 2.85, 2.52]][[3.15, 3.25, 3.32]][[86.88, 86.88, 86.88]][[2.53, 4.11]]
MSE(-DR):[[0.0, 0.01, -0.32]][[0.31, 0.41, 0.48]][[84.04, 84.04, 84.04]][[-0.31, 1.27]]
better than DR_NO_MARL
MC-based ATE = -0.4
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.38, -3.21, -2.84]][[-4.19, -4.21, -4.36]][[0.4, 0.4, 0.4]][-2.67]
std:[[1.02, 1.15, 0.11]][[0.05, 0.05, 0.06]][[0.0, 0.0, 0.0]][0.24]
MSE:[[3.53, 3.41, 2.84]][[4.19, 4.21, 4.36]][[0.4, 0.4, 0.4]][2.68]
MSE(-DR):[[0.0, -0.12, -0.69]][[0.66, 0.68, 0.83]][[-3.13, -3.13, -3.13]][-0.85]
```

```
better than DR_NO_MARL
==============


time spent until now: 28.5 mins


--------------------------------------
[pattern_seed, T, sd_R] = [5, 672, 10]

max(u_O) =  161.8
O_threshold = 80
means of Order:

108.7 93.1 161.8 94.6 101.7

136.5 82.9 88.4 103.3 93.1

78.4 95.5 92.6 112.2 71.3

86.5 125.2 144.2 73.5 113.2

81.8 83.8 83.6 91.4 121.4

target policy:

1 1 1 1 1

1 1 1 1 1

0 1 1 1 0

1 1 1 0 1

1 1 1 1 1

number of reward locations:  22
O_threshold = 85
target policy:

1 1 1 1 1

1 0 1 1 1

0 1 1 1 0

1 1 1 0 1

0 0 0 1 1

number of reward locations:  18
O_threshold = 90
target policy:

1 1 1 1 1

1 0 0 1 1

0 1 1 1 0

0 1 1 0 1

0 0 0 1 1

number of reward locations:  16
O_threshold = 95
target policy:

1 0 1 0 1

1 0 0 1 0

0 1 0 1 0

0 1 1 0 1

0 0 0 0 1

number of reward locations:  11
O_threshold = 100
target policy:

1 0 1 0 1

1 0 0 1 0

0 0 0 1 0

0 1 1 0 1
```

```
0 0 0 0 1

number of reward locations:  10
O_threshold = 105
target policy:

1 0 1 0 0

1 0 0 0 0

0 0 0 1 0

0 1 1 0 1

0 0 0 0 1

number of reward locations:  8
O_threshold = 110
target policy:

0 0 1 0 0

1 0 0 0 0

0 0 0 1 0

0 1 1 0 1

0 0 0 0 1

number of reward locations:  7
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:80.552
O_threshold = 80
MC for this TARGET:[84.27, 0.094]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.69, 0.59, 0.22]][[1.04, 0.94, 0.97]][[-84.27, -84.27, -84.27]][[0.13, -3.72]]
std:[[0.21, 0.17, 0.23]][[0.16, 0.15, 0.06]][[0.0, 0.0, 0.0]][[0.19, 0.04]]
MSE:[[0.72, 0.61, 0.32]][[1.05, 0.95, 0.97]][[84.27, 84.27, 84.27]][[0.23, 3.72]]
MSE(-DR):[[0.0, -0.11, -0.4]][[0.33, 0.23, 0.25]][[83.55, 83.55, 83.55]][[-0.49, 3.0]]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[84.94, 0.09]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.6, 0.49, -0.03]][[0.62, 0.47, 0.52]][[-84.94, -84.94, -84.94]][[-0.13, -4.39]]
std:[[0.1, 0.12, 0.02]][[0.19, 0.16, 0.14]][[0.0, 0.0, 0.0]][[0.01, 0.04]]
MSE:[[0.61, 0.5, 0.04]][[0.65, 0.5, 0.54]][[84.94, 84.94, 84.94]][[0.13, 4.39]]
MSE(-DR):[[0.0, -0.11, -0.57]][[0.04, -0.11, -0.07]][[84.33, 84.33, 84.33]][[-0.48, 3.78]]
MC-based ATE = 0.67
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.09, -0.1, -0.25]][[-0.42, -0.47, -0.45]][[-0.67, -0.67, -0.67]][-0.26]
std:[[0.3, 0.29, 0.21]][[0.03, 0.01, 0.09]][[0.0, 0.0, 0.0]][0.2]
MSE:[[0.31, 0.31, 0.33]][[0.42, 0.47, 0.46]][[0.67, 0.67, 0.67]][0.33]
MSE(-DR):[[0.0, 0.0, 0.02]][[0.11, 0.16, 0.15]][[0.36, 0.36, 0.36]][0.02]
******
==============


O_threshold = 90
MC for this TARGET:[85.928, 0.092]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.36, -0.5, -0.87]][[-0.12, -0.27, -0.29]][[-85.93, -85.93, -85.93]][[-1.01, -5.38]]
std:[[0.17, 0.2, 0.06]][[0.09, 0.08, 0.03]][[0.0, 0.0, 0.0]][[0.09, 0.04]]
MSE:[[0.4, 0.54, 0.87]][[0.15, 0.28, 0.29]][[85.93, 85.93, 85.93]][[1.01, 5.38]]
MSE(-DR):[[0.0, 0.14, 0.47]][[-0.25, -0.12, -0.11]][[85.53, 85.53, 85.53]][[0.61, 4.98]]
MC-based ATE = 1.66
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.05, -1.09, -1.09]][[-1.16, -1.21, -1.26]][[-1.66, -1.66, -1.66]][-1.14]
std:[[0.37, 0.36, 0.29]][[0.07, 0.07, 0.03]][[0.0, 0.0, 0.0]][0.29]
MSE:[[1.11, 1.15, 1.13]][[1.16, 1.21, 1.26]][[1.66, 1.66, 1.66]][1.18]
MSE(-DR):[[0.0, 0.04, 0.02]][[0.05, 0.1, 0.15]][[0.55, 0.55, 0.55]][0.07]
******
==============


O_threshold = 95
MC for this TARGET:[86.554, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.11, -1.25, -1.46]][[-1.42, -1.59, -1.67]][[-86.55, -86.55, -86.55]][[-1.6, -6.0]]
std:[[0.28, 0.26, 0.28]][[0.09, 0.09, 0.05]][[0.0, 0.0, 0.0]][[0.26, 0.04]]
MSE:[[1.14, 1.28, 1.49]][[1.42, 1.59, 1.67]][[86.55, 86.55, 86.55]][[1.62, 6.0]]
MSE(-DR):[[0.0, 0.14, 0.35]][[0.28, 0.45, 0.53]][[85.41, 85.41, 85.41]][[0.48, 4.86]]
******
MC-based ATE = 2.28
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.8, -1.84, -1.68]][[-2.46, -2.53, -2.64]][[-2.28, -2.28, -2.28]][-1.72]
std:[[0.07, 0.09, 0.05]][[0.07, 0.06, 0.0]][[0.0, 0.0, 0.0]][0.07]
MSE:[[1.8, 1.84, 1.68]][[2.46, 2.53, 2.64]][[2.28, 2.28, 2.28]][1.72]
MSE(-DR):[[0.0, 0.04, -0.12]][[0.66, 0.73, 0.84]][[0.48, 0.48, 0.48]][-0.08]
better than DR_NO_MARL
==============


O_threshold = 100
MC for this TARGET:[87.399, 0.091]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.52, -1.61, -1.61]][[-2.13, -2.3, -2.41]][[-87.4, -87.4, -87.4]][[-1.7, -6.85]]
std:[[0.11, 0.12, 0.15]][[0.14, 0.14, 0.07]][[0.0, 0.0, 0.0]][[0.16, 0.04]]
MSE:[[1.52, 1.61, 1.62]][[2.13, 2.3, 2.41]][[87.4, 87.4, 87.4]][[1.71, 6.85]]
MSE(-DR):[[0.0, 0.09, 0.1]][[0.61, 0.78, 0.89]][[85.88, 85.88, 85.88]][[0.19, 5.33]]
******
MC-based ATE = 3.13
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.21, -2.2, -1.83]][[-3.16, -3.24, -3.38]][[-3.13, -3.13, -3.13]][-1.83]
std:[[0.09, 0.05, 0.08]][[0.02, 0.0, 0.02]][[0.0, 0.0, 0.0]][0.03]
MSE:[[2.21, 2.2, 1.83]][[3.16, 3.24, 3.38]][[3.13, 3.13, 3.13]][1.83]
MSE(-DR):[[0.0, -0.01, -0.38]][[0.95, 1.03, 1.17]][[0.92, 0.92, 0.92]][-0.38]
better than DR_NO_MARL
==============


O_threshold = 105
MC for this TARGET:[87.788, 0.091]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.07, -2.16, -2.33]][[-3.33, -3.49, -3.55]][[-87.79, -87.79, -87.79]][[-2.42, -7.24]]
std:[[0.23, 0.21, 0.14]][[0.09, 0.09, 0.05]][[0.0, 0.0, 0.0]][[0.12, 0.04]]
MSE:[[2.08, 2.17, 2.33]][[3.33, 3.49, 3.55]][[87.79, 87.79, 87.79]][[2.42, 7.24]]
MSE(-DR):[[0.0, 0.09, 0.25]][[1.25, 1.41, 1.47]][[85.71, 85.71, 85.71]][[0.34, 5.16]]
******
MC-based ATE = 3.52
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.76, -2.75, -2.56]][[-4.36, -4.43, -4.52]][[-3.52, -3.52, -3.52]][-2.55]
std:[[0.02, 0.04, 0.09]][[0.07, 0.06, 0.01]][[0.0, 0.0, 0.0]][0.07]
MSE:[[2.76, 2.75, 2.56]][[4.36, 4.43, 4.52]][[3.52, 3.52, 3.52]][2.55]
MSE(-DR):[[0.0, -0.01, -0.2]][[1.6, 1.67, 1.76]][[0.76, 0.76, 0.76]][-0.21]
better than DR_NO_MARL
==============


O_threshold = 110
MC for this TARGET:[87.252, 0.091]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.36, -2.43, -2.57]][[-3.49, -3.65, -3.71]][[-87.25, -87.25, -87.25]][[-2.65, -6.7]]
std:[[0.15, 0.11, 0.09]][[0.08, 0.07, 0.03]][[0.0, 0.0, 0.0]][[0.05, 0.04]]
MSE:[[2.36, 2.43, 2.57]][[3.49, 3.65, 3.71]][[87.25, 87.25, 87.25]][[2.65, 6.7]]
MSE(-DR):[[0.0, 0.07, 0.21]][[1.13, 1.29, 1.35]][[84.89, 84.89, 84.89]][[0.29, 4.34]]
******
MC-based ATE = 2.98
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.04, -3.02, -2.8]][[-4.53, -4.59, -4.68]][[-2.98, -2.98, -2.98]][-2.77]
std:[[0.06, 0.05, 0.14]][[0.08, 0.08, 0.03]][[0.0, 0.0, 0.0]][0.14]
MSE:[[3.04, 3.02, 2.8]][[4.53, 4.59, 4.68]][[2.98, 2.98, 2.98]][2.77]
MSE(-DR):[[0.0, -0.02, -0.24]][[1.49, 1.55, 1.64]][[-0.06, -0.06, -0.06]][-0.27]
better than DR_NO_MARL
==============


time spent until now: 34.1 mins


--------------------------------------
[pattern_seed, T, sd_R] = [6, 672, 10]

max(u_O) =  168.4
O_threshold = 80
means of Order:

93.5 115.1 103.9 83.1 60.5

119.4 124.6 73.5 138.1 91.3

168.4 112.2 93.0 127.4 101.7

102.1 101.0 96.4 112.9 117.0

106.8 143.0 75.8 90.7 117.3

target policy:

1 1 1 1 0

1 1 0 1 1
```

```
1 1 1 1 1

1 1 1 1 1

1 1 0 1 1

number of reward locations:  22
O_threshold = 85
target policy:

1 1 1 0 0

1 1 0 1 1

1 1 1 1 1

1 1 1 1 1

1 1 0 1 1

number of reward locations:  21
O_threshold = 90
target policy:

1 1 1 0 0

1 1 0 1 1

1 1 1 1 1

1 1 1 1 1

1 1 0 1 1

number of reward locations:  21
O_threshold = 95
target policy:

0 1 1 0 0

1 1 0 1 0

1 1 0 1 1

1 1 1 1 1

1 1 0 0 1

number of reward locations:  17
O_threshold = 100
target policy:

0 1 1 0 0

1 1 0 1 0

1 1 0 1 1

1 1 0 1 1

1 1 0 0 1

number of reward locations:  16
O_threshold = 105
target policy:

0 1 0 0 0

1 1 0 1 0

1 1 0 1 0

0 0 0 1 1

1 1 0 0 1

number of reward locations:  12
O_threshold = 110
target policy:

0 1 0 0 0

1 1 0 1 0

1 1 0 1 0

0 0 0 1 1

0 1 0 0 1
```