```
Last login: Mon Mar 30 13:29:18 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-223-141-217.compute-1.amazonaws.com

ssh: connect to host ec2-3-223-141-217.compute-1.amazonaws.com port 22: Connection refused
Run-Mac:.ssh mac$
Run-Mac:.ssh mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-223-141-217.compute-1.amazonaws.com
The authenticity of host 'ec2-3-223-141-217.compute-1.amazonaws.com (3.223.141.217)' can't be established.
ECDSA key fingerprint is SHA256:fnERXPJu9ZIjnlvMR80ipmfOYxqHm8GTsj9tLvcJmBg.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-223-141-217.compute-1.amazonaws.com,3.223.141.217' (ECDSA) 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 21:30:17 UTC 2020

  System load:  0.72               Processes:           379
  Usage of /:   55.4% of 15.45GB   Users logged in:     0
  Memory usage: 0%                 IP address for ens5: 172.31.13.254
  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

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


Last login: Thu Mar  5 21:23:34 2020 from 107.13.161.147
ubuntu@ip-172-31-13-254:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1
ubuntu@ip-172-31-13-254:~$ python EC2.py
17:32, 03/30; num of cores:36

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, lam, simple, M_in_R, u_O_u_D, mean_reversion] = [672, 5, 5, 10, 0.2, 1, 1, 1e-05,
False, True, 10, False]


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

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

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1
```

```
1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 115
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 0 1 0 1

number of reward locations:  10
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
O_threshold = 130
target policy:

1 0 0 1 1

0 0 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  6
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:74.704
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:[[0.81, 0.72, 0.83]][[3.45, 3.11, 2.96]][[-83.93, -83.93, -83.93]][[0.74, -9.23]]
std:[[0.92, 0.9, 0.48]][[0.27, 0.28, 0.2]][[0.0, 0.0, 0.0]][[0.43, 0.2]]
MSE:[[1.23, 1.15, 0.96]][[3.46, 3.12, 2.97]][[83.93, 83.93, 83.93]][[0.86, 9.23]]
MSE(-DR):[[0.0, -0.08, -0.27]][[2.23, 1.89, 1.74]][[82.7, 82.7, 82.7]][[-0.37, 8.0]]
better than DR_NO_MARL
```

==========

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:[[0.76, 0.69, 0.36]][[3.82, 3.48, 3.27]][[-82.1, -82.1, -82.1]][[0.29, -7.39]]
std:[[0.89, 0.88, 0.46]][[0.28, 0.3, 0.21]][[0.0, 0.0, 0.0]][[0.43, 0.2]]
MSE:[[1.17, 1.12, 0.58]][[3.83, 3.49, 3.28]][[82.1, 82.1, 82.1]][[0.52, 7.39]]
MSE(-DR):[[0.0, -0.05, -0.59]][[2.66, 2.32, 2.11]][[80.93, 80.93, 80.93]][[-0.65, 6.22]]
better than DR_NO_MARL
MC-based ATE = -1.83
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.06, -0.03, -0.47]][[0.37, 0.37, 0.31]][[1.83, 1.83, 1.83]][-0.45]
std:[[0.39, 0.4, 0.19]][[0.1, 0.1, 0.07]][[0.0, 0.0, 0.0]][0.18]
MSE:[[0.39, 0.4, 0.51]][[0.38, 0.38, 0.32]][[1.83, 1.83, 1.83]][0.48]
MSE(-DR):[[0.0, 0.01, 0.12]][[-0.01, -0.01, -0.07]][[1.44, 1.44, 1.44]][0.09]
==========

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:[[-1.22, -1.34, -2.62]][[0.74, 0.33, -0.05]][[-85.64, -85.64, -85.64]][[-2.74, -10.94]]
std:[[0.72, 0.74, 0.35]][[0.29, 0.3, 0.23]][[0.0, 0.0, 0.0]][[0.38, 0.2]]
MSE:[[1.42, 1.53, 2.64]][[0.79, 0.45, 0.24]][[85.64, 85.64, 85.64]][[2.77, 10.94]]
MSE(-DR):[[0.0, 0.11, 1.22]][[-0.63, -0.97, -1.18]][[84.22, 84.22, 84.22]][[1.35, 9.52]]
MC-based ATE = 1.71
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.03, -2.06, -3.45]][[-2.72, -2.78, -3.01]][[-1.71, -1.71, -1.71]][-3.48]
std:[[0.84, 0.8, 0.55]][[0.16, 0.15, 0.11]][[0.0, 0.0, 0.0]][0.51]
MSE:[[2.2, 2.21, 3.49]][[2.72, 2.78, 3.01]][[1.71, 1.71, 1.71]][3.52]
MSE(-DR):[[0.0, 0.01, 1.29]][[0.52, 0.58, 0.81]][[-0.49, -0.49, -0.49]][1.32]
******
==========

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:[[-2.66, -2.78, -3.53]][[-3.02, -3.32, -3.76]][[-83.16, -83.16, -83.16]][[-3.64, -8.46]]
std:[[0.63, 0.64, 0.4]][[0.43, 0.45, 0.34]][[0.0, 0.0, 0.0]][[0.42, 0.2]]
MSE:[[2.73, 2.85, 3.55]][[3.05, 3.35, 3.78]][[83.16, 83.16, 83.16]][[3.66, 8.46]]
MSE(-DR):[[0.0, 0.12, 0.82]][[0.32, 0.62, 1.05]][[80.43, 80.43, 80.43]][[0.93, 5.73]]
******
MC-based ATE = -0.77
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.48, -3.5, -4.36]][[-6.47, -6.43, -6.71]][[0.77, 0.77, 0.77]][-4.38]
std:[[1.28, 1.28, 0.58]][[0.4, 0.4, 0.31]][[0.0, 0.0, 0.0]][0.57]
MSE:[[3.71, 3.73, 4.4]][[6.48, 6.44, 6.72]][[0.77, 0.77, 0.77]][4.42]
MSE(-DR):[[0.0, 0.02, 0.69]][[2.77, 2.73, 3.01]][[-2.94, -2.94, -2.94]][0.71]
******
==========

O_threshold = 115
MC for this TARGET:[82.398, 0.135]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.13, -3.19, -3.83]][[-3.61, -3.86, -4.29]][[-82.4, -82.4, -82.4]][[-3.9, -7.69]]
std:[[0.4, 0.42, 0.39]][[0.41, 0.42, 0.32]][[0.0, 0.0, 0.0]][[0.4, 0.2]]
MSE:[[3.16, 3.22, 3.85]][[3.63, 3.88, 4.3]][[82.4, 82.4, 82.4]][[3.92, 7.69]]
MSE(-DR):[[0.0, 0.06, 0.69]][[0.47, 0.72, 1.14]][[79.24, 79.24, 79.24]][[0.76, 4.53]]
******
MC-based ATE = -1.53
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.94, -3.92, -4.66]][[-7.06, -6.97, -7.25]][[1.53, 1.53, 1.53]][-4.64]
std:[[1.15, 1.14, 0.65]][[0.38, 0.39, 0.3]][[0.0, 0.0, 0.0]][0.61]
MSE:[[4.1, 4.08, 4.71]][[7.07, 6.98, 7.26]][[1.53, 1.53, 1.53]][4.68]
MSE(-DR):[[0.0, -0.02, 0.61]][[2.97, 2.88, 3.16]][[-2.57, -2.57, -2.57]][0.58]
******
==========

O_threshold = 120
MC for this TARGET:[83.847, 0.13]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-6.59, -6.65, -7.12]][[-7.15, -7.37, -7.78]][[-83.85, -83.85, -83.85]][[-7.17, -9.14]]
std:[[0.75, 0.78, 0.43]][[0.43, 0.43, 0.35]][[0.0, 0.0, 0.0]][[0.42, 0.2]]
MSE:[[6.63, 6.7, 7.13]][[7.16, 7.38, 7.79]][[83.85, 83.85, 83.85]][[7.18, 9.14]]
MSE(-DR):[[0.0, 0.07, 0.5]][[0.53, 0.75, 1.16]][[77.22, 77.22, 77.22]][[0.55, 2.51]]
******
MC-based ATE = -0.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-7.41, -7.37, -7.95]][[-10.6, -10.48, -10.74]][[0.09, 0.09, 0.09]][-7.91]
std:[[0.99, 1.03, 0.45]][[0.39, 0.41, 0.33]][[0.0, 0.0, 0.0]][0.43]
MSE:[[7.48, 7.44, 7.96]][[10.61, 10.49, 10.75]][[0.09, 0.09, 0.09]][7.92]
MSE(-DR):[[0.0, -0.04, 0.48]][[3.13, 3.01, 3.27]][[-7.39, -7.39, -7.39]][0.44]
******
==========

```
O_threshold = 130
MC for this TARGET:[86.096, 0.133]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.69, -9.72, -9.53]][[-11.39, -11.56, -12.02]][[-86.1, -86.1, -86.1]][[-9.57, -11.39]]
std:[[0.69, 0.72, 0.57]][[0.35, 0.36, 0.32]][[0.0, 0.0, 0.0]][[0.55, 0.2]]
MSE:[[9.71, 9.75, 9.55]][[11.4, 11.57, 12.02]][[86.1, 86.1, 86.1]][[9.59, 11.39]]
MSE(-DR):[[0.0, 0.04, -0.16]][[1.69, 1.86, 2.31]][[76.39, 76.39, 76.39]][[-0.12, 1.68]]
better than DR_NO_MARL
MC-based ATE = 2.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-10.5, -10.45, -10.36]][[-14.84, -14.67, -14.98]][[-2.16, -2.16, -2.16]][-10.31]
std:[[1.12, 1.15, 0.5]][[0.4, 0.42, 0.33]][[0.0, 0.0, 0.0]][0.48]
MSE:[[10.56, 10.51, 10.37]][[14.85, 14.68, 14.98]][[2.16, 2.16, 2.16]][10.32]
MSE(-DR):[[0.0, -0.05, -0.19]][[4.29, 4.12, 4.42]][[-8.4, -8.4, -8.4]][-0.24]
better than DR_NO_MARL
==============


time spent until now: 14.4 mins


---------------------------------------
[pattern_seed, T, sd_R] = [0, 480, 10]

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

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0
```

```
0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 115
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 0 1 0 1

number of reward locations:  10
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
O_threshold = 130
target policy:

1 0 0 1 1

0 0 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  6
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:74.741
O_threshold = 80
MC for this TARGET:[83.918, 0.107]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.06, 0.93, 0.98]][[3.29, 2.95, 2.93]][[-83.92, -83.92, -83.92]][[0.84, -9.18]]
std:[[1.04, 1.01, 0.5]][[0.26, 0.28, 0.18]][[0.0, 0.0, 0.0]][[0.45, 0.18]]
MSE:[[1.48, 1.37, 1.1]][[3.3, 2.96, 2.94]][[83.92, 83.92, 83.92]][[0.95, 9.18]]
MSE(-DR):[[0.0, -0.11, -0.38]][[1.82, 1.48, 1.46]][[82.44, 82.44, 82.44]][[-0.53, 7.7]]
better than DR_NO_MARL
==============


O_threshold = 90
MC for this TARGET:[82.085, 0.099]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.08, 0.94, 0.68]][[3.76, 3.41, 3.31]][[-82.08, -82.08, -82.08]][[0.53, -7.34]]
std:[[0.87, 0.85, 0.46]][[0.19, 0.21, 0.14]][[0.0, 0.0, 0.0]][[0.4, 0.18]]
MSE:[[1.39, 1.27, 0.82]][[3.76, 3.42, 3.31]][[82.08, 82.08, 82.08]][[0.66, 7.34]]
MSE(-DR):[[0.0, -0.12, -0.57]][[2.37, 2.03, 1.92]][[80.69, 80.69, 80.69]][[-0.73, 5.95]]
better than DR_NO_MARL
MC-based ATE = -1.83
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.02, 0.01, -0.3]][[0.46, 0.46, 0.38]][[1.83, 1.83, 1.83]][-0.31]
std:[[0.37, 0.38, 0.25]][[0.12, 0.12, 0.09]][[0.0, 0.0, 0.0]][0.23]
MSE:[[0.37, 0.38, 0.39]][[0.48, 0.48, 0.39]][[1.83, 1.83, 1.83]][0.39]
MSE(-DR):[[0.0, 0.01, 0.02]][[0.11, 0.11, 0.02]][[1.46, 1.46, 1.46]][0.02]
******
==============


O_threshold = 100
MC for this TARGET:[85.629, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.69, -0.84, -2.49]][[0.78, 0.37, 0.06]][[-85.63, -85.63, -85.63]][[-2.64, -10.89]]
std:[[0.55, 0.55, 0.3]][[0.18, 0.21, 0.12]][[0.0, 0.0, 0.0]][[0.29, 0.18]]
MSE:[[0.88, 1.0, 2.51]][[0.8, 0.43, 0.13]][[85.63, 85.63, 85.63]][[2.66, 10.89]]
MSE(-DR):[[0.0, 0.12, 1.63]][[-0.08, -0.45, -0.75]][[84.75, 84.75, 84.75]][[1.78, 10.01]]
MC-based ATE = 1.71
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.75, -1.77, -3.47]][[-2.52, -2.58, -2.87]][[-1.71, -1.71, -1.71]][-3.48]
std:[[0.8, 0.77, 0.42]][[0.2, 0.2, 0.1]][[0.0, 0.0, 0.0]][0.38]
MSE:[[1.92, 1.93, 3.5]][[2.53, 2.59, 2.87]][[1.71, 1.71, 1.71]][3.5]
MSE(-DR):[[0.0, 0.01, 1.58]][[0.61, 0.67, 0.95]][[-0.21, -0.21, -0.21]][1.58]
******
==============


O_threshold = 110
MC for this TARGET:[83.143, 0.101]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.23, -2.36, -3.45]][[-2.86, -3.15, -3.63]][[-83.14, -83.14, -83.14]][[-3.57, -8.4]]
std:[[0.41, 0.41, 0.34]][[0.26, 0.26, 0.27]][[0.0, 0.0, 0.0]][[0.33, 0.18]]
MSE:[[2.27, 2.4, 3.47]][[2.87, 3.16, 3.64]][[83.14, 83.14, 83.14]][[3.59, 8.4]]
MSE(-DR):[[0.0, 0.13, 1.2]][[0.6, 0.89, 1.37]][[80.87, 80.87, 80.87]][[1.32, 6.13]]
******
MC-based ATE = -0.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.3, -3.28, -4.43]][[-6.16, -6.09, -6.56]][[0.78, 0.78, 0.78]][-4.42]
std:[[1.13, 1.11, 0.59]][[0.34, 0.33, 0.23]][[0.0, 0.0, 0.0]][0.55]
MSE:[[3.49, 3.46, 4.47]][[6.17, 6.1, 6.56]][[0.78, 0.78, 0.78]][4.45]
MSE(-DR):[[0.0, -0.03, 0.98]][[2.68, 2.61, 3.07]][[-2.71, -2.71, -2.71]][0.96]
******
==============


O_threshold = 115
MC for this TARGET:[82.383, 0.099]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.69, -2.77, -3.68]][[-3.43, -3.68, -4.15]][[-82.38, -82.38, -82.38]][[-3.76, -7.64]]
std:[[0.39, 0.41, 0.34]][[0.26, 0.24, 0.25]][[0.0, 0.0, 0.0]][[0.34, 0.18]]
MSE:[[2.72, 2.8, 3.7]][[3.44, 3.69, 4.16]][[82.38, 82.38, 82.38]][[3.78, 7.64]]
MSE(-DR):[[0.0, 0.08, 0.98]][[0.72, 0.97, 1.44]][[79.66, 79.66, 79.66]][[1.06, 4.92]]
******
MC-based ATE = -1.54
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.75, -3.7, -4.66]][[-6.72, -6.63, -7.09]][[1.54, 1.54, 1.54]][-4.61]
std:[[1.18, 1.16, 0.56]][[0.35, 0.35, 0.23]][[0.0, 0.0, 0.0]][0.53]
MSE:[[3.93, 3.88, 4.69]][[6.73, 6.64, 7.09]][[1.54, 1.54, 1.54]][4.64]
MSE(-DR):[[0.0, -0.05, 0.76]][[2.8, 2.71, 3.16]][[-2.39, -2.39, -2.39]][0.71]
******
==============


O_threshold = 120
MC for this TARGET:[83.834, 0.1]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-6.4, -6.45, -7.23]][[-7.12, -7.33, -7.79]][[-83.83, -83.83, -83.83]][[-7.28, -9.09]]
std:[[0.37, 0.38, 0.34]][[0.24, 0.24, 0.25]][[0.0, 0.0, 0.0]][[0.34, 0.18]]
MSE:[[6.41, 6.46, 7.24]][[7.12, 7.33, 7.79]][[83.83, 83.83, 83.83]][[7.29, 9.09]]
MSE(-DR):[[0.0, 0.05, 0.83]][[0.71, 0.92, 1.38]][[77.42, 77.42, 77.42]][[0.88, 2.68]]
******
MC-based ATE = -0.08
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-7.47, -7.38, -8.21]][[-10.41, -10.27, -10.72]][[0.08, 0.08, 0.08]][-8.12]
std:[[0.96, 0.95, 0.66]][[0.33, 0.33, 0.22]][[0.0, 0.0, 0.0]][0.64]
MSE:[[7.53, 7.44, 8.24]][[10.42, 10.28, 10.72]][[0.08, 0.08, 0.08]][8.15]
MSE(-DR):[[0.0, -0.09, 0.71]][[2.89, 2.75, 3.19]][[-7.45, -7.45, -7.45]][0.62]
******
==============


O_threshold = 130
MC for this TARGET:[86.084, 0.102]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-10.26, -10.24, -9.96]][[-11.32, -11.5, -11.99]][[-86.08, -86.08, -86.08]][[-9.94, -11.34]]
std:[[0.51, 0.54, 0.47]][[0.22, 0.23, 0.22]][[0.0, 0.0, 0.0]][[0.47, 0.18]]
MSE:[[10.27, 10.25, 9.97]][[11.32, 11.5, 11.99]][[86.08, 86.08, 86.08]][[9.95, 11.34]]
MSE(-DR):[[0.0, -0.02, -0.3]][[1.05, 1.23, 1.72]][[75.81, 75.81, 75.81]][[-0.32, 1.07]]
better than DR_NO_MARL
MC-based ATE = 2.17
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-11.33, -11.16, -10.94]][[-14.62, -14.44, -14.93]][[-2.17, -2.17, -2.17]][-10.78]
std:[[1.16, 1.19, 0.68]][[0.36, 0.37, 0.22]][[0.0, 0.0, 0.0]][0.66]
MSE:[[11.39, 11.22, 10.96]][[14.62, 14.44, 14.93]][[2.17, 2.17, 2.17]][10.8]
MSE(-DR):[[0.0, -0.17, -0.43]][[3.23, 3.05, 3.54]][[-9.22, -9.22, -9.22]][-0.59]
better than DR_NO_MARL
==============


time spent until now: 28.9 mins


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

max(u_O) =  156.6
O_threshold = 80
```

means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations: 24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations: 21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations: 18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations: 11
O_threshold = 115
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 0 1 0 1

number of reward locations: 10
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

```
0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
O_threshold = 130
target policy:

1 0 0 1 1

0 0 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  6
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:74.787
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:[[1.57, 1.48, 0.71]][[3.52, 3.14, 3.07]][[-83.92, -83.92, -83.92]][[0.63, -9.14]]
std:[[0.43, 0.42, 0.35]][[0.22, 0.22, 0.18]][[0.0, 0.0, 0.0]][[0.33, 0.22]]
MSE:[[1.63, 1.54, 0.79]][[3.53, 3.15, 3.08]][[83.92, 83.92, 83.92]][[0.71, 9.14]]
MSE(-DR):[[0.0, -0.09, -0.84]][[1.9, 1.52, 1.45]][[82.29, 82.29, 82.29]][[-0.92, 7.51]]
better than DR_NO_MARL
==============


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:[[1.67, 1.58, 0.49]][[3.94, 3.57, 3.42]][[-82.09, -82.09, -82.09]][[0.4, -7.3]]
std:[[0.34, 0.34, 0.33]][[0.28, 0.27, 0.26]][[0.0, 0.0, 0.0]][[0.3, 0.22]]
MSE:[[1.7, 1.62, 0.59]][[3.95, 3.58, 3.43]][[82.09, 82.09, 82.09]][[0.5, 7.3]]
MSE(-DR):[[0.0, -0.08, -1.11]][[2.25, 1.88, 1.73]][[80.39, 80.39, 80.39]][[-1.2, 5.6]]
better than DR_NO_MARL
MC-based ATE = -1.84
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.1, 0.1, -0.22]][[0.42, 0.43, 0.35]][[1.84, 1.84, 1.84]][-0.22]
std:[[0.27, 0.29, 0.19]][[0.1, 0.1, 0.1]][[0.0, 0.0, 0.0]][0.19]
MSE:[[0.29, 0.31, 0.29]][[0.43, 0.44, 0.36]][[1.84, 1.84, 1.84]][0.29]
MSE(-DR):[[0.0, 0.02, 0.0]][[0.14, 0.15, 0.07]][[1.55, 1.55, 1.55]][0.0]
******
==============


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:[[-0.89, -1.05, -2.67]][[0.91, 0.48, 0.14]][[-85.63, -85.63, -85.63]][[-2.83, -10.84]]
std:[[0.4, 0.41, 0.31]][[0.29, 0.27, 0.3]][[0.0, 0.0, 0.0]][[0.3, 0.22]]
MSE:[[0.98, 1.13, 2.69]][[0.96, 0.55, 0.33]][[85.63, 85.63, 85.63]][[2.85, 10.84]]
MSE(-DR):[[0.0, 0.15, 1.71]][[-0.02, -0.43, -0.65]][[84.65, 84.65, 84.65]][[1.87, 9.86]]
MC-based ATE = 1.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.46, -2.53, -3.38]][[-2.61, -2.66, -2.93]][[-1.7, -1.7, -1.7]][-3.45]
std:[[0.69, 0.67, 0.43]][[0.15, 0.15, 0.16]][[0.0, 0.0, 0.0]][0.39]
MSE:[[2.55, 2.62, 3.41]][[2.61, 2.66, 2.93]][[1.7, 1.7, 1.7]][3.47]
MSE(-DR):[[0.0, 0.07, 0.86]][[0.06, 0.11, 0.38]][[-0.85, -0.85, -0.85]][0.92]
******
==============


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:[[-2.42, -2.54, -3.42]][[-2.87, -3.18, -3.59]][[-83.14, -83.14, -83.14]][[-3.54, -8.36]]
std:[[0.46, 0.46, 0.17]][[0.2, 0.18, 0.21]][[0.0, 0.0, 0.0]][[0.17, 0.22]]
MSE:[[2.46, 2.58, 3.42]][[2.88, 3.19, 3.6]][[83.14, 83.14, 83.14]][[3.54, 8.36]]
MSE(-DR):[[0.0, 0.12, 0.96]][[0.42, 0.73, 1.14]][[80.68, 80.68, 80.68]][[1.08, 5.9]]
******
MC-based ATE = -0.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.99, -4.02, -4.13]][[-6.39, -6.33, -6.66]][[0.78, 0.78, 0.78]][-4.17]
std:[[0.62, 0.61, 0.35]][[0.1, 0.13, 0.1]][[0.0, 0.0, 0.0]][0.32]
MSE:[[4.04, 4.07, 4.14]][[6.39, 6.33, 6.66]][[0.78, 0.78, 0.78]][4.18]
MSE(-DR):[[0.0, 0.03, 0.1]][[2.35, 2.29, 2.62]][[-3.26, -3.26, -3.26]][0.14]
******
==============


O_threshold = 115
```

```
MC for this TARGET:[82.382, 0.08]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.0, -3.07, -3.72]][[-3.4, -3.68, -4.08]][[-82.38, -82.38, -82.38]][[-3.79, -7.59]]
std:[[0.5, 0.49, 0.19]][[0.23, 0.2, 0.23]][[0.0, 0.0, 0.0]][[0.18, 0.22]]
MSE:[[3.04, 3.11, 3.72]][[3.41, 3.69, 4.09]][[82.38, 82.38, 82.38]][[3.79, 7.59]]
MSE(-DR):[[0.0, 0.07, 0.68]][[0.37, 0.65, 1.05]][[79.34, 79.34, 79.34]][[0.75, 4.55]]
******
MC-based ATE = -1.54
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-4.56, -4.56, -4.43]][[-6.92, -6.82, -7.15]][[1.54, 1.54, 1.54]][-4.42]
std:[[0.69, 0.67, 0.35]][[0.11, 0.13, 0.12]][[0.0, 0.0, 0.0]][0.33]
MSE:[[4.61, 4.61, 4.44]][[6.92, 6.82, 7.15]][[1.54, 1.54, 1.54]][4.43]
MSE(-DR):[[0.0, 0.0, -0.17]][[2.31, 2.21, 2.54]][[-3.07, -3.07, -3.07]][-0.18]
better than DR_NO_MARL
==============


O_threshold = 120
MC for this TARGET:[83.836, 0.079]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-6.43, -6.43, -6.99]][[-6.97, -7.21, -7.62]][[-83.84, -83.84, -83.84]][[-6.99, -9.05]]
std:[[0.45, 0.47, 0.18]][[0.18, 0.16, 0.21]][[0.0, 0.0, 0.0]][[0.18, 0.22]]
MSE:[[6.45, 6.45, 6.99]][[6.97, 7.21, 7.62]][[83.84, 83.84, 83.84]][[6.99, 9.05]]
MSE(-DR):[[0.0, 0.0, 0.54]][[0.52, 0.76, 1.17]][[77.39, 77.39, 77.39]][[0.54, 2.6]]
******
MC-based ATE = -0.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-7.99, -7.91, -7.7]][[-10.49, -10.35, -10.69]][[0.09, 0.09, 0.09]][-7.62]
std:[[0.56, 0.59, 0.4]][[0.13, 0.16, 0.13]][[0.0, 0.0, 0.0]][0.36]
MSE:[[8.01, 7.93, 7.71]][[10.49, 10.35, 10.69]][[0.09, 0.09, 0.09]][7.63]
MSE(-DR):[[0.0, -0.08, -0.3]][[2.48, 2.34, 2.68]][[-7.92, -7.92, -7.92]][-0.38]
better than DR_NO_MARL
==============


O_threshold = 130
MC for this TARGET:[86.088, 0.084]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-10.01, -9.96, -9.68]][[-11.22, -11.43, -11.85]][[-86.09, -86.09, -86.09]][[-9.63, -11.3]]
std:[[0.77, 0.8, 0.26]][[0.2, 0.19, 0.22]][[0.0, 0.0, 0.0]][[0.3, 0.22]]
MSE:[[10.04, 9.99, 9.68]][[11.22, 11.43, 11.85]][[86.09, 86.09, 86.09]][[9.63, 11.3]]
MSE(-DR):[[0.0, -0.05, -0.36]][[1.18, 1.39, 1.81]][[76.05, 76.05, 76.05]][[-0.41, 1.26]]
better than DR_NO_MARL
MC-based ATE = 2.16
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-11.57, -11.44, -10.39]][[-14.74, -14.57, -14.92]][[-2.16, -2.16, -2.16]][-10.26]
std:[[0.68, 0.73, 0.41]][[0.21, 0.25, 0.12]][[0.0, 0.0, 0.0]][0.4]
MSE:[[11.59, 11.46, 10.4]][[14.74, 14.57, 14.92]][[2.16, 2.16, 2.16]][10.27]
MSE(-DR):[[0.0, -0.13, -1.19]][[3.15, 2.98, 3.33]][[-9.43, -9.43, -9.43]][-1.32]
better than DR_NO_MARL
==============


time spent until now: 44.5 mins


---------------------------------------
[pattern_seed, T, sd_R] = [1, 336, 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 = 90
target policy:
```

```
1 0 0 0 1

0 1 0 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

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

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 1 1

0 1 1 1 1

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

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1

number of reward locations:  10
O_threshold = 115
target policy:

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 1 0 1

number of reward locations:  8
O_threshold = 120
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 0 0 0

number of reward locations:  5
O_threshold = 130
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  3
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:66.691
O_threshold = 80
MC for this TARGET:[73.133, 0.127]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.78, 1.71, 1.13]][[3.88, 3.54, 3.41]][[-73.13, -73.13, -73.13]][[1.06, -6.44]]
```

```
std:[[0.84, 0.87, 0.3]][[0.3, 0.32, 0.24]][[0.0, 0.0, 0.0]][[0.32, 0.22]]
MSE:[[1.97, 1.92, 1.17]][[3.89, 3.55, 3.42]][[73.13, 73.13, 73.13]][[1.11, 6.44]]
MSE(-DR):[[0.0, -0.05, -0.8]][[1.92, 1.58, 1.45]][[71.16, 71.16, 71.16]][[-0.86, 4.47]]
better than DR_NO_MARL
==============


O_threshold = 90
MC for this TARGET:[73.499, 0.122]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.44, -0.57, -1.33]][[1.3, 0.94, 0.73]][[-73.5, -73.5, -73.5]][[-1.46, -6.81]]
std:[[0.4, 0.41, 0.23]][[0.26, 0.28, 0.22]][[0.0, 0.0, 0.0]][[0.23, 0.22]]
MSE:[[0.59, 0.7, 1.35]][[1.33, 0.98, 0.76]][[73.5, 73.5, 73.5]][[1.48, 6.81]]
MSE(-DR):[[0.0, 0.11, 0.76]][[0.74, 0.39, 0.17]][[72.91, 72.91, 72.91]][[0.89, 6.22]]
******
MC-based ATE = 0.37
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.22, -2.28, -2.46]][[-2.59, -2.6, -2.68]][[-0.37, -0.37, -0.37]][-2.52]
std:[[0.55, 0.57, 0.36]][[0.22, 0.22, 0.22]][[0.0, 0.0, 0.0]][0.34]
MSE:[[2.29, 2.35, 2.49]][[2.6, 2.61, 2.69]][[0.37, 0.37, 0.37]][2.54]
MSE(-DR):[[0.0, 0.06, 0.2]][[0.31, 0.32, 0.4]][[-1.92, -1.92, -1.92]][0.25]
******
==============


O_threshold = 100
MC for this TARGET:[77.165, 0.128]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-4.39, -4.46, -5.02]][[-3.99, -4.34, -4.72]][[-77.17, -77.17, -77.17]][[-5.09, -10.47]]
std:[[0.64, 0.65, 0.29]][[0.25, 0.26, 0.24]][[0.0, 0.0, 0.0]][[0.3, 0.22]]
MSE:[[4.44, 4.51, 5.03]][[4.0, 4.35, 4.73]][[77.17, 77.17, 77.17]][[5.1, 10.47]]
MSE(-DR):[[0.0, 0.07, 0.59]][[-0.44, -0.09, 0.29]][[72.73, 72.73, 72.73]][[0.66, 6.03]]
MC-based ATE = 4.03
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-6.17, -6.17, -6.15]][[-7.87, -7.88, -8.13]][[-4.03, -4.03, -4.03]][-6.15]
std:[[0.98, 0.98, 0.39]][[0.29, 0.28, 0.3]][[0.0, 0.0, 0.0]][0.36]
MSE:[[6.25, 6.25, 6.16]][[7.88, 7.88, 8.14]][[4.03, 4.03, 4.03]][6.16]
MSE(-DR):[[0.0, 0.0, -0.09]][[1.63, 1.63, 1.89]][[-2.22, -2.22, -2.22]][-0.09]
better than DR_NO_MARL
==============


O_threshold = 110
MC for this TARGET:[80.265, 0.136]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-7.11, -7.13, -7.41]][[-7.72, -8.07, -8.6]][[-80.26, -80.26, -80.26]][[-7.43, -13.57]]
std:[[0.68, 0.73, 0.33]][[0.24, 0.27, 0.25]][[0.0, 0.0, 0.0]][[0.37, 0.22]]
MSE:[[7.14, 7.17, 7.42]][[7.72, 8.07, 8.6]][[80.26, 80.26, 80.26]][[7.44, 13.57]]
MSE(-DR):[[0.0, 0.03, 0.28]][[0.58, 0.93, 1.46]][[73.12, 73.12, 73.12]][[0.3, 6.43]]
******
MC-based ATE = 7.13
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-8.89, -8.84, -8.54]][[-11.6, -11.61, -12.01]][[-7.13, -7.13, -7.13]][-8.49]
std:[[1.08, 1.18, 0.3]][[0.32, 0.32, 0.31]][[0.0, 0.0, 0.0]][0.37]
MSE:[[8.96, 8.92, 8.55]][[11.6, 11.61, 12.01]][[7.13, 7.13, 7.13]][8.5]
MSE(-DR):[[0.0, -0.04, -0.41]][[2.64, 2.65, 3.05]][[-1.83, -1.83, -1.83]][-0.46]
better than DR_NO_MARL
==============


O_threshold = 115
MC for this TARGET:[80.245, 0.136]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.36, -9.33, -9.08]][[-10.31, -10.59, -11.08]][[-80.24, -80.24, -80.24]][[-9.04, -13.55]]
std:[[0.93, 0.97, 0.35]][[0.31, 0.35, 0.27]][[0.0, 0.0, 0.0]][[0.38, 0.22]]
MSE:[[9.41, 9.38, 9.09]][[10.31, 10.6, 11.08]][[80.24, 80.24, 80.24]][[9.05, 13.55]]
MSE(-DR):[[0.0, -0.03, -0.32]][[0.9, 1.19, 1.67]][[70.83, 70.83, 70.83]][[-0.36, 4.14]]
better than DR_NO_MARL
MC-based ATE = 7.11
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-11.14, -11.03, -10.21]][[-14.19, -14.13, -14.49]][[-7.11, -7.11, -7.11]][-10.11]
std:[[1.38, 1.47, 0.38]][[0.44, 0.42, 0.37]][[0.0, 0.0, 0.0]][0.45]
MSE:[[11.23, 11.13, 10.22]][[14.2, 14.14, 14.49]][[7.11, 7.11, 7.11]][10.12]
MSE(-DR):[[0.0, -0.1, -1.01]][[2.97, 2.91, 3.26]][[-4.12, -4.12, -4.12]][-1.11]
better than DR_NO_MARL
==============


O_threshold = 120
MC for this TARGET:[78.018, 0.136]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.03, -8.99, -9.33]][[-11.18, -11.34, -11.85]][[-78.02, -78.02, -78.02]][[-9.29, -11.33]]
std:[[1.19, 1.22, 0.34]][[0.39, 0.41, 0.3]][[0.0, 0.0, 0.0]][[0.36, 0.22]]
MSE:[[9.11, 9.07, 9.34]][[11.19, 11.35, 11.85]][[78.02, 78.02, 78.02]][[9.3, 11.33]]
MSE(-DR):[[0.0, -0.04, 0.23]][[2.08, 2.24, 2.74]][[68.91, 68.91, 68.91]][[0.19, 2.22]]
******
MC-based ATE = 4.89
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-10.81, -10.69, -10.46]][[-15.06, -14.88, -15.26]][[-4.89, -4.89, -4.89]][-10.35]
```

```
std:[[1.49, 1.55, 0.46]][[0.48, 0.45, 0.39]][[0.0, 0.0, 0.0]][0.49]
MSE:[[10.91, 10.8, 10.47]][[15.07, 14.89, 15.26]][[4.89, 4.89, 4.89]][10.36]
MSE(-DR):[[0.0, -0.11, -0.44]][[4.16, 3.98, 4.35]][[-6.02, -6.02, -6.02]][-0.55]
better than DR_NO_MARL
==============


O_threshold = 130
MC for this TARGET:[75.724, 0.134]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-8.53, -8.49, -9.35]][[-11.62, -11.67, -12.14]][[-75.72, -75.72, -75.72]][[-9.31, -9.03]]
std:[[1.24, 1.3, 0.44]][[0.37, 0.37, 0.29]][[0.0, 0.0, 0.0]][[0.47, 0.22]]
MSE:[[8.62, 8.59, 9.36]][[11.63, 11.68, 12.14]][[75.72, 75.72, 75.72]][[9.32, 9.03]]
MSE(-DR):[[0.0, -0.03, 0.74]][[3.01, 3.06, 3.52]][[67.1, 67.1, 67.1]][[0.7, 0.41]]
******
MC-based ATE = 2.59
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-10.31, -10.2, -10.49]][[-15.5, -15.21, -15.55]][[-2.59, -2.59, -2.59]][-10.37]
std:[[1.61, 1.69, 0.61]][[0.5, 0.48, 0.4]][[0.0, 0.0, 0.0]][0.64]
MSE:[[10.43, 10.34, 10.51]][[15.51, 15.22, 15.56]][[2.59, 2.59, 2.59]][10.39]
MSE(-DR):[[0.0, -0.09, 0.08]][[5.08, 4.79, 5.13]][[-7.84, -7.84, -7.84]][-0.04]
******
==============


time spent until now: 58.6 mins


---------------------------------------
[pattern_seed, T, sd_R] = [1, 480, 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 = 90
target policy:

1 0 0 0 1

0 1 0 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

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

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 1 1

0 1 1 1 1

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

```
1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1

number of reward locations:  10
O_threshold = 115
target policy:

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 1 0 1

number of reward locations:  8
O_threshold = 120
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 0 0 0

number of reward locations:  5
O_threshold = 130
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  3
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5
6 7
--------------------------------------
Value of Behaviour policy:66.66
O_threshold = 80
MC for this TARGET:[73.132, 0.107]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[2.05, 1.9, 1.14]][[3.81, 3.47, 3.37]][[-73.13, -73.13, -73.13]][[0.99, -6.47]]
std:[[0.9, 0.91, 0.46]][[0.21, 0.18, 0.16]][[0.0, 0.0, 0.0]][[0.45, 0.12]]
MSE:[[2.24, 2.11, 1.23]][[3.82, 3.47, 3.37]][[73.13, 73.13, 73.13]][[1.09, 6.47]]
MSE(-DR):[[0.0, -0.13, -1.01]][[1.58, 1.23, 1.13]][[70.89, 70.89, 70.89]][[-1.15, 4.23]]
better than DR_NO_MARL
==============


O_threshold = 90
MC for this TARGET:[73.503, 0.105]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.57, -0.68, -1.4]][[1.17, 0.82, 0.65]][[-73.5, -73.5, -73.5]][[-1.52, -6.84]]
std:[[0.46, 0.47, 0.25]][[0.22, 0.2, 0.11]][[0.0, 0.0, 0.0]][[0.28, 0.12]]
MSE:[[0.73, 0.83, 1.42]][[1.19, 0.84, 0.66]][[73.5, 73.5, 73.5]][[1.55, 6.84]]
MSE(-DR):[[0.0, 0.1, 0.69]][[0.46, 0.11, -0.07]][[72.77, 72.77, 72.77]][[0.82, 6.11]]
******
MC-based ATE = 0.37
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.62, -2.58, -2.55]][[-2.65, -2.65, -2.72]][[-0.37, -0.37, -0.37]][-2.51]
std:[[0.86, 0.88, 0.49]][[0.2, 0.2, 0.16]][[0.0, 0.0, 0.0]][0.5]
MSE:[[2.76, 2.73, 2.6]][[2.66, 2.66, 2.72]][[0.37, 0.37, 0.37]][2.56]
MSE(-DR):[[0.0, -0.03, -0.16]][[-0.1, -0.1, -0.04]][[-2.39, -2.39, -2.39]][-0.2]
==============


O_threshold = 100
MC for this TARGET:[77.155, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-4.33, -4.46, -4.93]][[-3.9, -4.24, -4.58]][[-77.16, -77.16, -77.16]][[-5.06, -10.5]]
```

```
std:[[0.75, 0.73, 0.21]][[0.21, 0.21, 0.13]][[0.0, 0.0, 0.0]][[0.18, 0.12]]
MSE:[[4.39, 4.52, 4.93]][[3.91, 4.25, 4.58]][[77.16, 77.16, 77.16]][[5.06, 10.5]]
MSE(-DR):[[0.0, 0.13, 0.54]][[-0.48, -0.14, 0.19]][[72.77, 72.77, 72.77]][[0.67, 6.11]]
MC-based ATE = 4.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-6.38, -6.36, -6.07]][[-7.71, -7.71, -7.95]][[-4.02, -4.02, -4.02]][-6.05]
std:[[1.16, 1.15, 0.58]][[0.21, 0.22, 0.17]][[0.0, 0.0, 0.0]][0.55]
MSE:[[6.48, 6.46, 6.1]][[7.71, 7.71, 7.95]][[4.02, 4.02, 4.02]][6.07]
MSE(-DR):[[0.0, -0.02, -0.38]][[1.23, 1.23, 1.47]][[-2.46, -2.46, -2.46]][-0.41]
better than DR_NO_MARL
==============


O_threshold = 110
MC for this TARGET:[80.256, 0.101]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-6.84, -6.95, -7.06]][[-7.57, -7.92, -8.39]][[-80.26, -80.26, -80.26]][[-7.18, -13.6]]
std:[[0.51, 0.48, 0.25]][[0.18, 0.2, 0.15]][[0.0, 0.0, 0.0]][[0.23, 0.12]]
MSE:[[6.86, 6.97, 7.06]][[7.57, 7.92, 8.39]][[80.26, 80.26, 80.26]][[7.18, 13.6]]
MSE(-DR):[[0.0, 0.11, 0.2]][[0.71, 1.06, 1.53]][[73.4, 73.4, 73.4]][[0.32, 6.74]]
******
MC-based ATE = 7.12
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-8.89, -8.85, -8.2]][[-11.38, -11.39, -11.76]][[-7.12, -7.12, -7.12]][-8.17]
std:[[1.04, 1.04, 0.63]][[0.23, 0.24, 0.18]][[0.0, 0.0, 0.0]][0.59]
MSE:[[8.95, 8.91, 8.22]][[11.38, 11.39, 11.76]][[7.12, 7.12, 7.12]][8.19]
MSE(-DR):[[0.0, -0.04, -0.73]][[2.43, 2.44, 2.81]][[-1.83, -1.83, -1.83]][-0.76]
better than DR_NO_MARL
==============


O_threshold = 115
MC for this TARGET:[80.235, 0.103]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.16, -9.24, -8.68]][[-10.16, -10.45, -10.88]][[-80.24, -80.24, -80.24]][[-8.76, -13.58]]
std:[[0.66, 0.65, 0.16]][[0.22, 0.24, 0.17]][[0.0, 0.0, 0.0]][[0.15, 0.12]]
MSE:[[9.18, 9.26, 8.68]][[10.16, 10.45, 10.88]][[80.24, 80.24, 80.24]][[8.76, 13.58]]
MSE(-DR):[[0.0, 0.08, -0.5]][[0.98, 1.27, 1.7]][[71.06, 71.06, 71.06]][[-0.42, 4.4]]
better than DR_NO_MARL
MC-based ATE = 7.1
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-11.21, -11.13, -9.82]][[-13.97, -13.92, -14.25]][[-7.1, -7.1, -7.1]][-9.75]
std:[[1.24, 1.24, 0.57]][[0.3, 0.29, 0.24]][[0.0, 0.0, 0.0]][0.53]
MSE:[[11.28, 11.2, 9.84]][[13.97, 13.92, 14.25]][[7.1, 7.1, 7.1]][9.76]
MSE(-DR):[[0.0, -0.08, -1.44]][[2.69, 2.64, 2.97]][[-4.18, -4.18, -4.18]][-1.52]
better than DR_NO_MARL
==============


O_threshold = 120
MC for this TARGET:[78.004, 0.105]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.26, -9.27, -9.35]][[-11.1, -11.25, -11.71]][[-78.0, -78.0, -78.0]][[-9.35, -11.34]]
std:[[0.78, 0.77, 0.15]][[0.28, 0.29, 0.22]][[0.0, 0.0, 0.0]][[0.17, 0.12]]
MSE:[[9.29, 9.3, 9.35]][[11.1, 11.25, 11.71]][[78.0, 78.0, 78.0]][[9.35, 11.34]]
MSE(-DR):[[0.0, 0.01, 0.06]][[1.81, 1.96, 2.42]][[68.71, 68.71, 68.71]][[0.06, 2.05]]
******
MC-based ATE = 4.87
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-11.31, -11.17, -10.49]][[-14.91, -14.72, -15.08]][[-4.87, -4.87, -4.87]][-10.34]
std:[[1.24, 1.24, 0.47]][[0.36, 0.34, 0.3]][[0.0, 0.0, 0.0]][0.44]
MSE:[[11.38, 11.24, 10.5]][[14.91, 14.72, 15.08]][[4.87, 4.87, 4.87]][10.35]
MSE(-DR):[[0.0, -0.14, -0.88]][[3.53, 3.34, 3.7]][[-6.51, -6.51, -6.51]][-1.03]
better than DR_NO_MARL
==============


O_threshold = 130
MC for this TARGET:[75.711, 0.102]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.5, -9.48, -9.66]][[-11.53, -11.57, -12.01]][[-75.71, -75.71, -75.71]][[-9.63, -9.05]]
std:[[0.52, 0.56, 0.2]][[0.34, 0.34, 0.27]][[0.0, 0.0, 0.0]][[0.23, 0.12]]
MSE:[[9.51, 9.5, 9.66]][[11.54, 11.57, 12.01]][[75.71, 75.71, 75.71]][[9.63, 9.05]]
MSE(-DR):[[0.0, -0.01, 0.15]][[2.03, 2.06, 2.5]][[66.2, 66.2, 66.2]][[0.12, -0.46]]
******
MC-based ATE = 2.58
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-11.55, -11.37, -10.8]][[-15.34, -15.04, -15.38]][[-2.58, -2.58, -2.58]][-10.62]
std:[[1.1, 1.13, 0.43]][[0.4, 0.38, 0.34]][[0.0, 0.0, 0.0]][0.44]
MSE:[[11.6, 11.43, 10.81]][[15.35, 15.04, 15.38]][[2.58, 2.58, 2.58]][10.63]
MSE(-DR):[[0.0, -0.17, -0.79]][[3.75, 3.44, 3.78]][[-9.02, -9.02, -9.02]][-0.97]
better than DR_NO_MARL
==============


time spent until now: 73.1 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 = 90
target policy:

1 0 0 0 1

0 1 0 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

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

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 1 1

0 1 1 1 1

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

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1

number of reward locations: 10
O_threshold = 115
target policy:

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 1 0 1

number of reward locations: 8
O_threshold = 120
target policy:

1 0 0 0 0

```
0 1 0 0 0

1 0 0 0 1

0 0 0 0 0

0 1 0 0 0

number of reward locations:  5
O_threshold = 130
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  3
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:66.671
O_threshold = 80
MC for this TARGET:[73.147, 0.087]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.87, 1.76, 1.19]][[3.89, 3.53, 3.44]][[-73.15, -73.15, -73.15]][[1.07, -6.48]]
std:[[0.74, 0.72, 0.49]][[0.15, 0.15, 0.18]][[0.0, 0.0, 0.0]][[0.46, 0.16]]
MSE:[[2.01, 1.9, 1.29]][[3.89, 3.53, 3.44]][[73.15, 73.15, 73.15]][[1.16, 6.48]]
MSE(-DR):[[0.0, -0.11, -0.72]][[1.88, 1.52, 1.43]][[71.14, 71.14, 71.14]][[-0.85, 4.47]]
better than DR_NO_MARL
==============


O_threshold = 90
MC for this TARGET:[73.511, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.21, -0.33, -1.31]][[1.28, 0.93, 0.74]][[-73.51, -73.51, -73.51]][[-1.43, -6.84]]
std:[[0.47, 0.47, 0.32]][[0.14, 0.13, 0.14]][[0.0, 0.0, 0.0]][[0.29, 0.16]]
MSE:[[0.51, 0.57, 1.35]][[1.29, 0.94, 0.75]][[73.51, 73.51, 73.51]][[1.46, 6.84]]
MSE(-DR):[[0.0, 0.06, 0.84]][[0.78, 0.43, 0.24]][[73.0, 73.0, 73.0]][[0.95, 6.33]]
******
MC-based ATE = 0.36
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.08, -2.09, -2.5]][[-2.61, -2.6, -2.7]][[-0.36, -0.36, -0.36]][-2.51]
std:[[0.49, 0.49, 0.4]][[0.13, 0.12, 0.09]][[0.0, 0.0, 0.0]][0.39]
MSE:[[2.14, 2.15, 2.53]][[2.61, 2.6, 2.7]][[0.36, 0.36, 0.36]][2.54]
MSE(-DR):[[0.0, 0.01, 0.39]][[0.47, 0.46, 0.56]][[-1.78, -1.78, -1.78]][0.4]
******
==============


O_threshold = 100
MC for this TARGET:[77.163, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-4.28, -4.4, -5.02]][[-3.89, -4.25, -4.56]][[-77.16, -77.16, -77.16]][[-5.15, -10.49]]
std:[[0.54, 0.55, 0.34]][[0.21, 0.2, 0.17]][[0.0, 0.0, 0.0]][[0.34, 0.16]]
MSE:[[4.31, 4.43, 5.03]][[3.9, 4.25, 4.56]][[77.16, 77.16, 77.16]][[5.16, 10.49]]
MSE(-DR):[[0.0, 0.12, 0.72]][[-0.41, -0.06, 0.25]][[72.85, 72.85, 72.85]][[0.85, 6.18]]
MC-based ATE = 4.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-6.15, -6.17, -6.21]][[-7.79, -7.78, -8.0]][[-4.02, -4.02, -4.02]][-6.23]
std:[[0.81, 0.8, 0.53]][[0.19, 0.2, 0.14]][[0.0, 0.0, 0.0]][0.49]
MSE:[[6.2, 6.22, 6.23]][[7.79, 7.78, 8.0]][[4.02, 4.02, 4.02]][6.25]
MSE(-DR):[[0.0, 0.02, 0.03]][[1.59, 1.58, 1.8]][[-2.18, -2.18, -2.18]][0.05]
******
==============


O_threshold = 110
MC for this TARGET:[80.264, 0.083]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-6.8, -6.93, -7.12]][[-7.55, -7.92, -8.38]][[-80.26, -80.26, -80.26]][[-7.25, -13.59]]
std:[[0.52, 0.51, 0.37]][[0.25, 0.24, 0.2]][[0.0, 0.0, 0.0]][[0.34, 0.16]]
MSE:[[6.82, 6.95, 7.13]][[7.55, 7.92, 8.38]][[80.26, 80.26, 80.26]][[7.26, 13.59]]
MSE(-DR):[[0.0, 0.13, 0.31]][[0.73, 1.1, 1.56]][[73.44, 73.44, 73.44]][[0.44, 6.77]]
******
MC-based ATE = 7.12
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-8.68, -8.69, -8.31]][[-11.44, -11.44, -11.83]][[-7.12, -7.12, -7.12]][-8.32]
std:[[0.98, 0.95, 0.59]][[0.23, 0.23, 0.16]][[0.0, 0.0, 0.0]][0.53]
MSE:[[8.74, 8.74, 8.33]][[11.44, 11.44, 11.83]][[7.12, 7.12, 7.12]][8.34]
MSE(-DR):[[0.0, 0.0, -0.41]][[2.7, 2.7, 3.09]][[-1.62, -1.62, -1.62]][-0.4]
better than DR_NO_MARL
```

==========

O_threshold = 115
MC for this TARGET:[80.243, 0.084]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.01, -9.07, -8.75]][[-10.09, -10.38, -10.82]][[-80.24, -80.24, -80.24]][[-8.81, -13.57]]
std:[[0.39, 0.39, 0.24]][[0.2, 0.2, 0.18]][[0.0, 0.0, 0.0]][[0.23, 0.16]]
MSE:[[9.02, 9.08, 8.75]][[10.09, 10.38, 10.82]][[80.24, 80.24, 80.24]][[8.81, 13.57]]
MSE(-DR):[[0.0, 0.06, -0.27]][[1.07, 1.36, 1.8]][[71.22, 71.22, 71.22]][[-0.21, 4.55]]
better than DR_NO_MARL
MC-based ATE = 7.1
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-10.88, -10.83, -9.94]][[-13.98, -13.91, -14.26]][[-7.1, -7.1, -7.1]][-9.89]
std:[[0.89, 0.86, 0.47]][[0.2, 0.2, 0.15]][[0.0, 0.0, 0.0]][0.41]
MSE:[[10.92, 10.86, 9.95]][[13.98, 13.91, 14.26]][[7.1, 7.1, 7.1]][9.9]
MSE(-DR):[[0.0, -0.06, -0.97]][[3.06, 2.99, 3.34]][[-3.82, -3.82, -3.82]][-1.02]
better than DR_NO_MARL
==========

O_threshold = 120
MC for this TARGET:[78.015, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.13, -9.14, -9.45]][[-11.03, -11.19, -11.64]][[-78.02, -78.02, -78.02]][[-9.46, -11.34]]
std:[[0.61, 0.62, 0.37]][[0.24, 0.23, 0.21]][[0.0, 0.0, 0.0]][[0.33, 0.16]]
MSE:[[9.15, 9.16, 9.46]][[11.03, 11.19, 11.64]][[78.02, 78.02, 78.02]][[9.47, 11.34]]
MSE(-DR):[[0.0, 0.01, 0.31]][[1.88, 2.04, 2.49]][[68.87, 68.87, 68.87]][[0.32, 2.19]]
******
MC-based ATE = 4.87
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-11.0, -10.91, -10.63]][[-14.92, -14.72, -15.08]][[-4.87, -4.87, -4.87]][-10.54]
std:[[1.07, 1.08, 0.57]][[0.25, 0.25, 0.19]][[0.0, 0.0, 0.0]][0.51]
MSE:[[11.05, 10.96, 10.65]][[14.92, 14.72, 15.08]][[4.87, 4.87, 4.87]][10.55]
MSE(-DR):[[0.0, -0.09, -0.4]][[3.87, 3.67, 4.03]][[-6.18, -6.18, -6.18]][-0.5]
better than DR_NO_MARL
==========

O_threshold = 130
MC for this TARGET:[75.726, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.19, -9.18, -9.7]][[-11.46, -11.52, -11.93]][[-75.73, -75.73, -75.73]][[-9.69, -9.05]]
std:[[0.76, 0.78, 0.35]][[0.28, 0.27, 0.22]][[0.0, 0.0, 0.0]][[0.33, 0.16]]
MSE:[[9.22, 9.21, 9.71]][[11.46, 11.52, 11.93]][[75.73, 75.73, 75.73]][[9.7, 9.05]]
MSE(-DR):[[0.0, -0.01, 0.49]][[2.24, 2.3, 2.71]][[66.51, 66.51, 66.51]][[0.48, -0.17]]
******
MC-based ATE = 2.58
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-11.06, -10.94, -10.89]][[-15.35, -15.05, -15.37]][[-2.58, -2.58, -2.58]][-10.77]
std:[[1.28, 1.29, 0.56]][[0.31, 0.31, 0.25]][[0.0, 0.0, 0.0]][0.51]
MSE:[[11.13, 11.02, 10.9]][[15.35, 15.05, 15.37]][[2.58, 2.58, 2.58]][10.78]
MSE(-DR):[[0.0, -0.11, -0.23]][[4.22, 3.92, 4.24]][[-8.55, -8.55, -8.55]][-0.35]
better than DR_NO_MARL
==========

time spent until now: 88.7 mins

[pattern_seed, T, sd_R] = [2, 336, 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 = 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 = 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 = 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
O_threshold = 115
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 120
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 130
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  2
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5
6 7
--------------------------------------
Value of Behaviour policy:65.148
```

```
O_threshold = 80
MC for this TARGET:[73.747, 0.134]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.16, -0.24, -0.69]][[3.06, 2.67, 2.49]][[-73.75, -73.75, -73.75]][[-0.77, -8.6]]
std:[[0.72, 0.75, 0.45]][[0.37, 0.39, 0.26]][[0.0, 0.0, 0.0]][[0.44, 0.15]]
MSE:[[0.74, 0.79, 0.82]][[3.08, 2.7, 2.5]][[73.75, 73.75, 73.75]][[0.89, 8.6]]
MSE(-DR):[[0.0, 0.05, 0.08]][[2.34, 1.96, 1.76]][[73.01, 73.01, 73.01]][[0.15, 7.86]]
******
==============


O_threshold = 90
MC for this TARGET:[73.237, 0.132]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.56, -0.73, -1.8]][[-0.37, -0.71, -1.02]][[-73.24, -73.24, -73.24]][[-1.97, -8.09]]
std:[[0.55, 0.55, 0.25]][[0.36, 0.38, 0.31]][[0.0, 0.0, 0.0]][[0.26, 0.15]]
MSE:[[0.78, 0.91, 1.82]][[0.52, 0.81, 1.07]][[73.24, 73.24, 73.24]][[1.99, 8.09]]
MSE(-DR):[[0.0, 0.13, 1.04]][[-0.26, 0.03, 0.29]][[72.46, 72.46, 72.46]][[1.21, 7.31]]
MC-based ATE = -0.51
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.4, -0.5, -1.12]][[-3.43, -3.38, -3.52]][[0.51, 0.51, 0.51]][-1.21]
std:[[0.7, 0.77, 0.51]][[0.24, 0.24, 0.2]][[0.0, 0.0, 0.0]][0.48]
MSE:[[0.81, 0.92, 1.23]][[3.44, 3.39, 3.53]][[0.51, 0.51, 0.51]][1.3]
MSE(-DR):[[0.0, 0.11, 0.42]][[2.63, 2.58, 2.72]][[-0.3, -0.3, -0.3]][0.49]
******
==============


O_threshold = 100
MC for this TARGET:[71.404, 0.129]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.47, -2.56, -3.23]][[-3.94, -4.13, -4.43]][[-71.4, -71.4, -71.4]][[-3.32, -6.26]]
std:[[0.44, 0.45, 0.34]][[0.3, 0.3, 0.27]][[0.0, 0.0, 0.0]][[0.37, 0.15]]
MSE:[[2.51, 2.6, 3.25]][[3.95, 4.14, 4.44]][[71.4, 71.4, 71.4]][[3.34, 6.26]]
MSE(-DR):[[0.0, 0.09, 0.74]][[1.44, 1.63, 1.93]][[68.89, 68.89, 68.89]][[0.83, 3.75]]
******
MC-based ATE = -2.34
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.31, -2.32, -2.54]][[-7.0, -6.8, -6.93]][[2.34, 2.34, 2.34]][-2.56]
std:[[1.02, 1.06, 0.45]][[0.33, 0.33, 0.23]][[0.0, 0.0, 0.0]][0.47]
MSE:[[2.53, 2.55, 2.58]][[7.01, 6.81, 6.93]][[2.34, 2.34, 2.34]][2.6]
MSE(-DR):[[0.0, 0.02, 0.05]][[4.48, 4.28, 4.4]][[-0.19, -0.19, -0.19]][0.07]
******
==============


O_threshold = 110
MC for this TARGET:[72.499, 0.132]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-5.25, -5.26, -5.99]][[-6.97, -7.1, -7.42]][[-72.5, -72.5, -72.5]][[-6.0, -7.35]]
std:[[0.85, 0.87, 0.51]][[0.38, 0.37, 0.31]][[0.0, 0.0, 0.0]][[0.52, 0.15]]
MSE:[[5.32, 5.33, 6.01]][[6.98, 7.11, 7.43]][[72.5, 72.5, 72.5]][[6.02, 7.35]]
MSE(-DR):[[0.0, 0.01, 0.69]][[1.66, 1.79, 2.11]][[67.18, 67.18, 67.18]][[0.7, 2.03]]
******
MC-based ATE = -1.25
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-5.09, -5.02, -5.3]][[-10.03, -9.77, -9.92]][[1.25, 1.25, 1.25]][-5.23]
std:[[1.4, 1.44, 0.71]][[0.46, 0.45, 0.29]][[0.0, 0.0, 0.0]][0.7]
MSE:[[5.28, 5.22, 5.35]][[10.04, 9.78, 9.92]][[1.25, 1.25, 1.25]][5.28]
MSE(-DR):[[0.0, -0.06, 0.07]][[4.76, 4.5, 4.64]][[-4.03, -4.03, -4.03]][0.0]
******
==============


O_threshold = 115
MC for this TARGET:[72.761, 0.126]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-7.24, -7.21, -7.89]][[-10.58, -10.61, -10.95]][[-72.76, -72.76, -72.76]][[-7.87, -7.61]]
std:[[1.15, 1.17, 0.63]][[0.36, 0.33, 0.27]][[0.0, 0.0, 0.0]][[0.64, 0.15]]
MSE:[[7.33, 7.3, 7.92]][[10.59, 10.62, 10.95]][[72.76, 72.76, 72.76]][[7.9, 7.61]]
MSE(-DR):[[0.0, -0.03, 0.59]][[3.26, 3.29, 3.62]][[65.43, 65.43, 65.43]][[0.57, 0.28]]
******
MC-based ATE = -0.99
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-7.08, -6.98, -7.2]][[-13.64, -13.28, -13.44]][[0.99, 0.99, 0.99]][-7.1]
std:[[1.64, 1.67, 0.88]][[0.43, 0.42, 0.27]][[0.0, 0.0, 0.0]][0.9]
MSE:[[7.27, 7.18, 7.25]][[13.65, 13.29, 13.44]][[0.99, 0.99, 0.99]][7.16]
MSE(-DR):[[0.0, -0.09, -0.02]][[6.38, 6.02, 6.17]][[-6.28, -6.28, -6.28]][-0.11]
better than DR_NO_MARL
==============


O_threshold = 120
MC for this TARGET:[72.761, 0.126]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-7.23, -7.21, -7.89]][[-10.58, -10.61, -10.95]][[-72.76, -72.76, -72.76]][[-7.87, -7.61]]
std:[[1.16, 1.17, 0.66]][[0.35, 0.33, 0.27]][[0.0, 0.0, 0.0]][[0.67, 0.15]]
MSE:[[7.32, 7.3, 7.92]][[10.59, 10.62, 10.95]][[72.76, 72.76, 72.76]][[7.9, 7.61]]
MSE(-DR):[[0.0, -0.02, 0.6]][[3.27, 3.3, 3.63]][[65.44, 65.44, 65.44]][[0.58, 0.29]]
```

******
MC-based ATE = -0.99
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-7.08, -6.98, -7.2]][[-13.64, -13.28, -13.45]][[0.99, 0.99, 0.99]][-7.11]
std:[[1.65, 1.67, 0.89]][[0.43, 0.42, 0.27]][[0.0, 0.0, 0.0]][0.91]
MSE:[[7.27, 7.18, 7.25]][[13.65, 13.29, 13.45]][[0.99, 0.99, 0.99]][7.17]
MSE(-DR):[[0.0, -0.09, -0.02]][[6.38, 6.02, 6.18]][[-6.28, -6.28, -6.28]][-0.1]
better than DR_NO_MARL
==============

O_threshold = 130
MC for this TARGET:[74.565, 0.13]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-10.28, -10.27, -10.9]][[-13.85, -13.86, -14.15]][[-74.56, -74.56, -74.56]][[-10.9, -9.42]]
std:[[1.27, 1.28, 0.64]][[0.37, 0.34, 0.29]][[0.0, 0.0, 0.0]][[0.66, 0.15]]
MSE:[[10.36, 10.35, 10.92]][[13.85, 13.86, 14.15]][[74.56, 74.56, 74.56]][[10.92, 9.42]]
MSE(-DR):[[0.0, -0.01, 0.56]][[3.49, 3.5, 3.79]][[64.2, 64.2, 64.2]][[0.56, -0.94]]
******
MC-based ATE = 0.82
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-10.12, -10.03, -10.22]][[-16.91, -16.53, -16.64]][[-0.82, -0.82, -0.82]][-10.13]
std:[[1.71, 1.73, 0.85]][[0.44, 0.43, 0.28]][[0.0, 0.0, 0.0]][0.88]
MSE:[[10.26, 10.18, 10.26]][[16.92, 16.54, 16.64]][[0.82, 0.82, 0.82]][10.17]
MSE(-DR):[[0.0, -0.08, 0.0]][[6.66, 6.28, 6.38]][[-9.44, -9.44, -9.44]][-0.09]
******
==============

time spent until now: 102.8 mins

[pattern_seed, T, sd_R] = [2, 480, 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 = 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 = 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 = 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
O_threshold = 115
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 120
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 130
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  2
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:65.186
O_threshold = 80
MC for this TARGET:[73.736, 0.101]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.6, 0.48, -0.43]][[3.12, 2.73, 2.66]][[-73.74, -73.74, -73.74]][[-0.55, -8.55]]
std:[[0.54, 0.55, 0.26]][[0.17, 0.18, 0.13]][[0.0, 0.0, 0.0]][[0.28, 0.14]]
MSE:[[0.81, 0.73, 0.5]][[3.12, 2.74, 2.66]][[73.74, 73.74, 73.74]][[0.62, 8.55]]
MSE(-DR):[[0.0, -0.08, -0.31]][[2.31, 1.93, 1.85]][[72.93, 72.93, 72.93]][[-0.19, 7.74]]
better than DR_NO_MARL
==============


O_threshold = 90
MC for this TARGET:[73.229, 0.095]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.44, -0.58, -1.76]][[-0.28, -0.63, -0.84]][[-73.23, -73.23, -73.23]][[-1.9, -8.04]]
std:[[0.44, 0.45, 0.27]][[0.27, 0.28, 0.19]][[0.0, 0.0, 0.0]][[0.25, 0.14]]
MSE:[[0.62, 0.73, 1.78]][[0.39, 0.69, 0.86]][[73.23, 73.23, 73.23]][[1.92, 8.04]]
MSE(-DR):[[0.0, 0.11, 1.16]][[-0.23, 0.07, 0.24]][[72.61, 72.61, 72.61]][[1.3, 7.42]]
MC-based ATE = -0.51
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.04, -1.06, -1.33]][[-3.41, -3.36, -3.5]][[0.51, 0.51, 0.51]][-1.35]
std:[[0.71, 0.73, 0.23]][[0.22, 0.22, 0.18]][[0.0, 0.0, 0.0]][0.23]
MSE:[[1.26, 1.29, 1.35]][[3.42, 3.37, 3.5]][[0.51, 0.51, 0.51]][1.37]
MSE(-DR):[[0.0, 0.03, 0.09]][[2.16, 2.11, 2.24]][[-0.75, -0.75, -0.75]][0.11]
******
==============
```

```
O_threshold = 100
MC for this TARGET:[71.393, 0.103]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.56, -2.6, -3.2]][[-3.84, -4.03, -4.33]][[-71.39, -71.39, -71.39]][[-3.24, -6.21]]
std:[[0.45, 0.46, 0.2]][[0.26, 0.27, 0.21]][[0.0, 0.0, 0.0]][[0.22, 0.14]]
MSE:[[2.6, 2.64, 3.21]][[3.85, 4.04, 4.34]][[71.39, 71.39, 71.39]][[3.25, 6.21]]
MSE(-DR):[[0.0, 0.04, 0.61]][[1.25, 1.44, 1.74]][[68.79, 68.79, 68.79]][[0.65, 3.61]]
******
MC-based ATE = -2.34
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.16, -3.08, -2.77]][[-6.97, -6.76, -6.98]][[2.34, 2.34, 2.34]][-2.69]
std:[[0.55, 0.56, 0.39]][[0.21, 0.2, 0.22]][[0.0, 0.0, 0.0]][0.42]
MSE:[[3.21, 3.13, 2.8]][[6.97, 6.76, 6.98]][[2.34, 2.34, 2.34]][2.72]
MSE(-DR):[[0.0, -0.08, -0.41]][[3.76, 3.55, 3.77]][[-0.87, -0.87, -0.87]][-0.49]
better than DR_NO_MARL
==============


O_threshold = 110
MC for this TARGET:[72.485, 0.106]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-5.37, -5.36, -6.06]][[-6.92, -7.06, -7.35]][[-72.48, -72.48, -72.48]][[-6.05, -7.3]]
std:[[0.56, 0.57, 0.31]][[0.25, 0.25, 0.2]][[0.0, 0.0, 0.0]][[0.28, 0.14]]
MSE:[[5.4, 5.39, 6.07]][[6.92, 7.06, 7.35]][[72.48, 72.48, 72.48]][[6.06, 7.3]]
MSE(-DR):[[0.0, -0.01, 0.67]][[1.52, 1.66, 1.95]][[67.08, 67.08, 67.08]][[0.66, 1.9]]
******
MC-based ATE = -1.25
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-5.98, -5.84, -5.63]][[-10.05, -9.79, -10.01]][[1.25, 1.25, 1.25]][-5.49]
std:[[0.71, 0.73, 0.46]][[0.23, 0.23, 0.24]][[0.0, 0.0, 0.0]][0.45]
MSE:[[6.02, 5.89, 5.65]][[10.05, 9.79, 10.01]][[1.25, 1.25, 1.25]][5.51]
MSE(-DR):[[0.0, -0.13, -0.37]][[4.03, 3.77, 3.99]][[-4.77, -4.77, -4.77]][-0.51]
better than DR_NO_MARL
==============


O_threshold = 115
MC for this TARGET:[72.733, 0.104]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-8.2, -8.12, -8.02]][[-10.51, -10.55, -10.82]][[-72.73, -72.73, -72.73]][[-7.94, -7.55]]
std:[[0.63, 0.62, 0.4]][[0.24, 0.23, 0.18]][[0.0, 0.0, 0.0]][[0.38, 0.14]]
MSE:[[8.22, 8.14, 8.03]][[10.51, 10.55, 10.82]][[72.73, 72.73, 72.73]][[7.95, 7.55]]
MSE(-DR):[[0.0, -0.08, -0.19]][[2.29, 2.33, 2.6]][[64.51, 64.51, 64.51]][[-0.27, -0.67]]
better than DR_NO_MARL
MC-based ATE = -1.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-8.8, -8.6, -7.59]][[-13.63, -13.28, -13.47]][[1.0, 1.0, 1.0]][-7.39]
std:[[0.64, 0.63, 0.47]][[0.24, 0.24, 0.23]][[0.0, 0.0, 0.0]][0.49]
MSE:[[8.82, 8.62, 7.6]][[13.63, 13.28, 13.47]][[1.0, 1.0, 1.0]][7.41]
MSE(-DR):[[0.0, -0.2, -1.22]][[4.81, 4.46, 4.65]][[-7.82, -7.82, -7.82]][-1.41]
better than DR_NO_MARL
==============


O_threshold = 120
MC for this TARGET:[72.733, 0.104]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-8.18, -8.12, -8.01]][[-10.51, -10.55, -10.82]][[-72.73, -72.73, -72.73]][[-7.95, -7.55]]
std:[[0.63, 0.62, 0.36]][[0.24, 0.23, 0.17]][[0.0, 0.0, 0.0]][[0.34, 0.14]]
MSE:[[8.2, 8.14, 8.02]][[10.51, 10.55, 10.82]][[72.73, 72.73, 72.73]][[7.96, 7.55]]
MSE(-DR):[[0.0, -0.06, -0.18]][[2.31, 2.35, 2.62]][[64.53, 64.53, 64.53]][[-0.24, -0.65]]
better than DR_NO_MARL
MC-based ATE = -1.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-8.78, -8.6, -7.58]][[-13.63, -13.28, -13.47]][[1.0, 1.0, 1.0]][-7.39]
std:[[0.63, 0.63, 0.45]][[0.24, 0.24, 0.22]][[0.0, 0.0, 0.0]][0.46]
MSE:[[8.8, 8.62, 7.59]][[13.63, 13.28, 13.47]][[1.0, 1.0, 1.0]][7.4]
MSE(-DR):[[0.0, -0.18, -1.21]][[4.83, 4.48, 4.67]][[-7.8, -7.8, -7.8]][-1.4]
better than DR_NO_MARL
==============


O_threshold = 130
MC for this TARGET:[74.541, 0.108]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-11.23, -11.15, -11.02]][[-13.73, -13.75, -14.0]][[-74.54, -74.54, -74.54]][[-10.93, -9.35]]
std:[[0.69, 0.65, 0.36]][[0.25, 0.25, 0.2]][[0.0, 0.0, 0.0]][[0.33, 0.14]]
MSE:[[11.25, 11.17, 11.03]][[13.73, 13.75, 14.0]][[74.54, 74.54, 74.54]][[10.93, 9.35]]
MSE(-DR):[[0.0, -0.08, -0.22]][[2.48, 2.5, 2.75]][[63.29, 63.29, 63.29]][[-0.32, -1.9]]
better than DR_NO_MARL
MC-based ATE = 0.8
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-11.84, -11.63, -10.59]][[-16.86, -16.49, -16.65]][[-0.8, -0.8, -0.8]][-10.38]
std:[[0.71, 0.69, 0.46]][[0.24, 0.23, 0.24]][[0.0, 0.0, 0.0]][0.49]
MSE:[[11.86, 11.65, 10.6]][[16.86, 16.49, 16.65]][[0.8, 0.8, 0.8]][10.39]
MSE(-DR):[[0.0, -0.21, -1.26]][[5.0, 4.63, 4.79]][[-11.06, -11.06, -11.06]][-1.47]
better than DR_NO_MARL
==============
```

time spent until now: 117.3 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 = 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 = 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 = 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
O_threshold = 115
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

```
number of reward locations:  3
O_threshold = 120
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 130
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 0 0 0

0 0 0 0 0

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