```
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
--------------------------------------
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
---------------------------------------
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
```