```
Last login: Sun Mar 29 23:01:32 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-228-4-227.compute-1.amazonaws.com
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1060-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

 System information disabled due to load higher than 16.0

 * Kubernetes 1.18 GA is now available! See https://microk8s.io for docs or
   install it with:

     sudo snap install microk8s --channel=1.18 --classic

 * Multipass 1.1 adds proxy support for developers behind enterprise
   firewalls. Rapid prototyping for cloud operations just got easier.

     https://multipass.run/

 * Canonical Livepatch is available for installation.
   - Reduce system reboots and improve kernel security. Activate at:
     https://ubuntu.com/livepatch

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


*** System restart required ***
Last login: Mon Mar 30 03:02:17 2020 from 107.13.161.147
ubuntu@ip-172-31-6-17:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1
ubuntu@ip-172-31-6-17:~$ python EC2.py
23:24, 03/29; num of cores:16

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


--------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 5]

max(u_O) =  156.6
O_threshold = 100
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 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 = 80
target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 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 = 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
1 2 3 4 5 1 2 3 4 5
----------------------------------------
Value of Behaviour policy:90.286
O_threshold = 100
MC for this TARGET:[98.071, 0.042]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.59, -0.69, -1.14]][[-1.05, -1.17, -1.25]][[-98.07, -98.07, -98.07]][[-1.24, -7.78]]
std:[[0.22, 0.2, 0.04]][[0.08, 0.08, 0.07]][[0.0, 0.0, 0.0]][[0.06, 0.05]]

MSE:[[0.63, 0.72, 1.14]][[1.05, 1.17, 1.25]][[98.07, 98.07, 98.07]][[1.24, 7.78]]
MSE(-DR):[[0.0, 0.09, 0.51]][[0.42, 0.54, 0.62]][[97.44, 97.44, 97.44]][[0.61, 7.15]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
==============
O_threshold = 90
MC for this TARGET:[96.124, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.1, -0.12, -0.37]][[-0.28, -0.34, -0.39]][[-96.12, -96.12, -96.12]][[-0.39, -5.84]]
std:[[0.23, 0.21, 0.06]][[0.14, 0.12, 0.1]][[0.0, 0.0, 0.0]][[0.04, 0.05]]

MSE:[[0.25, 0.24, 0.37]][[0.31, 0.36, 0.4]][[96.12, 96.12, 96.12]][[0.39, 5.84]]
MSE(-DR):[[0.0, -0.01, 0.12]][[0.06, 0.11, 0.15]][[95.87, 95.87, 95.87]][[0.14, 5.59]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -1.95
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.49, 0.57, 0.77]][[0.77, 0.83, 0.86]][[1.95, 1.95, 1.95]][0.85]
std:[[0.02, 0.01, 0.1]][[0.06, 0.05, 0.02]][[0.0, 0.0, 0.0]][0.1]

MSE:[[0.49, 0.57, 0.78]][[0.77, 0.83, 0.86]][[1.95, 1.95, 1.95]][0.86]
MSE(-DR):[[0.0, 0.08, 0.29]][[0.28, 0.34, 0.37]][[1.46, 1.46, 1.46]][0.37]
***** BETTER THAN [IS, DR_NO_MARL] *****
==============
O_threshold = 80
```

```
MC for this TARGET:[94.632, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.34, 0.33, 0.3]][[0.32, 0.32, 0.25]][[-94.63, -94.63, -94.63]][[0.3, -4.35]]
std:[[0.31, 0.29, 0.09]][[0.11, 0.09, 0.07]][[0.0, 0.0, 0.0]][[0.07, 0.05]]

MSE:[[0.46, 0.44, 0.31]][[0.34, 0.33, 0.26]][[94.63, 94.63, 94.63]][[0.31, 4.35]]
MSE(-DR):[[0.0, -0.02, -0.15]][[-0.12, -0.13, -0.2]][[94.17, 94.17, 94.17]][[-0.15, 3.89]]
MC-based ATE = -3.44
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.93, 1.02, 1.44]][[1.37, 1.5, 1.5]][[3.44, 3.44, 3.44]][1.54]
std:[[0.1, 0.09, 0.13]][[0.03, 0.01, 0.01]][[0.0, 0.0, 0.0]][0.13]

MSE:[[0.94, 1.02, 1.45]][[1.37, 1.5, 1.5]][[3.44, 3.44, 3.44]][1.55]
MSE(-DR):[[0.0, 0.08, 0.51]][[0.43, 0.56, 0.56]][[2.5, 2.5, 2.5]][0.61]
***** BETTER THAN [IS, DR_NO_MARL] *****
==============
O_threshold = 110
MC for this TARGET:[97.173, 0.042]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.85, -0.94, -1.4]][[-1.51, -1.66, -1.8]][[-97.17, -97.17, -97.17]][[-1.49, -6.89]]
std:[[0.14, 0.16, 0.0]][[0.09, 0.08, 0.08]][[0.0, 0.0, 0.0]][[0.02, 0.05]]

MSE:[[0.86, 0.95, 1.4]][[1.51, 1.66, 1.8]][[97.17, 97.17, 97.17]][[1.49, 6.89]]
MSE(-DR):[[0.0, 0.09, 0.54]][[0.65, 0.8, 0.94]][[96.31, 96.31, 96.31]][[0.63, 6.03]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -0.9
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.26, -0.25, -0.26]][[-0.46, -0.49, -0.56]][[0.9, 0.9, 0.9]][-0.25]
std:[[0.36, 0.36, 0.04]][[0.01, 0.0, 0.01]][[0.0, 0.0, 0.0]][0.04]

MSE:[[0.44, 0.44, 0.26]][[0.46, 0.49, 0.56]][[0.9, 0.9, 0.9]][0.25]
MSE(-DR):[[0.0, 0.0, -0.18]][[0.02, 0.05, 0.12]][[0.46, 0.46, 0.46]][-0.19]
better than DR_NO_MARL
==============
O_threshold = 120
MC for this TARGET:[96.466, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.39, -2.44, -2.83]][[-2.9, -3.03, -3.15]][[-96.47, -96.47, -96.47]][[-2.87, -6.18]]
std:[[0.42, 0.43, 0.13]][[0.12, 0.11, 0.13]][[0.0, 0.0, 0.0]][[0.14, 0.05]]

MSE:[[2.43, 2.48, 2.83]][[2.9, 3.03, 3.15]][[96.47, 96.47, 96.47]][[2.87, 6.18]]
MSE(-DR):[[0.0, 0.05, 0.4]][[0.47, 0.6, 0.72]][[94.04, 94.04, 94.04]][[0.44, 3.75]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -1.61
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.8, -1.75, -1.68]][[-1.85, -1.85, -1.9]][[1.61, 1.61, 1.61]][-1.63]
std:[[0.63, 0.63, 0.09]][[0.03, 0.03, 0.06]][[0.0, 0.0, 0.0]][0.09]

MSE:[[1.91, 1.86, 1.68]][[1.85, 1.85, 1.9]][[1.61, 1.61, 1.61]][1.63]
MSE(-DR):[[0.0, -0.05, -0.23]][[-0.06, -0.06, -0.01]][[-0.3, -0.3, -0.3]][-0.28]
==============
time spent until now: 4.0 mins


---------------------------------------
[pattern_seed, T, sd_R] = [1, 672, 5]

max(u_O) =  141.0
O_threshold = 100
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 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 = 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 = 80
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 = 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 = 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
1 2 3 4 5 1 2 3 4 5
--------------------------------------
Value of Behaviour policy:81.144
O_threshold = 100
MC for this TARGET:[90.096, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.22, -2.34, -2.71]][[-2.52, -2.72, -2.81]][[-90.1, -90.1, -90.1]][[-2.83, -8.95]]
std:[[0.1, 0.12, 0.1]][[0.12, 0.11, 0.17]][[0.0, 0.0, 0.0]][[0.11, 0.12]]

MSE:[[2.22, 2.34, 2.71]][[2.52, 2.72, 2.82]][[90.1, 90.1, 90.1]][[2.83, 8.95]]
MSE(-DR):[[0.0, 0.12, 0.49]][[0.3, 0.5, 0.6]][[87.88, 87.88, 87.88]][[0.61, 6.73]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
==============
O_threshold = 90
MC for this TARGET:[86.664, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
```

```
bias:[[-0.37, -0.46, -1.03]][[0.15, 0.01, -0.05]][[-86.66, -86.66, -86.66]][[-1.12, -5.52]]
std:[[0.34, 0.37, 0.18]][[0.13, 0.12, 0.17]][[0.0, 0.0, 0.0]][[0.21, 0.12]]

MSE:[[0.5, 0.59, 1.05]][[0.2, 0.12, 0.18]][[86.66, 86.66, 86.66]][[1.14, 5.52]]
MSE(-DR):[[0.0, 0.09, 0.55]][[-0.3, -0.38, -0.32]][[86.16, 86.16, 86.16]][[0.64, 5.02]]
MC-based ATE = -3.43
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.85, 1.88, 1.68]][[2.66, 2.72, 2.76]][[3.43, 3.43, 3.43]][1.71]
std:[[0.24, 0.25, 0.08]][[0.01, 0.01, 0.0]][[0.0, 0.0, 0.0]][0.1]

MSE:[[1.87, 1.9, 1.68]][[2.66, 2.72, 2.76]][[3.43, 3.43, 3.43]][1.71]
MSE(-DR):[[0.0, 0.03, -0.19]][[0.79, 0.85, 0.89]][[1.56, 1.56, 1.56]][-0.16]
better than DR_NO_MARL
==============
O_threshold = 80
MC for this TARGET:[83.89, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.07, 1.03, 0.95]][[1.97, 1.9, 1.88]][[-83.89, -83.89, -83.89]][[0.91, -2.75]]
std:[[0.41, 0.38, 0.18]][[0.19, 0.18, 0.18]][[0.0, 0.0, 0.0]][[0.15, 0.12]]

MSE:[[1.15, 1.1, 0.97]][[1.98, 1.91, 1.89]][[83.89, 83.89, 83.89]][[0.92, 2.75]]
MSE(-DR):[[0.0, -0.05, -0.18]][[0.83, 0.76, 0.74]][[82.74, 82.74, 82.74]][[-0.23, 1.6]]
better than DR_NO_MARL
MC-based ATE = -6.21
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[3.29, 3.38, 3.65]][[4.49, 4.62, 4.69]][[6.21, 6.21, 6.21]][3.74]
std:[[0.31, 0.26, 0.09]][[0.07, 0.07, 0.01]][[0.0, 0.0, 0.0]][0.04]

MSE:[[3.3, 3.39, 3.65]][[4.49, 4.62, 4.69]][[6.21, 6.21, 6.21]][3.74]
MSE(-DR):[[0.0, 0.09, 0.35]][[1.19, 1.32, 1.39]][[2.91, 2.91, 2.91]][0.44]
***** BETTER THAN [IS, DR_NO_MARL] *****
==============
O_threshold = 110
MC for this TARGET:[92.267, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.08, -3.17, -3.51]][[-4.28, -4.51, -4.64]][[-92.27, -92.27, -92.27]][[-3.6, -11.12]]
std:[[0.09, 0.07, 0.05]][[0.12, 0.1, 0.17]][[0.0, 0.0, 0.0]][[0.04, 0.12]]

MSE:[[3.08, 3.17, 3.51]][[4.28, 4.51, 4.64]][[92.27, 92.27, 92.27]][[3.6, 11.12]]
MSE(-DR):[[0.0, 0.09, 0.43]][[1.2, 1.43, 1.56]][[89.19, 89.19, 89.19]][[0.52, 8.04]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = 2.17
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.86, -0.83, -0.8]][[-1.76, -1.79, -1.84]][[-2.17, -2.17, -2.17]][-0.77]
std:[[0.19, 0.18, 0.15]][[0.0, 0.01, 0.0]][[0.0, 0.0, 0.0]][0.14]

MSE:[[0.88, 0.85, 0.81]][[1.76, 1.79, 1.84]][[2.17, 2.17, 2.17]][0.78]
MSE(-DR):[[0.0, -0.03, -0.07]][[0.88, 0.91, 0.96]][[1.29, 1.29, 1.29]][-0.1]
better than DR_NO_MARL
==============
O_threshold = 120
MC for this TARGET:[88.904, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.21, -3.29, -3.53]][[-5.35, -5.47, -5.59]][[-88.9, -88.9, -88.9]][[-3.61, -7.76]]
std:[[0.49, 0.46, 0.17]][[0.19, 0.17, 0.27]][[0.0, 0.0, 0.0]][[0.15, 0.12]]

MSE:[[3.25, 3.32, 3.53]][[5.35, 5.47, 5.6]][[88.9, 88.9, 88.9]][[3.61, 7.76]]
MSE(-DR):[[0.0, 0.07, 0.28]][[2.1, 2.22, 2.35]][[85.65, 85.65, 85.65]][[0.36, 4.51]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -1.19
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.98, -0.95, -0.82]][[-2.84, -2.76, -2.78]][[1.19, 1.19, 1.19]][-0.78]
std:[[0.59, 0.58, 0.26]][[0.07, 0.06, 0.09]][[0.0, 0.0, 0.0]][0.25]

MSE:[[1.14, 1.11, 0.86]][[2.84, 2.76, 2.78]][[1.19, 1.19, 1.19]][0.82]
MSE(-DR):[[0.0, -0.03, -0.28]][[1.7, 1.62, 1.64]][[0.05, 0.05, 0.05]][-0.32]
better than DR_NO_MARL
==============
time spent until now: 8.0 mins


---------------------------------------
[pattern_seed, T, sd_R] = [2, 672, 5]

max(u_O) =  157.3
O_threshold = 100
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:

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 = 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 = 80
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 = 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 = 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
1 2 ^CProcess Process-518:
Traceback (most recent call last):
  File "EC2.py", line 69, in <module>
Process Process-514:
    file = file, print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 62, in simu
    value_reps = rep_seeds(once, OPE_rep_times)
  File "/home/ubuntu/_uti_basic.py", line 119, in rep_seeds
Process Process-520:
    return list(map(fun, range(rep_times)))
  File "/home/ubuntu/simu_funs.py", line 58, in once
Process Process-517:
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 192, in simu_once
Process Process-522:
Process Process-525:
Process Process-513:
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 130, in V_DR
    r = arr(parmap(getOneRegionValue, range(N), n_cores))
  File "/home/ubuntu/_uti_basic.py", line 74, in parmap
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
Process Process-523:
    sent = [q_in.put((i, x)) for i, x in enumerate(X)]
  File "/home/ubuntu/_uti_basic.py", line 74, in <listcomp>
Process Process-515:
Process Process-521:
Process Process-526:
    sent = [q_in.put((i, x)) for i, x in enumerate(X)]
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
Process Process-519:
Traceback (most recent call last):
Process Process-516:
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
Traceback (most recent call last):
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
Traceback (most recent call last):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
```

```
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
Traceback (most recent call last):
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
```

```
    spatial = False)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
KeyboardInterrupt
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
Traceback (most recent call last):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
```

```
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
Process Process-527:
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
Process Process-524:
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
Process Process-528:
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
```

```
ubuntu@ip-172-31-6-17:~$ python EC2.py

23:33, 03/29; num of cores:16

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


---------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 5]

max(u_O) =  156.6
O_threshold = 100
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

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

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

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

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 105
target policy:

1 1 1 1 1

0 1 0 0 1

0 1 1 0 1

1 1 0 1 0

0 1 1 0 1
```

```
number of reward locations:  16
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
1 2 3 4 5 1 2 3 4 5
--------------------------------------
Value of Behaviour policy:90.286
O_threshold = 100
MC for this TARGET:[98.071, 0.042]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.6, -0.69, -1.15]][[-1.05, -1.17, -1.25]][[-98.07, -98.07, -98.07]][[-1.24, -7.78]]
std:[[0.21, 0.2, 0.04]][[0.09, 0.08, 0.08]][[0.0, 0.0, 0.0]][[0.05, 0.05]]

MSE:[[0.64, 0.72, 1.15]][[1.05, 1.17, 1.25]][[98.07, 98.07, 98.07]][[1.24, 7.78]]
MSE(-DR):[[0.0, 0.08, 0.51]][[0.41, 0.53, 0.61]][[97.43, 97.43, 97.43]][[0.6, 7.14]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
==============
O_threshold = 90
MC for this TARGET:[96.124, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.1, -0.12, -0.36]][[-0.28, -0.34, -0.38]][[-96.12, -96.12, -96.12]][[-0.38, -5.84]]
std:[[0.22, 0.21, 0.05]][[0.15, 0.12, 0.1]][[0.0, 0.0, 0.0]][[0.04, 0.05]]

MSE:[[0.24, 0.24, 0.36]][[0.32, 0.36, 0.39]][[96.12, 96.12, 96.12]][[0.38, 5.84]]
MSE(-DR):[[0.0, 0.0, 0.12]][[0.08, 0.12, 0.15]][[95.88, 95.88, 95.88]][[0.14, 5.6]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -1.95
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.51, 0.57, 0.79]][[0.78, 0.83, 0.87]][[1.95, 1.95, 1.95]][0.85]
std:[[0.01, 0.01, 0.09]][[0.05, 0.05, 0.02]][[0.0, 0.0, 0.0]][0.09]

MSE:[[0.51, 0.57, 0.8]][[0.78, 0.83, 0.87]][[1.95, 1.95, 1.95]][0.85]
MSE(-DR):[[0.0, 0.06, 0.29]][[0.27, 0.32, 0.36]][[1.44, 1.44, 1.44]][0.34]
***** BETTER THAN [IS, DR_NO_MARL] *****
==============
O_threshold = 95
MC for this TARGET:[96.124, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.09, -0.12, -0.35]][[-0.28, -0.34, -0.38]][[-96.12, -96.12, -96.12]][[-0.38, -5.84]]
std:[[0.24, 0.21, 0.07]][[0.15, 0.12, 0.1]][[0.0, 0.0, 0.0]][[0.05, 0.05]]

MSE:[[0.26, 0.24, 0.36]][[0.32, 0.36, 0.39]][[96.12, 96.12, 96.12]][[0.38, 5.84]]
MSE(-DR):[[0.0, -0.02, 0.1]][[0.06, 0.1, 0.13]][[95.86, 95.86, 95.86]][[0.12, 5.58]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -1.95
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.52, 0.57, 0.8]][[0.78, 0.83, 0.86]][[1.95, 1.95, 1.95]][0.86]
std:[[0.03, 0.01, 0.12]][[0.06, 0.05, 0.02]][[0.0, 0.0, 0.0]][0.1]

MSE:[[0.52, 0.57, 0.81]][[0.78, 0.83, 0.86]][[1.95, 1.95, 1.95]][0.87]
MSE(-DR):[[0.0, 0.05, 0.29]][[0.26, 0.31, 0.34]][[1.43, 1.43, 1.43]][0.35]
***** BETTER THAN [IS, DR_NO_MARL] *****
==============
O_threshold = 105
MC for this TARGET:[98.037, 0.041]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.15, -1.26, -1.75]][[-1.0, -1.14, -1.22]][[-98.04, -98.04, -98.04]][[-1.86, -7.75]]
std:[[0.18, 0.21, 0.01]][[0.12, 0.1, 0.1]][[0.0, 0.0, 0.0]][[0.01, 0.05]]

MSE:[[1.16, 1.28, 1.75]][[1.01, 1.14, 1.22]][[98.04, 98.04, 98.04]][[1.86, 7.75]]
MSE(-DR):[[0.0, 0.12, 0.59]][[-0.15, -0.02, 0.06]][[96.88, 96.88, 96.88]][[0.7, 6.59]]
MC-based ATE = -0.03
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.55, -0.57, -0.6]][[0.06, 0.03, 0.03]][[0.03, 0.03, 0.03]][-0.62]
std:[[0.02, 0.01, 0.03]][[0.03, 0.02, 0.02]][[0.0, 0.0, 0.0]][0.06]

MSE:[[0.55, 0.57, 0.6]][[0.07, 0.04, 0.04]][[0.03, 0.03, 0.03]][0.62]
```

```
MSE(-DR):[[0.0, 0.02, 0.05]][[-0.48, -0.51, -0.51]][[-0.52, -0.52, -0.52]][0.07]
==============
O_threshold = 110
MC for this TARGET:[97.173, 0.042]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.84, -0.94, -1.39]][[-1.51, -1.66, -1.8]][[-97.17, -97.17, -97.17]][[-1.49, -6.89]]
std:[[0.15, 0.16, 0.0]][[0.09, 0.08, 0.08]][[0.0, 0.0, 0.0]][[0.02, 0.05]]

MSE:[[0.85, 0.95, 1.39]][[1.51, 1.66, 1.8]][[97.17, 97.17, 97.17]][[1.49, 6.89]]
MSE(-DR):[[0.0, 0.1, 0.54]][[0.66, 0.81, 0.95]][[96.32, 96.32, 96.32]][[0.64, 6.04]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -0.9
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.23, -0.25, -0.24]][[-0.45, -0.49, -0.56]][[0.9, 0.9, 0.9]][-0.25]
std:[[0.35, 0.36, 0.04]][[0.0, 0.0, 0.0]][[0.0, 0.0, 0.0]][0.03]

MSE:[[0.42, 0.44, 0.24]][[0.45, 0.49, 0.56]][[0.9, 0.9, 0.9]][0.25]
MSE(-DR):[[0.0, 0.02, -0.18]][[0.03, 0.07, 0.14]][[0.48, 0.48, 0.48]][-0.17]
better than DR_NO_MARL
==============
time spent until now: 4.0 mins


---------------------------------------
[pattern_seed, T, sd_R] = [1, 672, 5]

max(u_O) =  141.0
O_threshold = 100
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 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 = 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 = 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 1 2 3 4 5
--------------------------------------
Value of Behaviour policy:81.144
O_threshold = 100
MC for this TARGET:[90.096, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.23, -2.34, -2.7]][[-2.52, -2.72, -2.81]][[-90.1, -90.1, -90.1]][[-2.82, -8.95]]
std:[[0.1, 0.12, 0.09]][[0.13, 0.11, 0.18]][[0.0, 0.0, 0.0]][[0.11, 0.12]]

MSE:[[2.23, 2.34, 2.7]][[2.52, 2.72, 2.82]][[90.1, 90.1, 90.1]][[2.82, 8.95]]
MSE(-DR):[[0.0, 0.11, 0.47]][[0.29, 0.49, 0.59]][[87.87, 87.87, 87.87]][[0.59, 6.72]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
==============
O_threshold = 90
MC for this TARGET:[86.664, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.35, -0.46, -1.01]][[0.14, 0.01, -0.05]][[-86.66, -86.66, -86.66]][[-1.12, -5.52]]
std:[[0.35, 0.37, 0.18]][[0.13, 0.12, 0.17]][[0.0, 0.0, 0.0]][[0.2, 0.12]]

MSE:[[0.49, 0.59, 1.03]][[0.19, 0.12, 0.18]][[86.66, 86.66, 86.66]][[1.14, 5.52]]
MSE(-DR):[[0.0, 0.1, 0.54]][[-0.3, -0.37, -0.31]][[86.17, 86.17, 86.17]][[0.65, 5.03]]
MC-based ATE = -3.43
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.87, 1.88, 1.7]][[2.66, 2.72, 2.76]][[3.43, 3.43, 3.43]][1.7]
std:[[0.25, 0.25, 0.09]][[0.0, 0.01, 0.01]][[0.0, 0.0, 0.0]][0.09]

MSE:[[1.89, 1.9, 1.7]][[2.66, 2.72, 2.76]][[3.43, 3.43, 3.43]][1.7]
MSE(-DR):[[0.0, 0.01, -0.19]][[0.77, 0.83, 0.87]][[1.54, 1.54, 1.54]][-0.19]
better than DR_NO_MARL
==============
O_threshold = 95
MC for this TARGET:[88.928, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.97, -2.08, -2.72]][[-1.73, -1.9, -1.98]][[-88.93, -88.93, -88.93]][[-2.82, -7.78]]
std:[[0.12, 0.12, 0.2]][[0.14, 0.14, 0.19]][[0.0, 0.0, 0.0]][[0.21, 0.12]]

MSE:[[1.97, 2.08, 2.73]][[1.74, 1.91, 1.99]][[88.93, 88.93, 88.93]][[2.83, 7.78]]
MSE(-DR):[[0.0, 0.11, 0.76]][[-0.23, -0.06, 0.02]][[86.96, 86.96, 86.96]][[0.86, 5.81]]
MC-based ATE = -1.17
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.26, 0.27, -0.01]][[0.79, 0.82, 0.82]][[1.17, 1.17, 1.17]][0.0]
std:[[0.02, 0.01, 0.11]][[0.02, 0.02, 0.01]][[0.0, 0.0, 0.0]][0.1]

MSE:[[0.26, 0.27, 0.11]][[0.79, 0.82, 0.82]][[1.17, 1.17, 1.17]][0.1]
MSE(-DR):[[0.0, 0.01, -0.15]][[0.53, 0.56, 0.56]][[0.91, 0.91, 0.91]][-0.16]
better than DR_NO_MARL
==============
```

```
O_threshold = 105
MC for this TARGET:[91.173, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.66, -2.77, -3.18]][[-3.3, -3.52, -3.64]][[-91.17, -91.17, -91.17]][[-3.29, -10.03]]
std:[[0.0, 0.03, 0.02]][[0.11, 0.1, 0.16]][[0.0, 0.0, 0.0]][[0.05, 0.12]]

MSE:[[2.66, 2.77, 3.18]][[3.3, 3.52, 3.64]][[91.17, 91.17, 91.17]][[3.29, 10.03]]
MSE(-DR):[[0.0, 0.11, 0.52]][[0.64, 0.86, 0.98]][[88.51, 88.51, 88.51]][[0.63, 7.37]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = 1.08
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.43, -0.42, -0.48]][[-0.78, -0.8, -0.83]][[-1.08, -1.08, -1.08]][-0.47]
std:[[0.1, 0.09, 0.08]][[0.01, 0.01, 0.03]][[0.0, 0.0, 0.0]][0.06]

MSE:[[0.44, 0.43, 0.49]][[0.78, 0.8, 0.83]][[1.08, 1.08, 1.08]][0.47]
MSE(-DR):[[0.0, -0.01, 0.05]][[0.34, 0.36, 0.39]][[0.64, 0.64, 0.64]][0.03]
***** BETTER THAN [IS, DR_NO_MARL] *****
==============
O_threshold = 110
MC for this TARGET:[92.267, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.07, -3.17, -3.49]][[-4.28, -4.51, -4.65]][[-92.27, -92.27, -92.27]][[-3.59, -11.12]]
std:[[0.09, 0.07, 0.06]][[0.12, 0.1, 0.17]][[0.0, 0.0, 0.0]][[0.04, 0.12]]

MSE:[[3.07, 3.17, 3.49]][[4.28, 4.51, 4.65]][[92.27, 92.27, 92.27]][[3.59, 11.12]]
MSE(-DR):[[0.0, 0.1, 0.42]][[1.21, 1.44, 1.58]][[89.2, 89.2, 89.2]][[0.52, 8.05]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = 2.17
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.85, -0.83, -0.79]][[-1.76, -1.79, -1.84]][[-2.17, -2.17, -2.17]][-0.77]
std:[[0.19, 0.18, 0.15]][[0.01, 0.01, 0.01]][[0.0, 0.0, 0.0]][0.15]

MSE:[[0.87, 0.85, 0.8]][[1.76, 1.79, 1.84]][[2.17, 2.17, 2.17]][0.78]
MSE(-DR):[[0.0, -0.02, -0.07]][[0.89, 0.92, 0.97]][[1.3, 1.3, 1.3]][-0.09]
better than DR_NO_MARL
==============
time spent until now: 8.0 mins


---------------------------------------
[pattern_seed, T, sd_R] = [2, 672, 5]

max(u_O) =  157.3
O_threshold = 100
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:

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 = 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 = 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 1 2 3 4 5
--------------------------------------
Value of Behaviour policy:79.848
O_threshold = 100
MC for this TARGET:[86.009, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.23, -1.32, -1.52]][[-2.7, -2.84, -2.95]][[-86.01, -86.01, -86.01]][[-1.61, -6.16]]
std:[[0.22, 0.19, 0.17]][[0.22, 0.21, 0.24]][[0.0, 0.0, 0.0]][[0.14, 0.02]]

MSE:[[1.25, 1.33, 1.53]][[2.71, 2.85, 2.96]][[86.01, 86.01, 86.01]][[1.62, 6.16]]
MSE(-DR):[[0.0, 0.08, 0.28]][[1.46, 1.6, 1.71]][[84.76, 84.76, 84.76]][[0.37, 4.91]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
==============
O_threshold = 90
MC for this TARGET:[86.894, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.16, -0.26, -0.67]][[-0.44, -0.6, -0.71]][[-86.89, -86.89, -86.89]][[-0.77, -7.05]]
std:[[0.17, 0.16, 0.06]][[0.16, 0.14, 0.15]][[0.0, 0.0, 0.0]][[0.06, 0.02]]

MSE:[[0.23, 0.31, 0.67]][[0.47, 0.62, 0.73]][[86.89, 86.89, 86.89]][[0.77, 7.05]]
MSE(-DR):[[0.0, 0.08, 0.44]][[0.24, 0.39, 0.5]][[86.66, 86.66, 86.66]][[0.54, 6.82]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = 0.89
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.06, 1.06, 0.85]][[2.26, 2.24, 2.24]][[-0.89, -0.89, -0.89]][0.84]
std:[[0.05, 0.03, 0.1]][[0.07, 0.06, 0.09]][[0.0, 0.0, 0.0]][0.08]

MSE:[[1.06, 1.06, 0.86]][[2.26, 2.24, 2.24]][[0.89, 0.89, 0.89]][0.84]
MSE(-DR):[[0.0, 0.0, -0.2]][[1.2, 1.18, 1.18]][[-0.17, -0.17, -0.17]][-0.22]
better than DR_NO_MARL
==============
```

```
O_threshold = 95
MC for this TARGET:[85.653, 0.041]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.64, -0.72, -0.93]][[-0.64, -0.76, -0.88]][[-85.65, -85.65, -85.65]][[-1.01, -5.81]]
std:[[0.14, 0.15, 0.03]][[0.18, 0.18, 0.17]][[0.0, 0.0, 0.0]][[0.04, 0.02]]

MSE:[[0.66, 0.74, 0.93]][[0.66, 0.78, 0.9]][[85.65, 85.65, 85.65]][[1.01, 5.81]]
MSE(-DR):[[0.0, 0.08, 0.27]][[0.0, 0.12, 0.24]][[84.99, 84.99, 84.99]][[0.35, 5.15]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -0.36
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.58, 0.59, 0.58]][[2.07, 2.08, 2.07]][[0.36, 0.36, 0.36]][0.59]
std:[[0.08, 0.05, 0.13]][[0.04, 0.03, 0.07]][[0.0, 0.0, 0.0]][0.1]

MSE:[[0.59, 0.59, 0.59]][[2.07, 2.08, 2.07]][[0.36, 0.36, 0.36]][0.6]
MSE(-DR):[[0.0, 0.0, 0.0]][[1.48, 1.49, 1.48]][[-0.23, -0.23, -0.23]][0.01]
***** BETTER THAN [IS, DR_NO_MARL] *****
==============
O_threshold = 105
MC for this TARGET:[85.977, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.45, -2.53, -2.77]][[-4.0, -4.12, -4.22]][[-85.98, -85.98, -85.98]][[-2.85, -6.13]]
std:[[0.13, 0.13, 0.02]][[0.23, 0.22, 0.26]][[0.0, 0.0, 0.0]][[0.03, 0.02]]

MSE:[[2.45, 2.53, 2.77]][[4.01, 4.13, 4.23]][[85.98, 85.98, 85.98]][[2.85, 6.13]]
MSE(-DR):[[0.0, 0.08, 0.32]][[1.56, 1.68, 1.78]][[83.53, 83.53, 83.53]][[0.4, 3.68]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -0.03
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.22, -1.21, -1.25]][[-1.3, -1.28, -1.27]][[0.03, 0.03, 0.03]][-1.24]
std:[[0.35, 0.32, 0.14]][[0.0, 0.01, 0.02]][[0.0, 0.0, 0.0]][0.11]

MSE:[[1.27, 1.25, 1.26]][[1.3, 1.28, 1.27]][[0.03, 0.03, 0.03]][1.24]
MSE(-DR):[[0.0, -0.02, -0.01]][[0.03, 0.01, 0.0]][[-1.24, -1.24, -1.24]][-0.03]
better than DR_NO_MARL
==============
O_threshold = 110
MC for this TARGET:[85.977, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.44, -2.53, -2.77]][[-4.0, -4.12, -4.22]][[-85.98, -85.98, -85.98]][[-2.85, -6.13]]
std:[[0.11, 0.13, 0.05]][[0.23, 0.22, 0.26]][[0.0, 0.0, 0.0]][[0.03, 0.02]]

MSE:[[2.44, 2.53, 2.77]][[4.01, 4.13, 4.23]][[85.98, 85.98, 85.98]][[2.85, 6.13]]
MSE(-DR):[[0.0, 0.09, 0.33]][[1.57, 1.69, 1.79]][[83.54, 83.54, 83.54]][[0.41, 3.69]]
***** BETTER THAN [QV, IS, DR_NO_MARL] *****
MC-based ATE = -0.03
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.22, -1.21, -1.25]][[-1.3, -1.28, -1.27]][[0.03, 0.03, 0.03]][-1.24]
std:[[0.33, 0.32, 0.12]][[0.01, 0.01, 0.02]][[0.0, 0.0, 0.0]][0.1]

MSE:[[1.26, 1.25, 1.26]][[1.3, 1.28, 1.27]][[0.03, 0.03, 0.03]][1.24]
MSE(-DR):[[0.0, -0.01, 0.0]][[0.04, 0.02, 0.01]][[-1.23, -1.23, -1.23]][-0.02]
***** BETTER THAN [IS, DR_NO_MARL] *****
==============
time spent until now: 11.9 mins


--------------------------------------
[pattern_seed, T, sd_R] = [3, 672, 5]

max(u_O) =  142.3
O_threshold = 100
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 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 = 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 = 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 5 1 2
```