```
Last login: Tue Mar 31 18:48:52 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-221-170-144.compute-1.amazonaws.com
^C
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-228-10-241.compute-1.amazonaws.com
Warning: Permanently added the ED25519 host key for IP address '3.228.10.241' 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 Tue Mar 31 23:52:45 UTC 2020

  System load:  1.31               Processes:           233
  Usage of /:   56.4% of 15.45GB   Users logged in:     0
  Memory usage: 1%                 IP address for ens5: 172.31.14.85
  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-172-31-14-85:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
  File "EC2.py", line 48
    for sd_OD in [.5, 5, 10, 20]
                               ^
SyntaxError: invalid syntax
ubuntu@ip-172-31-14-85:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
Traceback (most recent call last):
  File "EC2.py", line 5, in <module>
    from simu_funs import *
  File "/home/ubuntu/simu_funs.py", line 6, in <module>
    from simu_DGP import *
  File "/home/ubuntu/simu_DGP.py", line 41
    if poisO = True:
             ^
SyntaxError: invalid syntax
ubuntu@ip-172-31-14-85:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
19:57, 03/31; num of cores:16

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, day_range, thre_ran
ge, poisO] = [None, 10, 10, 5, 0.2, 1, 1, 0.0001, False, True, 0, False, [3, 7, 14], [80, 90, 100, 110, 120], False]


--------------------------------------
[pattern_seed, sd_OD] = [0, 0.5]

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

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:
```

```
1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 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 = 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 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

--------------------------------------
Value of Behaviour policy:79.076
O_threshold = 80
MC for this TARGET:[88.835, 0.036]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.19, 0.15, -1.04]][[0.42, 0.35, -0.25]][[-88.84, -88.84, -88.84]][-9.76]
std:[[0.32, 0.33, 0.26]][[0.12, 0.12, 0.07]][[0.0, 0.0, 0.0]][0.02]
```

```
MSE:[[0.37, 0.36, 1.07]][[0.44, 0.37, 0.26]][[88.84, 88.84, 88.84]][9.76]
MSE(-DR):[[0.0, -0.01, 0.7]][[0.07, 0.0, -0.11]][[88.47, 88.47, 88.47]][9.39]
***
==============


O_threshold = 90
MC for this TARGET:[87.434, 0.037]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.87, 0.81, -1.2]][[1.58, 1.5, 0.58]][[-87.43, -87.43, -87.43]][-8.36]
std:[[0.24, 0.25, 0.22]][[0.14, 0.13, 0.1]][[0.0, 0.0, 0.0]][0.02]
MSE:[[0.9, 0.85, 1.22]][[1.59, 1.51, 0.59]][[87.43, 87.43, 87.43]][8.36]
MSE(-DR):[[0.0, -0.05, 0.32]][[0.69, 0.61, -0.31]][[86.53, 86.53, 86.53]][7.46]
***
MC-based ATE = -1.4
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.67, 0.66, -0.16]][[1.16, 1.14, 0.83]][[1.4, 1.4, 1.4]][1.4]
std:[[0.08, 0.08, 0.04]][[0.01, 0.01, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.67, 0.66, 0.16]][[1.16, 1.14, 0.83]][[1.4, 1.4, 1.4]][1.4]
MSE(-DR):[[0.0, -0.01, -0.51]][[0.49, 0.47, 0.16]][[0.73, 0.73, 0.73]][0.73]
**
==============


O_threshold = 100
MC for this TARGET:[91.774, 0.037]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.61, -1.68, -4.64]][[-1.16, -1.28, -2.7]][[-91.77, -91.77, -91.77]][-12.7]
std:[[0.08, 0.08, 0.04]][[0.13, 0.13, 0.04]][[0.0, 0.0, 0.0]][0.02]
MSE:[[1.61, 1.68, 4.64]][[1.17, 1.29, 2.7]][[91.77, 91.77, 91.77]][12.7]
MSE(-DR):[[0.0, 0.07, 3.03]][[-0.44, -0.32, 1.09]][[90.16, 90.16, 90.16]][11.09]
MC-based ATE = 2.94
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.81, -1.83, -3.6]][[-1.58, -1.63, -2.45]][[-2.94, -2.94, -2.94]][-2.94]
std:[[0.4, 0.41, 0.3]][[0.01, 0.01, 0.03]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.85, 1.88, 3.61]][[1.58, 1.63, 2.45]][[2.94, 2.94, 2.94]][2.94]
MSE(-DR):[[0.0, 0.03, 1.76]][[-0.27, -0.22, 0.6]][[1.09, 1.09, 1.09]][1.09]
==============


O_threshold = 110
MC for this TARGET:[88.749, 0.036]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.74, -1.8, -3.48]][[-2.53, -2.63, -4.26]][[-88.75, -88.75, -88.75]][-9.67]
std:[[0.07, 0.08, 0.07]][[0.09, 0.07, 0.05]][[0.0, 0.0, 0.0]][0.02]
MSE:[[1.74, 1.8, 3.48]][[2.53, 2.63, 4.26]][[88.75, 88.75, 88.75]][9.67]
MSE(-DR):[[0.0, 0.06, 1.74]][[0.79, 0.89, 2.52]][[87.01, 87.01, 87.01]][7.93]
***
MC-based ATE = -0.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.94, -1.95, -2.45]][[-2.95, -2.98, -4.02]][[0.09, 0.09, 0.09]][0.09]
std:[[0.25, 0.25, 0.33]][[0.04, 0.05, 0.03]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.96, 1.97, 2.47]][[2.95, 2.98, 4.02]][[0.09, 0.09, 0.09]][0.09]
MSE(-DR):[[0.0, 0.01, 0.51]][[0.99, 1.02, 2.06]][[-1.87, -1.87, -1.87]][-1.87]
*
==============


O_threshold = 120
MC for this TARGET:[90.87, 0.036]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.01, -7.04, -7.92]][[-7.42, -7.52, -9.13]][[-90.87, -90.87, -90.87]][-11.79]
std:[[0.08, 0.1, 0.02]][[0.1, 0.08, 0.0]][[0.0, 0.0, 0.0]][0.02]
MSE:[[7.01, 7.04, 7.92]][[7.42, 7.52, 9.13]][[90.87, 90.87, 90.87]][11.79]
MSE(-DR):[[0.0, 0.03, 0.91]][[0.41, 0.51, 2.12]][[83.86, 83.86, 83.86]][4.78]
***
MC-based ATE = 2.04
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.21, -7.19, -6.89]][[-7.84, -7.87, -8.89]][[-2.04, -2.04, -2.04]][-2.04]
std:[[0.24, 0.23, 0.28]][[0.03, 0.04, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.21, 7.19, 6.9]][[7.84, 7.87, 8.89]][[2.04, 2.04, 2.04]][2.04]
MSE(-DR):[[0.0, -0.02, -0.31]][[0.63, 0.66, 1.68]][[-5.17, -5.17, -5.17]][-5.17]
**
==============


Traceback (most recent call last):
  File "EC2.py", line 79, in <module>
```

```
    res_real.append(arr([a[2] for a in N_targets]))
NameError: name 'N_targets' is not defined
ubuntu@ip-172-31-14-85:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py

20:05, 03/31; num of cores:16

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, day_range, thre_ran
ge, poisO] = [None, 10, 10, 5, 0.2, 1, 1, 0.0001, False, True, 0, False, [3, 7, 14], [80, 90, 100, 110, 120], False]


----------------------------------------
[pattern_seed, sd_OD] = [0, 0.5]

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

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 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 = 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 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

---------------------------------------
Value of Behaviour policy:79.076
O_threshold = 80
MC for this TARGET:[88.835, 0.036]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.19, 0.15, -1.03]][[0.42, 0.35, -0.23]][[-88.84, -88.84, -88.84]][-9.76]
std:[[0.32, 0.33, 0.27]][[0.14, 0.12, 0.1]]][[0.0, 0.0, 0.0]][0.02]
MSE:[[0.37, 0.36, 1.06]][[0.44, 0.37, 0.25]][[88.84, 88.84, 88.84]][9.76]
MSE(-DR):[[0.0, -0.01, 0.69]][[0.07, 0.0, -0.12]][[88.47, 88.47, 88.47]][9.39]
***
==============


O_threshold = 90
MC for this TARGET:[87.434, 0.037]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.86, 0.81, -1.21]][[1.58, 1.5, 0.61]][[-87.43, -87.43, -87.43]][-8.36]
std:[[0.23, 0.25, 0.21]][[0.15, 0.13, 0.12]]][[0.0, 0.0, 0.0]][0.02]
MSE:[[0.89, 0.85, 1.23]][[1.59, 1.51, 0.62]][[87.43, 87.43, 87.43]][8.36]
MSE(-DR):[[0.0, -0.04, 0.34]][[0.7, 0.62, -0.27]][[86.54, 86.54, 86.54]][7.47]
***
MC-based ATE = -1.4
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.66, 0.66, -0.18]][[1.16, 1.14, 0.83]][[1.4, 1.4, 1.4]][1.4]
std:[[0.09, 0.08, 0.05]][[0.01, 0.01, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.67, 0.66, 0.19]][[1.16, 1.14, 0.83]][[1.4, 1.4, 1.4]][1.4]
MSE(-DR):[[0.0, -0.01, -0.48]][[0.49, 0.47, 0.16]][[0.73, 0.73, 0.73]][0.73]
**
==============


O_threshold = 100
MC for this TARGET:[91.774, 0.037]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.62, -1.68, -4.66]][[-1.15, -1.28, -2.67]][[-91.77, -91.77, -91.77]][-12.7]
std:[[0.07, 0.08, 0.04]][[0.15, 0.13, 0.09]][[0.0, 0.0, 0.0]][0.02]
MSE:[[1.62, 1.68, 4.66]][[1.16, 1.29, 2.67]][[91.77, 91.77, 91.77]][12.7]
MSE(-DR):[[0.0, 0.06, 3.04]][[-0.46, -0.33, 1.05]][[90.15, 90.15, 90.15]][11.08]
MC-based ATE = 2.94
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.81, -1.83, -3.63]][[-1.57, -1.63, -2.44]][[-2.94, -2.94, -2.94]][-2.94]
std:[[0.39, 0.41, 0.31]][[0.01, 0.01, 0.01]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.85, 1.88, 3.64]][[1.57, 1.63, 2.44]][[2.94, 2.94, 2.94]][2.94]
MSE(-DR):[[0.0, 0.03, 1.79]][[-0.28, -0.22, 0.59]][[1.09, 1.09, 1.09]][1.09]
==============


O_threshold = 110
MC for this TARGET:[88.749, 0.036]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.76, -1.8, -3.48]][[-2.52, -2.63, -4.22]][[-88.75, -88.75, -88.75]][-9.67]
std:[[0.09, 0.08, 0.04]][[0.09, 0.07, 0.0]][[0.0, 0.0, 0.0]][0.02]
MSE:[[1.76, 1.8, 3.48]][[2.52, 2.63, 4.22]][[88.75, 88.75, 88.75]][9.67]
MSE(-DR):[[0.0, 0.04, 1.72]][[0.76, 0.87, 2.46]][[86.99, 86.99, 86.99]][7.91]
***
MC-based ATE = -0.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

```
bias:[[-1.95, -1.95, -2.45]][[-2.94, -2.98, -3.99]][[0.09, 0.09, 0.09]][0.09]
std:[[0.23, 0.25, 0.31]][[0.05, 0.05, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.96, 1.97, 2.47]][[2.94, 2.98, 3.99]][[0.09, 0.09, 0.09]][0.09]
MSE(-DR):[[0.0, 0.01, 0.51]][[0.98, 1.02, 2.03]][[-1.87, -1.87, -1.87]][-1.87]
*
==============


O_threshold = 120
MC for this TARGET:[90.87, 0.036]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.01, -7.04, -7.93]][[-7.43, -7.52, -9.18]][[-90.87, -90.87, -90.87]][-11.79]
std:[[0.08, 0.1, 0.08]][[0.11, 0.08, 0.03]][[0.0, 0.0, 0.0]][0.02]
MSE:[[7.01, 7.04, 7.93]][[7.43, 7.52, 9.18]][[90.87, 90.87, 90.87]][11.79]
MSE(-DR):[[0.0, 0.03, 0.92]][[0.42, 0.51, 2.17]][[83.86, 83.86, 83.86]][4.78]
***
MC-based ATE = 2.04
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.2, -7.19, -6.91]][[-7.85, -7.87, -8.95]][[-2.04, -2.04, -2.04]][-2.04]
std:[[0.24, 0.23, 0.34]][[0.04, 0.04, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.2, 7.19, 6.92]][[7.85, 7.87, 8.95]][[2.04, 2.04, 2.04]][2.04]
MSE(-DR):[[0.0, -0.01, -0.28]][[0.65, 0.67, 1.75]][[-5.16, -5.16, -5.16]][-5.16]
**
==============


[array([[ 0.37,  0.36,  1.06,  0.44,  0.37,  0.25, 88.84, 88.84, 88.84,
         9.76],
       [ 0.89,  0.85,  1.23,  1.59,  1.51,  0.62, 87.43, 87.43, 87.43,
         8.36],
       [ 1.62,  1.68,  4.66,  1.16,  1.29,  2.67, 91.77, 91.77, 91.77,
        12.7 ],
       [ 1.76,  1.8 ,  3.48,  2.52,  2.63,  4.22, 88.75, 88.75, 88.75,
         9.67],
       [ 7.01,  7.04,  7.93,  7.43,  7.52,  9.18, 90.87, 90.87, 90.87,
        11.79]])]
time spent until now: 6.0 mins


--------------------------------------
[pattern_seed, sd_OD] = [0, 5]

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

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 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 = 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 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

--------------------------------------
Value of Behaviour policy:79.126
O_threshold = 80
MC for this TARGET:[88.8, 0.041]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.33, 0.27, -0.61]][[0.41, 0.32, -0.17]][[-88.8, -88.8, -88.8]][-9.67]
std:[[0.08, 0.08, 0.19]][[0.13, 0.1, 0.12]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.34, 0.28, 0.64]][[0.43, 0.34, 0.21]][[88.8, 88.8, 88.8]][9.67]
MSE(-DR):[[0.0, -0.06, 0.3]][[0.09, 0.0, -0.13]][[88.46, 88.46, 88.46]][9.33]
***
==============


O_threshold = 90
MC for this TARGET:[87.325, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.16, 1.12, -0.71]][[1.58, 1.49, 0.68]][[-87.32, -87.32, -87.32]][-8.2]
std:[[0.14, 0.13, 0.19]][[0.17, 0.17, 0.18]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.17, 1.13, 0.73]][[1.59, 1.5, 0.7]][[87.32, 87.32, 87.32]][8.2]
MSE(-DR):[[0.0, -0.04, -0.44]][[0.42, 0.33, -0.47]][[86.15, 86.15, 86.15]][7.03]
**
MC-based ATE = -1.47
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.83, 0.85, -0.1]][[1.17, 1.17, 0.84]][[1.47, 1.47, 1.47]][1.47]
std:[[0.06, 0.05, 0.0]][[0.05, 0.07, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.83, 0.85, 0.1]][[1.17, 1.17, 0.84]][[1.47, 1.47, 1.47]][1.47]
MSE(-DR):[[0.0, 0.02, -0.73]][[0.34, 0.34, 0.01]][[0.64, 0.64, 0.64]][0.64]
**
```

==========

O_threshold = 100
MC for this TARGET:[91.569, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.38, -1.45, -4.5]][[-1.11, -1.25, -2.48]][[-91.57, -91.57, -91.57]][-12.44]
std:[[0.25, 0.24, 0.17]][[0.19, 0.17, 0.26]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.4, 1.47, 4.5]][[1.13, 1.26, 2.49]][[91.57, 91.57, 91.57]][12.44]
MSE(-DR):[[0.0, 0.07, 3.1]][[-0.27, -0.14, 1.09]][[90.17, 90.17, 90.17]][11.04]
MC-based ATE = 2.77
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.71, -1.72, -3.89]][[-1.52, -1.57, -2.32]][[-2.77, -2.77, -2.77]][-2.77]
std:[[0.16, 0.16, 0.02]][[0.06, 0.07, 0.14]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.72, 1.73, 3.89]][[1.52, 1.57, 2.32]][[2.77, 2.77, 2.77]][2.77]
MSE(-DR):[[0.0, 0.01, 2.17]][[-0.2, -0.15, 0.6]][[1.05, 1.05, 1.05]][1.05]
==========

O_threshold = 110
MC for this TARGET:[88.701, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.71, -1.79, -3.37]][[-2.52, -2.63, -4.14]][[-88.7, -88.7, -88.7]][-9.58]
std:[[0.2, 0.21, 0.03]][[0.13, 0.13, 0.22]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.72, 1.8, 3.37]][[2.52, 2.63, 4.15]][[88.7, 88.7, 88.7]][9.58]
MSE(-DR):[[0.0, 0.08, 1.65]][[0.8, 0.91, 2.43]][[86.98, 86.98, 86.98]][7.86]
***
MC-based ATE = -0.1
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.04, -2.06, -2.76]][[-2.93, -2.94, -3.98]][[0.1, 0.1, 0.1]][0.1]
std:[[0.28, 0.3, 0.16]][[0.0, 0.03, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.06, 2.08, 2.76]][[2.93, 2.94, 3.98]][[0.1, 0.1, 0.1]][0.1]
MSE(-DR):[[0.0, 0.02, 0.7]][[0.87, 0.88, 1.92]][[-1.96, -1.96, -1.96]][-1.96]
*
==========

O_threshold = 120
MC for this TARGET:[90.814, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.03, -7.07, -7.84]][[-7.43, -7.51, -9.01]][[-90.81, -90.81, -90.81]][-11.69]
std:[[0.39, 0.39, 0.15]][[0.2, 0.2, 0.27]][[0.0, 0.0, 0.0]][0.06]
MSE:[[7.04, 7.08, 7.84]][[7.43, 7.51, 9.01]][[90.81, 90.81, 90.81]][11.69]
MSE(-DR):[[0.0, 0.04, 0.8]][[0.39, 0.47, 1.97]][[83.77, 83.77, 83.77]][4.65]
***
MC-based ATE = 2.01
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.35, -7.34, -7.23]][[-7.83, -7.83, -8.84]][[-2.01, -2.01, -2.01]][-2.01]
std:[[0.47, 0.48, 0.34]][[0.07, 0.1, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.37, 7.36, 7.24]][[7.83, 7.83, 8.84]][[2.01, 2.01, 2.01]][2.01]
MSE(-DR):[[0.0, -0.01, -0.13]][[0.46, 0.46, 1.47]][[-5.36, -5.36, -5.36]][-5.36]
**
==========

[array([[0.37, 0.36, 1.06, 0.44, 0.37, 0.25, 88.84, 88.84, 88.84,
         9.76],
       [0.89, 0.85, 1.23, 1.59, 1.51, 0.62, 87.43, 87.43, 87.43,
         8.36],
       [1.62, 1.68, 4.66, 1.16, 1.29, 2.67, 91.77, 91.77, 91.77,
        12.7],
       [1.76, 1.8 , 3.48, 2.52, 2.63, 4.22, 88.75, 88.75, 88.75,
         9.67],
       [7.01, 7.04, 7.93, 7.43, 7.52, 9.18, 90.87, 90.87, 90.87,
        11.79]]), array([[0.34, 0.28, 0.64, 0.43, 0.34, 0.21, 88.8 , 88.8 , 88.8 ,
         9.67],
       [1.17, 1.13, 0.73, 1.59, 1.5 , 0.7 , 87.32, 87.32, 87.32,
         8.2],
       [1.4 , 1.47, 4.5 , 1.13, 1.26, 2.49, 91.57, 91.57, 91.57,
        12.44],
       [1.72, 1.8 , 3.37, 2.52, 2.63, 4.15, 88.7 , 88.7 , 88.7 ,
         9.58],
       [7.04, 7.08, 7.84, 7.43, 7.51, 9.01, 90.81, 90.81, 90.81,
        11.69]])]
time spent until now: 12.0 mins

```
[pattern_seed, sd_OD] = [0, 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 = 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 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:78.908
O_threshold = 80
MC for this TARGET:[88.737, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.01, -0.03, -0.95]][[0.54, 0.46, -0.15]][[-88.74, -88.74, -88.74]][-9.83]
std:[[0.27, 0.24, 0.25]][[0.09, 0.12, 0.15]][[0.0, 0.0, 0.0]][0.11]
MSE:[[0.27, 0.24, 0.98]][[0.55, 0.48, 0.21]][[88.74, 88.74, 88.74]][9.83]
MSE(-DR):[[0.0, -0.03, 0.71]][[0.28, 0.21, -0.06]][[88.47, 88.47, 88.47]][9.56]
***
==============


O_threshold = 90
MC for this TARGET:[87.229, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.02, 0.97, -0.81]][[1.77, 1.68, 0.79]][[-87.23, -87.23, -87.23]][-8.32]
std:[[0.23, 0.21, 0.24]][[0.12, 0.13, 0.15]][[0.0, 0.0, 0.0]][0.11]
MSE:[[1.05, 0.99, 0.84]][[1.77, 1.69, 0.8]][[87.23, 87.23, 87.23]][8.32]
MSE(-DR):[[0.0, -0.06, -0.21]][[0.72, 0.64, -0.25]][[86.18, 86.18, 86.18]][7.27]
**
MC-based ATE = -1.51
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.02, 1.0, 0.14]][[1.23, 1.22, 0.94]][[1.51, 1.51, 1.51]][1.51]
std:[[0.04, 0.02, 0.01]][[0.03, 0.01, 0.0]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.02, 1.0, 0.14]][[1.23, 1.22, 0.94]][[1.51, 1.51, 1.51]][1.51]
MSE(-DR):[[0.0, -0.02, -0.88]][[0.21, 0.2, -0.08]][[0.49, 0.49, 0.49]][0.49]
**
==============


O_threshold = 100
MC for this TARGET:[91.412, 0.041]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.36, -1.43, -4.3]][[-0.82, -0.95, -2.33]][[-91.41, -91.41, -91.41]][-12.5]
std:[[0.18, 0.19, 0.11]][[0.16, 0.15, 0.18]][[0.0, 0.0, 0.0]][0.11]
MSE:[[1.37, 1.44, 4.3]][[0.84, 0.96, 2.34]][[91.41, 91.41, 91.41]][12.5]
MSE(-DR):[[0.0, 0.07, 2.93]][[-0.53, -0.41, 0.97]][[90.04, 90.04, 90.04]][11.13]
MC-based ATE = 2.68
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.37, -1.4, -3.35]][[-1.36, -1.41, -2.18]][[-2.68, -2.68, -2.68]][-2.68]
std:[[0.09, 0.05, 0.14]][[0.07, 0.04, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.37, 1.4, 3.35]][[1.36, 1.41, 2.18]][[2.68, 2.68, 2.68]][2.68]
MSE(-DR):[[0.0, 0.03, 1.98]][[-0.01, 0.04, 0.81]][[1.31, 1.31, 1.31]][1.31]
==============


O_threshold = 110
MC for this TARGET:[88.655, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.75, -1.82, -3.38]][[-2.48, -2.6, -4.16]][[-88.66, -88.66, -88.66]][-9.75]
std:[[0.63, 0.6, 0.32]][[0.16, 0.17, 0.13]][[0.0, 0.0, 0.0]][0.11]
MSE:[[1.86, 1.92, 3.4]][[2.49, 2.61, 4.16]][[88.66, 88.66, 88.66]][9.75]
MSE(-DR):[[0.0, 0.06, 1.54]][[0.63, 0.75, 2.3]][[86.8, 86.8, 86.8]][7.89]
***
MC-based ATE = -0.08
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.76, -1.79, -2.43]][[-3.02, -3.06, -4.01]][[0.08, 0.08, 0.08]][0.08]
std:[[0.36, 0.37, 0.07]][[0.07, 0.05, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.8, 1.83, 2.43]][[3.02, 3.06, 4.01]][[0.08, 0.08, 0.08]][0.08]
MSE(-DR):[[0.0, 0.03, 0.63]][[1.22, 1.26, 2.21]][[-1.72, -1.72, -1.72]][-1.72]
*
==============


O_threshold = 120
MC for this TARGET:[90.724, 0.038]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

```
bias:[[-7.01, -7.05, -8.0]][[-7.46, -7.56, -9.09]][[-90.72, -90.72, -90.72]][-11.82]
std:[[0.4, 0.41, 0.22]][[0.19, 0.21, 0.13]][[0.0, 0.0, 0.0]][0.11]
MSE:[[7.02, 7.06, 8.0]][[7.46, 7.56, 9.09]][[90.72, 90.72, 90.72]][11.82]
MSE(-DR):[[0.0, 0.04, 0.98]][[0.44, 0.54, 2.07]][[83.7, 83.7, 83.7]][4.8]
***
MC-based ATE = 1.99
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.01, -7.02, -7.06]][[-8.0, -8.02, -8.94]][[-1.99, -1.99, -1.99]][-1.99]
std:[[0.14, 0.18, 0.03]][[0.1, 0.09, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.01, 7.02, 7.06]][[8.0, 8.02, 8.94]][[1.99, 1.99, 1.99]][1.99]
MSE(-DR):[[0.0, 0.01, 0.05]][[0.99, 1.01, 1.93]][[-5.02, -5.02, -5.02]][-5.02]
*
==============


[array([[ 0.37,  0.36,  1.06,  0.44,  0.37,  0.25, 88.84, 88.84, 88.84,
         9.76],
       [ 0.89,  0.85,  1.23,  1.59,  1.51,  0.62, 87.43, 87.43, 87.43,
         8.36],
       [ 1.62,  1.68,  4.66,  1.16,  1.29,  2.67, 91.77, 91.77, 91.77,
        12.7 ],
       [ 1.76,  1.8 ,  3.48,  2.52,  2.63,  4.22, 88.75, 88.75, 88.75,
         9.67],
       [ 7.01,  7.04,  7.93,  7.43,  7.52,  9.18, 90.87, 90.87, 90.87,
        11.79]]), array([[ 0.34,  0.28,  0.64,  0.43,  0.34,  0.21, 88.8 , 88.8 , 88.8 ,
         9.67],
       [ 1.17,  1.13,  0.73,  1.59,  1.5 ,  0.7 , 87.32, 87.32, 87.32,
         8.2 ],
       [ 1.4 ,  1.47,  4.5 ,  1.13,  1.26,  2.49, 91.57, 91.57, 91.57,
        12.44],
       [ 1.72,  1.8 ,  3.37,  2.52,  2.63,  4.15, 88.7 , 88.7 , 88.7 ,
         9.58],
       [ 7.04,  7.08,  7.84,  7.43,  7.51,  9.01, 90.81, 90.81, 90.81,
        11.69]]), array([[ 0.27,  0.24,  0.98,  0.55,  0.48,  0.21, 88.74, 88.74, 88.74,
         9.83],
       [ 1.05,  0.99,  0.84,  1.77,  1.69,  0.8 , 87.23, 87.23, 87.23,
         8.32],
       [ 1.37,  1.44,  4.3 ,  0.84,  0.96,  2.34, 91.41, 91.41, 91.41,
        12.5 ],
       [ 1.86,  1.92,  3.4 ,  2.49,  2.61,  4.16, 88.66, 88.66, 88.66,
         9.75],
       [ 7.02,  7.06,  8.  ,  7.46,  7.56,  9.09, 90.72, 90.72, 90.72,
        11.82]])]
time spent until now: 18.0 mins


--------------------------------------
[pattern_seed, sd_OD] = [0, 20]

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 = 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 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

---------------------------------------
Value of Behaviour policy:78.737
O_threshold = 80
MC for this TARGET:[88.63, 0.042]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.05, 0.01, -0.75]][[0.37, 0.29, -0.21]][[-88.63, -88.63, -88.63]][-9.89]
std:[[0.04, 0.02, 0.28]][[0.12, 0.11, 0.11]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.06, 0.02, 0.8]][[0.39, 0.31, 0.24]][[88.63, 88.63, 88.63]][9.89]
MSE(-DR):[[0.0, -0.04, 0.74]][[0.33, 0.25, 0.18]][[88.57, 88.57, 88.57]][9.83]
***
==============


O_threshold = 90
MC for this TARGET:[87.075, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.86, 0.82, -0.7]][[1.51, 1.44, 0.7]][[-87.08, -87.08, -87.08]][-8.34]
std:[[0.08, 0.09, 0.21]][[0.05, 0.05, 0.0]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.86, 0.82, 0.73]][[1.51, 1.44, 0.7]][[87.08, 87.08, 87.08]][8.34]
MSE(-DR):[[0.0, -0.04, -0.13]][[0.65, 0.58, -0.16]][[86.22, 86.22, 86.22]][7.48]
**
```

```
MC-based ATE = -1.55
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.8, 0.82, 0.05]][[1.13, 1.15, 0.91]][[1.55, 1.55, 1.55]][1.55]
std:[[0.12, 0.11, 0.08]][[0.07, 0.06, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.81, 0.83, 0.09]][[1.13, 1.15, 0.92]][[1.55, 1.55, 1.55]][1.55]
MSE(-DR):[[0.0, 0.02, -0.72]][[0.32, 0.34, 0.11]][[0.74, 0.74, 0.74]][0.74]
**
==============


O_threshold = 100
MC for this TARGET:[91.162, 0.042]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.57, -1.63, -4.35]][[-1.17, -1.29, -2.46]][[-91.16, -91.16, -91.16]][-12.43]
std:[[0.19, 0.18, 0.1]][[0.06, 0.05, 0.01]][[0.0, 0.0, 0.0]][0.05]
MSE:[[1.58, 1.64, 4.35]][[1.17, 1.29, 2.46]][[91.16, 91.16, 91.16]][12.43]
MSE(-DR):[[0.0, 0.06, 2.77]][[-0.41, -0.29, 0.88]][[89.58, 89.58, 89.58]][10.85]
MC-based ATE = 2.53
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.62, -1.64, -3.6]][[-1.55, -1.58, -2.24]][[-2.53, -2.53, -2.53]][-2.53]
std:[[0.15, 0.16, 0.19]][[0.06, 0.07, 0.12]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.63, 1.65, 3.61]][[1.55, 1.58, 2.24]][[2.53, 2.53, 2.53]][2.53]
MSE(-DR):[[0.0, 0.02, 1.98]][[-0.08, -0.05, 0.61]][[0.9, 0.9, 0.9]][0.9]
==============


O_threshold = 110
MC for this TARGET:[88.564, 0.041]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.16, -2.23, -3.63]][[-2.9, -2.99, -4.41]][[-88.56, -88.56, -88.56]][-9.83]
std:[[0.07, 0.06, 0.04]][[0.12, 0.11, 0.04]][[0.0, 0.0, 0.0]][0.05]
MSE:[[2.16, 2.23, 3.63]][[2.9, 2.99, 4.41]][[88.56, 88.56, 88.56]][9.83]
MSE(-DR):[[0.0, 0.07, 1.47]][[0.74, 0.83, 2.25]][[86.4, 86.4, 86.4]][7.67]
***
MC-based ATE = -0.07
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.22, -2.23, -2.88]][[-3.27, -3.28, -4.2]][[0.07, 0.07, 0.07]][0.07]
std:[[0.11, 0.08, 0.32]][[0.0, 0.0, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.22, 2.23, 2.9]][[3.27, 3.28, 4.2]][[0.07, 0.07, 0.07]][0.07]
MSE(-DR):[[0.0, 0.01, 0.68]][[1.05, 1.06, 1.98]][[-2.15, -2.15, -2.15]][-2.15]
*
==============


O_threshold = 120
MC for this TARGET:[90.539, 0.045]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.26, -7.31, -8.08]][[-7.69, -7.75, -9.18]][[-90.54, -90.54, -90.54]][-11.8]
std:[[0.11, 0.11, 0.04]][[0.06, 0.07, 0.03]][[0.0, 0.0, 0.0]][0.05]
MSE:[[7.26, 7.31, 8.08]][[7.69, 7.75, 9.18]][[90.54, 90.54, 90.54]][11.8]
MSE(-DR):[[0.0, 0.05, 0.82]][[0.43, 0.49, 1.92]][[83.28, 83.28, 83.28]][4.54]
***
MC-based ATE = 1.91
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.32, -7.32, -7.33]][[-8.06, -8.04, -8.97]][[-1.91, -1.91, -1.91]][-1.91]
std:[[0.15, 0.13, 0.33]][[0.06, 0.05, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.32, 7.32, 7.34]][[8.06, 8.04, 8.97]][[1.91, 1.91, 1.91]][1.91]
MSE(-DR):[[0.0, 0.0, 0.02]][[0.74, 0.72, 1.65]][[-5.41, -5.41, -5.41]][-5.41]
*
==============


[array([[ 0.37,  0.36,  1.06,  0.44,  0.37,  0.25, 88.84, 88.84, 88.84,
         9.76],
       [ 0.89,  0.85,  1.23,  1.59,  1.51,  0.62, 87.43, 87.43, 87.43,
         8.36],
       [ 1.62,  1.68,  4.66,  1.16,  1.29,  2.67, 91.77, 91.77, 91.77,
        12.7 ],
       [ 1.76,  1.8 ,  3.48,  2.52,  2.63,  4.22, 88.75, 88.75, 88.75,
         9.67],
       [ 7.01,  7.04,  7.93,  7.43,  7.52,  9.18, 90.87, 90.87, 90.87,
        11.79]]), array([[ 0.34,  0.28,  0.64,  0.43,  0.34,  0.21, 88.8 , 88.8 , 88.8 ,
         9.67],
       [ 1.17,  1.13,  0.73,  1.59,  1.5 ,  0.7 , 87.32, 87.32, 87.32,
         8.2 ],
       [ 1.4 ,  1.47,  4.5 ,  1.13,  1.26,  2.49, 91.57, 91.57, 91.57,
        12.44],
       [ 1.72,  1.8 ,  3.37,  2.52,  2.63,  4.15, 88.7 , 88.7 , 88.7 ,
```

```
         9.58],
        [ 7.04,  7.08,  7.84,  7.43,  7.51,  9.01, 90.81, 90.81, 90.81,
         11.69]]), array([[ 0.27,  0.24,  0.98,  0.55,  0.48,  0.21, 88.74, 88.74, 88.74,
          9.83],
        [ 1.05,  0.99,  0.84,  1.77,  1.69,  0.8 , 87.23, 87.23, 87.23,
          8.32],
        [ 1.37,  1.44,  4.3 ,  0.84,  0.96,  2.34, 91.41, 91.41, 91.41,
         12.5 ],
        [ 1.86,  1.92,  3.4 ,  2.49,  2.61,  4.16, 88.66, 88.66, 88.66,
          9.75],
        [ 7.02,  7.06,  8.  ,  7.46,  7.56,  9.09, 90.72, 90.72, 90.72,
         11.82]]), array([[6.000e-02, 2.000e-02, 8.000e-01, 3.900e-01, 3.100e-01, 2.400e-01,
         8.863e+01, 8.863e+01, 8.863e+01, 9.890e+00],
        [8.600e-01, 8.200e-01, 7.300e-01, 1.510e+00, 1.440e+00, 7.000e-01,
         8.708e+01, 8.708e+01, 8.708e+01, 8.340e+00],
        [1.580e+00, 1.640e+00, 4.350e+00, 1.170e+00, 1.290e+00, 2.460e+00,
         9.116e+01, 9.116e+01, 9.116e+01, 1.243e+01],
        [2.160e+00, 2.230e+00, 3.630e+00, 2.900e+00, 2.990e+00, 4.410e+00,
         8.856e+01, 8.856e+01, 8.856e+01, 9.830e+00],
        [7.260e+00, 7.310e+00, 8.080e+00, 7.690e+00, 7.750e+00, 9.180e+00,
         9.054e+01, 9.054e+01, 9.054e+01, 1.180e+01]])]
time spent until now: 24.0 mins


---------------------------------------
[pattern_seed, sd_OD] = [1, 0.5]

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 = 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 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

---------------------------------------
Value of Behaviour policy:71.547
O_threshold = 80
MC for this TARGET:[78.381, 0.036]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[2.06, 2.02, 0.3]][[2.35, 2.26, 1.59]][[-78.38, -78.38, -78.38]][-6.83]
std:[[0.1, 0.07, 0.0]][[0.05, 0.06, 0.02]][[0.0, 0.0, 0.0]][0.03]
MSE:[[2.06, 2.02, 0.3]][[2.35, 2.26, 1.59]][[78.38, 78.38, 78.38]][6.83]
MSE(-DR):[[0.0, -0.04, -1.76]][[0.29, 0.2, -0.47]][[76.32, 76.32, 76.32]][4.77]
**
==============


O_threshold = 90
MC for this TARGET:[79.717, 0.036]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.37, 0.31, -2.48]][[0.36, 0.25, -0.78]][[-79.72, -79.72, -79.72]][-8.17]
std:[[0.23, 0.21, 0.04]][[0.05, 0.04, 0.02]][[0.0, 0.0, 0.0]][0.03]
MSE:[[0.44, 0.37, 2.48]][[0.36, 0.25, 0.78]][[79.72, 79.72, 79.72]][8.17]
MSE(-DR):[[0.0, -0.07, 2.04]][[-0.08, -0.19, 0.34]][[79.28, 79.28, 79.28]][7.73]
MC-based ATE = 1.34
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.69, -1.72, -2.78]][[-1.99, -2.01, -2.37]][[-1.34, -1.34, -1.34]][-1.34]
std:[[0.13, 0.14, 0.04]][[0.0, 0.02, 0.0]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.69, 1.73, 2.78]][[1.99, 2.01, 2.37]][[1.34, 1.34, 1.34]][1.34]
MSE(-DR):[[0.0, 0.04, 1.09]][[0.3, 0.32, 0.68]][[-0.35, -0.35, -0.35]][-0.35]
*
==============


O_threshold = 100
MC for this TARGET:[84.426, 0.035]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.73, -3.81, -5.99]][[-4.12, -4.24, -5.8]][[-84.43, -84.43, -84.43]][-12.88]
std:[[0.11, 0.09, 0.02]][[0.08, 0.07, 0.05]][[0.0, 0.0, 0.0]][0.03]
MSE:[[3.73, 3.81, 5.99]][[4.12, 4.24, 5.8]][[84.43, 84.43, 84.43]][12.88]
MSE(-DR):[[0.0, 0.08, 2.26]][[0.39, 0.51, 2.07]][[80.7, 80.7, 80.7]][9.15]
***
MC-based ATE = 6.05
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.78, -5.83, -6.29]][[-6.47, -6.5, -7.39]][[-6.05, -6.05, -6.05]][-6.05]
std:[[0.02, 0.02, 0.02]][[0.03, 0.01, 0.03]][[0.0, 0.0, 0.0]][0.0]
MSE:[[5.78, 5.83, 6.29]][[6.47, 6.5, 7.39]][[6.05, 6.05, 6.05]][6.05]
```

```
MSE(-DR):[[0.0, 0.05, 0.51]][[0.69, 0.72, 1.61]][[0.27, 0.27, 0.27]][0.27]
*
==============


O_threshold = 110
MC for this TARGET:[88.018, 0.036]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.21, -6.28, -7.97]][[-7.13, -7.29, -9.18]][[-88.02, -88.02, -88.02]][-16.47]
std:[[0.13, 0.11, 0.07]][[0.12, 0.09, 0.01]][[0.0, 0.0, 0.0]][0.03]
MSE:[[6.21, 6.28, 7.97]][[7.13, 7.29, 9.18]][[88.02, 88.02, 88.02]][16.47]
MSE(-DR):[[0.0, 0.07, 1.76]][[0.92, 1.08, 2.97]][[81.81, 81.81, 81.81]][10.26]
***
MC-based ATE = 9.64
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.27, -8.31, -8.27]][[-9.48, -9.55, -10.77]][[-9.64, -9.64, -9.64]][-9.64]
std:[[0.03, 0.05, 0.06]][[0.07, 0.03, 0.01]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.27, 8.31, 8.27]][[9.48, 9.55, 10.77]][[9.64, 9.64, 9.64]][9.64]
MSE(-DR):[[0.0, 0.04, 0.0]][[1.21, 1.28, 2.5]][[1.37, 1.37, 1.37]][1.37]
*
==============


O_threshold = 120
MC for this TARGET:[83.813, 0.037]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.22, -8.26, -8.52]][[-9.21, -9.27, -10.97]][[-83.81, -83.81, -83.81]][-12.27]
std:[[0.1, 0.12, 0.11]][[0.16, 0.14, 0.03]][[0.0, 0.0, 0.0]][0.03]
MSE:[[8.22, 8.26, 8.52]][[9.21, 9.27, 10.97]][[83.81, 83.81, 83.81]][12.27]
MSE(-DR):[[0.0, 0.04, 0.3]][[0.99, 1.05, 2.75]][[75.59, 75.59, 75.59]][4.05]
***
MC-based ATE = 5.43
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.28, -10.28, -8.82]][[-11.56, -11.53, -12.56]][[-5.43, -5.43, -5.43]][-5.43]
std:[[0.2, 0.19, 0.1]][[0.11, 0.08, 0.01]][[0.0, 0.0, 0.0]][0.0]
MSE:[[10.28, 10.28, 8.82]][[11.56, 11.53, 12.56]][[5.43, 5.43, 5.43]][5.43]
MSE(-DR):[[0.0, 0.0, -1.46]][[1.28, 1.25, 2.28]][[-4.85, -4.85, -4.85]][-4.85]
**
==============


[array([[ 0.37,  0.36,  1.06,  0.44,  0.37,  0.25, 88.84, 88.84, 88.84,
         9.76],
       [ 0.89,  0.85,  1.23,  1.59,  1.51,  0.62, 87.43, 87.43, 87.43,
         8.36],
       [ 1.62,  1.68,  4.66,  1.16,  1.29,  2.67, 91.77, 91.77, 91.77,
        12.7 ],
       [ 1.76,  1.8 ,  3.48,  2.52,  2.63,  4.22, 88.75, 88.75, 88.75,
         9.67],
       [ 7.01,  7.04,  7.93,  7.43,  7.52,  9.18, 90.87, 90.87, 90.87,
        11.79]]), array([[ 0.34,  0.28,  0.64,  0.43,  0.34,  0.21, 88.8 , 88.8 , 88.8 ,
         9.67],
       [ 1.17,  1.13,  0.73,  1.59,  1.5 ,  0.7 , 87.32, 87.32, 87.32,
         8.2 ],
       [ 1.4 ,  1.47,  4.5 ,  1.13,  1.26,  2.49, 91.57, 91.57, 91.57,
        12.44],
       [ 1.72,  1.8 ,  3.37,  2.52,  2.63,  4.15, 88.7 , 88.7 , 88.7 ,
         9.58],
       [ 7.04,  7.08,  7.84,  7.43,  7.51,  9.01, 90.81, 90.81, 90.81,
        11.69]]), array([[ 0.27,  0.24,  0.98,  0.55,  0.48,  0.21, 88.74, 88.74, 88.74,
         9.83],
       [ 1.05,  0.99,  0.84,  1.77,  1.69,  0.8 , 87.23, 87.23, 87.23,
         8.32],
       [ 1.37,  1.44,  4.3 ,  0.84,  0.96,  2.34, 91.41, 91.41, 91.41,
        12.5 ],
       [ 1.86,  1.92,  3.4 ,  2.49,  2.61,  4.16, 88.66, 88.66, 88.66,
         9.75],
       [ 7.02,  7.06,  8.  ,  7.46,  7.56,  9.09, 90.72, 90.72, 90.72,
        11.82]]), array([[6.000e-02, 2.000e-02, 8.000e-01, 3.900e-01, 3.100e-01, 2.400e-01,
        8.863e+01, 8.863e+01, 8.863e+01, 9.890e+00],
       [8.600e-01, 8.200e-01, 7.300e-01, 1.510e+00, 1.440e+00, 7.000e-01,
        8.708e+01, 8.708e+01, 8.708e+01, 8.340e+00],
       [1.580e+00, 1.640e+00, 4.350e+00, 1.170e+00, 1.290e+00, 2.460e+00,
        9.116e+01, 9.116e+01, 9.116e+01, 1.243e+01],
       [2.160e+00, 2.230e+00, 3.630e+00, 2.900e+00, 2.990e+00, 4.410e+00,
        8.856e+01, 8.856e+01, 8.856e+01, 9.830e+00],
       [7.260e+00, 7.310e+00, 8.080e+00, 7.690e+00, 7.750e+00, 9.180e+00,
        9.054e+01, 9.054e+01, 9.054e+01, 1.180e+01]]), array([[ 2.06,  2.02,  0.3 ,  2.35,  2.26,  1.59, 78.38, 78.38,
```

```
78.38,
         6.83],
       [ 0.44,  0.37,  2.48,  0.36,  0.25,  0.78, 79.72, 79.72, 79.72,
         8.17],
       [ 3.73,  3.81,  5.99,  4.12,  4.24,  5.8 , 84.43, 84.43, 84.43,
        12.88],
       [ 6.21,  6.28,  7.97,  7.13,  7.29,  9.18, 88.02, 88.02, 88.02,
        16.47],
       [ 8.22,  8.26,  8.52,  9.21,  9.27, 10.97, 83.81, 83.81, 83.81,
        12.27]])]
time spent until now: 30.0 mins


--------------------------------------
[pattern_seed, sd_OD] = [1, 5]

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 = 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 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

--------------------------------------
Value of Behaviour policy:71.33
O_threshold = 80
MC for this TARGET:[78.33, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[2.06, 2.0, 0.37]][[2.22, 2.1, 1.56]][[-78.33, -78.33, -78.33]][-7.0]
std:[[0.35, 0.35, 0.37]][[0.12, 0.12, 0.15]][[0.0, 0.0, 0.0]][0.17]
MSE:[[2.09, 2.03, 0.52]][[2.22, 2.1, 1.57]][[78.33, 78.33, 78.33]][7.0]
MSE(-DR):[[0.0, -0.06, -1.57]][[0.13, 0.01, -0.52]][[76.24, 76.24, 76.24]][4.91]
**
==============


O_threshold = 90
MC for this TARGET:[79.701, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.15, 0.09, -2.44]][[0.26, 0.13, -0.8]][[-79.7, -79.7, -79.7]][-8.37]
std:[[0.31, 0.32, 0.26]][[0.12, 0.12, 0.24]][[0.0, 0.0, 0.0]][0.17]
MSE:[[0.34, 0.33, 2.45]][[0.29, 0.18, 0.84]][[79.7, 79.7, 79.7]][8.37]
MSE(-DR):[[0.0, -0.01, 2.11]][[-0.05, -0.16, 0.5]][[79.36, 79.36, 79.36]][8.03]
MC-based ATE = 1.37
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.91, -1.91, -2.81]][[-1.96, -1.97, -2.37]][[-1.37, -1.37, -1.37]][-1.37]
std:[[0.04, 0.03, 0.11]][[0.0, 0.01, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.91, 1.91, 2.81]][[1.96, 1.97, 2.37]][[1.37, 1.37, 1.37]][1.37]
MSE(-DR):[[0.0, 0.0, 0.9]][[0.05, 0.06, 0.46]][[-0.54, -0.54, -0.54]][-0.54]
*
==============


O_threshold = 100
MC for this TARGET:[84.329, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.78, -3.87, -6.0]][[-4.17, -4.29, -5.7]][[-84.33, -84.33, -84.33]][-13.0]
std:[[0.19, 0.18, 0.16]][[0.17, 0.18, 0.29]][[0.0, 0.0, 0.0]][0.17]
MSE:[[3.78, 3.87, 6.0]][[4.17, 4.29, 5.71]][[84.33, 84.33, 84.33]][13.0]
MSE(-DR):[[0.0, 0.09, 2.22]][[0.39, 0.51, 1.93]][[80.55, 80.55, 80.55]][9.22]
***
MC-based ATE = 6.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.83, -5.87, -6.37]][[-6.39, -6.39, -7.26]][[-6.0, -6.0, -6.0]][-6.0]
std:[[0.16, 0.17, 0.21]][[0.05, 0.05, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[5.83, 5.87, 6.37]][[6.39, 6.39, 7.26]][[6.0, 6.0, 6.0]][6.0]
MSE(-DR):[[0.0, 0.04, 0.54]][[0.56, 0.56, 1.43]][[0.17, 0.17, 0.17]][0.17]
*
==============


O_threshold = 110
MC for this TARGET:[87.923, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.38, -6.47, -7.9]][[-7.26, -7.44, -9.18]][[-87.92, -87.92, -87.92]][-16.59]
std:[[0.15, 0.15, 0.12]][[0.13, 0.12, 0.29]][[0.0, 0.0, 0.0]][0.17]
MSE:[[6.38, 6.47, 7.9]][[7.26, 7.44, 9.18]][[87.92, 87.92, 87.92]][16.59]
```

```
MSE(-DR):[[0.0, 0.09, 1.52]][[0.88, 1.06, 2.8]][[81.54, 81.54, 81.54]][10.21]
***
MC-based ATE = 9.59
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.44, -8.46, -8.27]][[-9.48, -9.55, -10.74]][[-9.59, -9.59, -9.59]][-9.59]
std:[[0.5, 0.5, 0.49]][[0.01, 0.0, 0.14]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.45, 8.47, 8.28]][[9.48, 9.55, 10.74]][[9.59, 9.59, 9.59]][9.59]
MSE(-DR):[[0.0, 0.02, -0.17]][[1.03, 1.1, 2.29]][[1.14, 1.14, 1.14]][1.14]
**
==============


O_threshold = 120
MC for this TARGET:[83.789, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.05, -8.1, -8.52]][[-9.29, -9.35, -10.88]][[-83.79, -83.79, -83.79]][-12.46]
std:[[0.54, 0.55, 0.17]][[0.27, 0.25, 0.36]][[0.0, 0.0, 0.0]][0.17]
MSE:[[8.07, 8.12, 8.52]][[9.29, 9.35, 10.89]][[83.79, 83.79, 83.79]][12.46]
MSE(-DR):[[0.0, 0.05, 0.45]][[1.22, 1.28, 2.82]][[75.72, 75.72, 75.72]][4.39]
***
MC-based ATE = 5.46
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.11, -10.09, -8.89]][[-11.52, -11.45, -12.44]][[-5.46, -5.46, -5.46]][-5.46]
std:[[0.89, 0.9, 0.55]][[0.15, 0.13, 0.2]][[0.0, 0.0, 0.0]][0.0]
MSE:[[10.15, 10.13, 8.91]][[11.52, 11.45, 12.44]][[5.46, 5.46, 5.46]][5.46]
MSE(-DR):[[0.0, -0.02, -1.24]][[1.37, 1.3, 2.29]][[-4.69, -4.69, -4.69]][-4.69]
**
==============


[array([[ 0.37,  0.36,  1.06,  0.44,  0.37,  0.25, 88.84, 88.84, 88.84,
         9.76],
       [ 0.89,  0.85,  1.23,  1.59,  1.51,  0.62, 87.43, 87.43, 87.43,
         8.36],
       [ 1.62,  1.68,  4.66,  1.16,  1.29,  2.67, 91.77, 91.77, 91.77,
        12.7 ],
       [ 1.76,  1.8 ,  3.48,  2.52,  2.63,  4.22, 88.75, 88.75, 88.75,
         9.67],
       [ 7.01,  7.04,  7.93,  7.43,  7.52,  9.18, 90.87, 90.87, 90.87,
        11.79]]), array([[ 0.34,  0.28,  0.64,  0.43,  0.34,  0.21, 88.8 , 88.8 , 88.8 ,
         9.67],
       [ 1.17,  1.13,  0.73,  1.59,  1.5 ,  0.7 , 87.32, 87.32, 87.32,
         8.2 ],
       [ 1.4 ,  1.47,  4.5 ,  1.13,  1.26,  2.49, 91.57, 91.57, 91.57,
        12.44],
       [ 1.72,  1.8 ,  3.37,  2.52,  2.63,  4.15, 88.7 , 88.7 , 88.7 ,
         9.58],
       [ 7.04,  7.08,  7.84,  7.43,  7.51,  9.01, 90.81, 90.81, 90.81,
        11.69]]), array([[ 0.27,  0.24,  0.98,  0.55,  0.48,  0.21, 88.74, 88.74, 88.74,
         9.83],
       [ 1.05,  0.99,  0.84,  1.77,  1.69,  0.8 , 87.23, 87.23, 87.23,
         8.32],
       [ 1.37,  1.44,  4.3 ,  0.84,  0.96,  2.34, 91.41, 91.41, 91.41,
        12.5 ],
       [ 1.86,  1.92,  3.4 ,  2.49,  2.61,  4.16, 88.66, 88.66, 88.66,
         9.75],
       [ 7.02,  7.06,  8.  ,  7.46,  7.56,  9.09, 90.72, 90.72, 90.72,
        11.82]]), array([[6.000e-02, 2.000e-02, 8.000e-01, 3.900e-01, 3.100e-01, 2.400e-01,
        8.863e+01, 8.863e+01, 8.863e+01, 9.890e+00],
       [8.600e-01, 8.200e-01, 7.300e-01, 1.510e+00, 1.440e+00, 7.000e-01,
        8.708e+01, 8.708e+01, 8.708e+01, 8.340e+00],
       [1.580e+00, 1.640e+00, 4.350e+00, 1.170e+00, 1.290e+00, 2.460e+00,
        9.116e+01, 9.116e+01, 9.116e+01, 1.243e+01],
       [2.160e+00, 2.230e+00, 3.630e+00, 2.900e+00, 2.990e+00, 4.410e+00,
        8.856e+01, 8.856e+01, 8.856e+01, 9.830e+00],
       [7.260e+00, 7.310e+00, 8.080e+00, 7.690e+00, 7.750e+00, 9.180e+00,
        9.054e+01, 9.054e+01, 9.054e+01, 1.180e+01]]), array([[ 2.06,  2.02,  0.3 ,  2.35,  2.26,  1.59, 78.38, 78.38,
78.38,
         6.83],
       [ 0.44,  0.37,  2.48,  0.36,  0.25,  0.78, 79.72, 79.72, 79.72,
         8.17],
       [ 3.73,  3.81,  5.99,  4.12,  4.24,  5.8 , 84.43, 84.43, 84.43,
        12.88],
       [ 6.21,  6.28,  7.97,  7.13,  7.29,  9.18, 88.02, 88.02, 88.02,
        16.47],
       [ 8.22,  8.26,  8.52,  9.21,  9.27, 10.97, 83.81, 83.81, 83.81,
        12.27]]), array([[ 2.09,  2.03,  0.52,  2.22,  2.1 ,  1.57, 78.33, 78.33, 78.33,
         7.  ],
```

```
       [ 0.34,  0.33,  2.45,  0.29,  0.18,  0.84, 79.7 , 79.7 , 79.7 ,
         8.37],
       [ 3.78,  3.87,  6.  ,  4.17,  4.29,  5.71, 84.33, 84.33, 84.33,
        13.  ],
       [ 6.38,  6.47,  7.9 ,  7.26,  7.44,  9.18, 87.92, 87.92, 87.92,
        16.59],
       [ 8.07,  8.12,  8.52,  9.29,  9.35, 10.89, 83.79, 83.79, 83.79,
        12.46]])]
time spent until now: 36.1 mins


---------------------------------------
[pattern_seed, sd_OD] = [1, 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 = 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 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

--------------------------------------
Value of Behaviour policy:71.313
O_threshold = 80
MC for this TARGET:[78.265, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.84, 1.78, 0.26]][[2.23, 2.14, 1.56]][[-78.26, -78.26, -78.26]][-6.95]
std:[[0.14, 0.14, 0.3]][[0.06, 0.07, 0.16]][[0.0, 0.0, 0.0]][0.13]
MSE:[[1.85, 1.79, 0.4]][[2.23, 2.14, 1.57]][[78.26, 78.26, 78.26]][6.95]
MSE(-DR):[[0.0, -0.06, -1.45]][[0.38, 0.29, -0.28]][[76.41, 76.41, 76.41]][5.1]
**
==============


O_threshold = 90
MC for this TARGET:[79.632, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.16, -0.22, -2.43]][[0.28, 0.16, -0.76]][[-79.63, -79.63, -79.63]][-8.32]
std:[[0.04, 0.05, 0.2]][[0.06, 0.08, 0.15]][[0.0, 0.0, 0.0]][0.13]
MSE:[[0.16, 0.23, 2.44]][[0.29, 0.18, 0.77]][[79.63, 79.63, 79.63]][8.32]
MSE(-DR):[[0.0, 0.07, 2.28]][[0.13, 0.02, 0.61]][[79.47, 79.47, 79.47]][8.16]
***
MC-based ATE = 1.37
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.0, -2.0, -2.69]][[-1.94, -1.98, -2.31]][[-1.37, -1.37, -1.37]][-1.37]
std:[[0.1, 0.09, 0.1]][[0.01, 0.01, 0.0]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.0, 2.0, 2.69]][[1.94, 1.98, 2.31]][[1.37, 1.37, 1.37]][1.37]
MSE(-DR):[[0.0, 0.0, 0.69]][[-0.06, -0.02, 0.31]][[-0.63, -0.63, -0.63]][-0.63]
==============


O_threshold = 100
MC for this TARGET:[84.212, 0.039]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.42, -4.52, -6.16]][[-4.21, -4.35, -5.75]][[-84.21, -84.21, -84.21]][-12.9]
std:[[0.09, 0.09, 0.24]][[0.04, 0.06, 0.18]][[0.0, 0.0, 0.0]][0.13]
MSE:[[4.42, 4.52, 6.16]][[4.21, 4.35, 5.75]][[84.21, 84.21, 84.21]][12.9]
MSE(-DR):[[0.0, 0.1, 1.74]][[-0.21, -0.07, 1.33]][[79.79, 79.79, 79.79]][8.48]
MC-based ATE = 5.95
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.26, -6.3, -6.42]][[-6.43, -6.49, -7.31]][[-5.95, -5.95, -5.95]][-5.95]
std:[[0.04, 0.05, 0.05]][[0.02, 0.01, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.26, 6.3, 6.42]][[6.43, 6.49, 7.31]][[5.95, 5.95, 5.95]][5.95]
MSE(-DR):[[0.0, 0.04, 0.16]][[0.17, 0.23, 1.05]][[-0.31, -0.31, -0.31]][-0.31]
*
==============


O_threshold = 110
MC for this TARGET:[87.804, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.9, -7.01, -7.72]][[-7.25, -7.41, -9.14]][[-87.8, -87.8, -87.8]][-16.49]
std:[[0.01, 0.01, 0.04]][[0.11, 0.14, 0.16]][[0.0, 0.0, 0.0]][0.13]
MSE:[[6.9, 7.01, 7.72]][[7.25, 7.41, 9.14]][[87.8, 87.8, 87.8]][16.49]
MSE(-DR):[[0.0, 0.11, 0.82]][[0.35, 0.51, 2.24]][[80.9, 80.9, 80.9]][9.59]
***
MC-based ATE = 9.54
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.74, -8.79, -7.98]][[-9.47, -9.55, -10.69]][[-9.54, -9.54, -9.54]][-9.54]
std:[[0.13, 0.13, 0.26]][[0.06, 0.07, 0.0]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.74, 8.79, 7.98]][[9.47, 9.55, 10.69]][[9.54, 9.54, 9.54]][9.54]
MSE(-DR):[[0.0, 0.05, -0.76]][[0.73, 0.81, 1.95]][[0.8, 0.8, 0.8]][0.8]
**
==============


O_threshold = 120
MC for this TARGET:[83.728, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.73, -8.79, -8.46]][[-9.39, -9.45, -10.99]][[-83.73, -83.73, -83.73]][-12.41]
std:[[0.03, 0.03, 0.11]][[0.05, 0.05, 0.06]][[0.0, 0.0, 0.0]][0.13]
MSE:[[8.73, 8.79, 8.46]][[9.39, 9.45, 10.99]][[83.73, 83.73, 83.73]][12.41]
MSE(-DR):[[0.0, 0.06, -0.27]][[0.66, 0.72, 2.26]][[75.0, 75.0, 75.0]][3.68]
**
MC-based ATE = 5.46
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.57, -10.57, -8.72]][[-11.61, -11.6, -12.54]][[-5.46, -5.46, -5.46]][-5.46]
std:[[0.11, 0.11, 0.41]][[0.01, 0.02, 0.09]][[0.0, 0.0, 0.0]][0.0]
MSE:[[10.57, 10.57, 8.73]][[11.61, 11.6, 12.54]][[5.46, 5.46, 5.46]][5.46]
MSE(-DR):[[0.0, 0.0, -1.84]][[1.04, 1.03, 1.97]][[-5.11, -5.11, -5.11]][-5.11]
**
==============


[array([[ 0.37,  0.36,  1.06,  0.44,  0.37,  0.25, 88.84, 88.84, 88.84,
         9.76],
       [ 0.89,  0.85,  1.23,  1.59,  1.51,  0.62, 87.43, 87.43, 87.43,
         8.36],
       [ 1.62,  1.68,  4.66,  1.16,  1.29,  2.67, 91.77, 91.77, 91.77,
        12.7 ],
       [ 1.76,  1.8 ,  3.48,  2.52,  2.63,  4.22, 88.75, 88.75, 88.75,
         9.67],
       [ 7.01,  7.04,  7.93,  7.43,  7.52,  9.18, 90.87, 90.87, 90.87,
        11.79]]), array([[ 0.34,  0.28,  0.64,  0.43,  0.34,  0.21, 88.8 , 88.8 , 88.8 ,
         9.67],
       [ 1.17,  1.13,  0.73,  1.59,  1.5 ,  0.7 , 87.32, 87.32, 87.32,
         8.2 ],
       [ 1.4 ,  1.47,  4.5 ,  1.13,  1.26,  2.49, 91.57, 91.57, 91.57,
        12.44],
       [ 1.72,  1.8 ,  3.37,  2.52,  2.63,  4.15, 88.7 , 88.7 , 88.7 ,
         9.58],
       [ 7.04,  7.08,  7.84,  7.43,  7.51,  9.01, 90.81, 90.81, 90.81,
        11.69]]), array([[ 0.27,  0.24,  0.98,  0.55,  0.48,  0.21, 88.74, 88.74, 88.74,
         9.83],
       [ 1.05,  0.99,  0.84,  1.77,  1.69,  0.8 , 87.23, 87.23, 87.23,
         8.32],
       [ 1.37,  1.44,  4.3 ,  0.84,  0.96,  2.34, 91.41, 91.41, 91.41,
        12.5 ],
       [ 1.86,  1.92,  3.4 ,  2.49,  2.61,  4.16, 88.66, 88.66, 88.66,
         9.75],
       [ 7.02,  7.06,  8.  ,  7.46,  7.56,  9.09, 90.72, 90.72, 90.72,
        11.82]]), array([[6.000e-02, 2.000e-02, 8.000e-01, 3.900e-01, 3.100e-01, 2.400e-01,
        8.863e+01, 8.863e+01, 8.863e+01, 9.890e+00],
       [8.600e-01, 8.200e-01, 7.300e-01, 1.510e+00, 1.440e+00, 7.000e-01,
        8.708e+01, 8.708e+01, 8.708e+01, 8.340e+00],
       [1.580e+00, 1.640e+00, 4.350e+00, 1.170e+00, 1.290e+00, 2.460e+00,
        9.116e+01, 9.116e+01, 9.116e+01, 1.243e+01],
       [2.160e+00, 2.230e+00, 3.630e+00, 2.900e+00, 2.990e+00, 4.410e+00,
        8.856e+01, 8.856e+01, 8.856e+01, 9.830e+00],
       [7.260e+00, 7.310e+00, 8.080e+00, 7.690e+00, 7.750e+00, 9.180e+00,
        9.054e+01, 9.054e+01, 9.054e+01, 1.180e+01]]), array([[ 2.06,  2.02,  0.3 ,  2.35,  2.26,  1.59, 78.38, 78.38,
78.38,
         6.83],
       [ 0.44,  0.37,  2.48,  0.36,  0.25,  0.78, 79.72, 79.72, 79.72,
         8.17],
       [ 3.73,  3.81,  5.99,  4.12,  4.24,  5.8 , 84.43, 84.43, 84.43,
        12.88],
       [ 6.21,  6.28,  7.97,  7.13,  7.29,  9.18, 88.02, 88.02, 88.02,
        16.47],
       [ 8.22,  8.26,  8.52,  9.21,  9.27, 10.97, 83.81, 83.81, 83.81,
        12.27]]), array([[ 2.09,  2.03,  0.52,  2.22,  2.1 ,  1.57, 78.33, 78.33, 78.33,
         7.  ],
       [ 0.34,  0.33,  2.45,  0.29,  0.18,  0.84, 79.7 , 79.7 , 79.7 ,
         8.37],
       [ 3.78,  3.87,  6.  ,  4.17,  4.29,  5.71, 84.33, 84.33, 84.33,
```

```
        13.  ],
       [ 6.38,  6.47,  7.9 ,  7.26,  7.44,  9.18, 87.92, 87.92, 87.92,
        16.59],
       [ 8.07,  8.12,  8.52,  9.29,  9.35, 10.89, 83.79, 83.79, 83.79,
        12.46]]), array([[ 1.85,  1.79,  0.4 ,  2.23,  2.14,  1.57, 78.26, 78.26, 78.26,
         6.95],
       [ 0.16,  0.23,  2.44,  0.29,  0.18,  0.77, 79.63, 79.63, 79.63,
         8.32],
       [ 4.42,  4.52,  6.16,  4.21,  4.35,  5.75, 84.21, 84.21, 84.21,
        12.9 ],
       [ 6.9 ,  7.01,  7.72,  7.25,  7.41,  9.14, 87.8 , 87.8 , 87.8 ,
        16.49],
       [ 8.73,  8.79,  8.46,  9.39,  9.45, 10.99, 83.73, 83.73, 83.73,
        12.41]])]
time spent until now: 42.1 mins


--------------------------------------
[pattern_seed, sd_OD] = [1, 20]

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 = 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 -th target;  2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:70.986
O_threshold = 80
MC for this TARGET:[78.134, 0.045]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.5, 1.45, 0.38]][[2.1, 2.0, 1.47]][[-78.13, -78.13, -78.13]][-7.15]
std:[[0.11, 0.11, 0.02]][[0.05, 0.03, 0.03]][[0.0, 0.0, 0.0]][0.01]
MSE:[[1.5, 1.45, 0.38]][[2.1, 2.0, 1.47]][[78.13, 78.13, 78.13]][7.15]
MSE(-DR):[[0.0, -0.05, -1.12]][[0.6, 0.5, -0.03]][[76.63, 76.63, 76.63]][5.65]
**
==============


O_threshold = 90
MC for this TARGET:[79.466, 0.04]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.5, -0.56, -2.4]][[0.08, -0.04, -0.9]][[-79.47, -79.47, -79.47]][-8.48]
std:[[0.05, 0.05, 0.09]][[0.19, 0.17, 0.1]][[0.0, 0.0, 0.0]][0.01]
MSE:[[0.5, 0.56, 2.4]][[0.21, 0.17, 0.91]][[79.47, 79.47, 79.47]][8.48]
MSE(-DR):[[0.0, 0.06, 1.9]][[-0.29, -0.33, 0.41]][[78.97, 78.97, 78.97]][7.98]
MC-based ATE = 1.33
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.0, -2.01, -2.78]][[-2.03, -2.05, -2.37]][[-1.33, -1.33, -1.33]][-1.33]
std:[[0.06, 0.06, 0.07]][[0.14, 0.13, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.0, 2.01, 2.78]][[2.03, 2.05, 2.37]][[1.33, 1.33, 1.33]][1.33]
MSE(-DR):[[0.0, 0.01, 0.78]][[0.03, 0.05, 0.37]][[-0.67, -0.67, -0.67]][-0.67]
*
==============


O_threshold = 100
MC for this TARGET:[83.992, 0.041]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.19, -4.29, -6.01]][[-4.27, -4.4, -5.72]][[-83.99, -83.99, -83.99]][-13.01]
std:[[0.24, 0.22, 0.2]][[0.16, 0.16, 0.04]][[0.0, 0.0, 0.0]][0.01]
MSE:[[4.2, 4.3, 6.01]][[4.27, 4.4, 5.72]][[83.99, 83.99, 83.99]][13.01]
MSE(-DR):[[0.0, 0.1, 1.81]][[0.07, 0.2, 1.52]][[79.79, 79.79, 79.79]][8.81]
***
MC-based ATE = 5.86
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.69, -5.74, -6.39]][[-6.38, -6.4, -7.19]][[-5.86, -5.86, -5.86]][-5.86]
std:[[0.13, 0.11, 0.17]][[0.11, 0.12, 0.01]][[0.0, 0.0, 0.0]][0.0]
MSE:[[5.69, 5.74, 6.39]][[6.38, 6.4, 7.19]][[5.86, 5.86, 5.86]][5.86]
MSE(-DR):[[0.0, 0.05, 0.7]][[0.69, 0.71, 1.5]][[0.17, 0.17, 0.17]][0.17]
*
==============


O_threshold = 110
MC for this TARGET:[87.563, 0.04]
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.07, -7.16, -8.02]][[-7.48, -7.63, -9.23]][[-87.56, -87.56, -87.56]][-16.58]
std:[[0.14, 0.12, 0.18]][[0.13, 0.12, 0.01]][[0.0, 0.0, 0.0]][0.01]
MSE:[[7.07, 7.16, 8.02]][[7.48, 7.63, 9.23]][[87.56, 87.56, 87.56]][16.58]
MSE(-DR):[[0.0, 0.09, 0.95]][[0.41, 0.56, 2.16]][[80.49, 80.49, 80.49]][9.51]
***
MC-based ATE = 9.43
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.57, -8.61, -8.4]][[-9.58, -9.63, -10.71]][[-9.43, -9.43, -9.43]][-9.43]
std:[[0.03, 0.01, 0.16]][[0.08, 0.09, 0.03]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.57, 8.61, 8.4]][[9.58, 9.63, 10.71]][[9.43, 9.43, 9.43]][9.43]
MSE(-DR):[[0.0, 0.04, -0.17]][[1.01, 1.06, 2.14]][[0.86, 0.86, 0.86]][0.86]
**
==============


O_threshold = 120
MC for this TARGET:[83.596, 0.043]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.79, -8.84, -8.57]][[-9.77, -9.83, -11.23]][[-83.6, -83.6, -83.6]][-12.61]
std:[[0.16, 0.16, 0.02]][[0.02, 0.01, 0.09]][[0.0, 0.0, 0.0]][0.01]
MSE:[[8.79, 8.84, 8.57]][[9.77, 9.83, 11.23]][[83.6, 83.6, 83.6]][12.61]
MSE(-DR):[[0.0, 0.05, -0.22]][[0.98, 1.04, 2.44]][[74.81, 74.81, 74.81]][3.82]
**
MC-based ATE = 5.46
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.29, -10.29, -8.95]][[-11.88, -11.83, -12.7]][[-5.46, -5.46, -5.46]][-5.46]
std:[[0.06, 0.05, 0.04]][[0.06, 0.05, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[10.29, 10.29, 8.95]][[11.88, 11.83, 12.7]][[5.46, 5.46, 5.46]][5.46]
MSE(-DR):[[0.0, 0.0, -1.34]][[1.59, 1.54, 2.41]][[-4.83, -4.83, -4.83]][-4.83]
**
==============


[array([[ 0.37,  0.36,  1.06,  0.44,  0.37,  0.25, 88.84, 88.84, 88.84,
         9.76],
       [ 0.89,  0.85,  1.23,  1.59,  1.51,  0.62, 87.43, 87.43, 87.43,
         8.36],
       [ 1.62,  1.68,  4.66,  1.16,  1.29,  2.67, 91.77, 91.77, 91.77,
        12.7 ],
       [ 1.76,  1.8 ,  3.48,  2.52,  2.63,  4.22, 88.75, 88.75, 88.75,
         9.67],
       [ 7.01,  7.04,  7.93,  7.43,  7.52,  9.18, 90.87, 90.87, 90.87,
        11.79]]), array([[ 0.34,  0.28,  0.64,  0.43,  0.34,  0.21, 88.8 , 88.8 , 88.8 ,
         9.67],
       [ 1.17,  1.13,  0.73,  1.59,  1.5 ,  0.7 , 87.32, 87.32, 87.32,
         8.2 ],
       [ 1.4 ,  1.47,  4.5 ,  1.13,  1.26,  2.49, 91.57, 91.57, 91.57,
        12.44],
       [ 1.72,  1.8 ,  3.37,  2.52,  2.63,  4.15, 88.7 , 88.7 , 88.7 ,
         9.58],
       [ 7.04,  7.08,  7.84,  7.43,  7.51,  9.01, 90.81, 90.81, 90.81,
        11.69]]), array([[ 0.27,  0.24,  0.98,  0.55,  0.48,  0.21, 88.74, 88.74, 88.74,
         9.83],
       [ 1.05,  0.99,  0.84,  1.77,  1.69,  0.8 , 87.23, 87.23, 87.23,
         8.32],
       [ 1.37,  1.44,  4.3 ,  0.84,  0.96,  2.34, 91.41, 91.41, 91.41,
        12.5 ],
       [ 1.86,  1.92,  3.4 ,  2.49,  2.61,  4.16, 88.66, 88.66, 88.66,
         9.75],
       [ 7.02,  7.06,  8.  ,  7.46,  7.56,  9.09, 90.72, 90.72, 90.72,
        11.82]]), array([[6.000e-02, 2.000e-02, 8.000e-01, 3.900e-01, 3.100e-01, 2.400e-01,
        8.863e+01, 8.863e+01, 8.863e+01, 9.890e+00],
       [8.600e-01, 8.200e-01, 7.300e-01, 1.510e+00, 1.440e+00, 7.000e-01,
        8.708e+01, 8.708e+01, 8.708e+01, 8.340e+00],
       [1.580e+00, 1.640e+00, 4.350e+00, 1.170e+00, 1.290e+00, 2.460e+00,
        9.116e+01, 9.116e+01, 9.116e+01, 1.243e+01],
       [2.160e+00, 2.230e+00, 3.630e+00, 2.900e+00, 2.990e+00, 4.410e+00,
        8.856e+01, 8.856e+01, 8.856e+01, 9.830e+00],
       [7.260e+00, 7.310e+00, 8.080e+00, 7.690e+00, 7.750e+00, 9.180e+00,
        9.054e+01, 9.054e+01, 9.054e+01, 1.180e+01]]), array([[ 2.06,  2.02,  0.3 ,  2.35,  2.26,  1.59, 78.38, 78.38,
78.38,
         6.83],
       [ 0.44,  0.37,  2.48,  0.36,  0.25,  0.78, 79.72, 79.72, 79.72,
         8.17],
       [ 3.73,  3.81,  5.99,  4.12,  4.24,  5.8 , 84.43, 84.43, 84.43,
        12.88],
       [ 6.21,  6.28,  7.97,  7.13,  7.29,  9.18, 88.02, 88.02, 88.02,
```

```
        16.47],
       [ 8.22,  8.26,  8.52,  9.21,  9.27, 10.97, 83.81, 83.81, 83.81,
        12.27]]), array([[ 2.09,  2.03,  0.52,  2.22,  2.1 ,  1.57, 78.33, 78.33, 78.33,
         7.  ],
       [ 0.34,  0.33,  2.45,  0.29,  0.18,  0.84, 79.7 , 79.7 , 79.7 ,
         8.37],
       [ 3.78,  3.87,  6.  ,  4.17,  4.29,  5.71, 84.33, 84.33, 84.33,
        13.  ],
       [ 6.38,  6.47,  7.9 ,  7.26,  7.44,  9.18, 87.92, 87.92, 87.92,
        16.59],
       [ 8.07,  8.12,  8.52,  9.29,  9.35, 10.89, 83.79, 83.79, 83.79,
        12.46]]), array([[ 1.85,  1.79,  0.4 ,  2.23,  2.14,  1.57, 78.26, 78.26, 78.26,
         6.95],
       [ 0.16,  0.23,  2.44,  0.29,  0.18,  0.77, 79.63, 79.63, 79.63,
         8.32],
       [ 4.42,  4.52,  6.16,  4.21,  4.35,  5.75, 84.21, 84.21, 84.21,
        12.9 ],
       [ 6.9 ,  7.01,  7.72,  7.25,  7.41,  9.14, 87.8 , 87.8 , 87.8 ,
        16.49],
       [ 8.73,  8.79,  8.46,  9.39,  9.45, 10.99, 83.73, 83.73, 83.73,
        12.41]]), array([[ 1.5 ,  1.45,  0.38,  2.1 ,  2.  ,  1.47, 78.13, 78.13, 78.13,
         7.15],
       [ 0.5 ,  0.56,  2.4 ,  0.21,  0.17,  0.91, 79.47, 79.47, 79.47,
         8.48],
       [ 4.2 ,  4.3 ,  6.01,  4.27,  4.4 ,  5.72, 83.99, 83.99, 83.99,
        13.01],
       [ 7.07,  7.16,  8.02,  7.48,  7.63,  9.23, 87.56, 87.56, 87.56,
        16.58],
       [ 8.79,  8.84,  8.57,  9.77,  9.83, 11.23, 83.6 , 83.6 , 83.6 ,
        12.61]])]
time spent until now: 48.1 mins


--------------------------------------
[pattern_seed, sd_OD] = [2, 0.5]

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 = 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 -th target; 2 -th target; 3 -th target;
```