```
Last login: Mon Mar 30 22:09:33 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-219-215-112.compute-1.amazonaws.com
The authenticity of host 'ec2-3-219-215-112.compute-1.amazonaws.com (3.219.215.112)' can't be established.
ECDSA key fingerprint is SHA256:zwNf6JCVuX3r8uy1flyIAhbA8xtXEVlntelIPHrPvjg.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-219-215-112.compute-1.amazonaws.com,3.219.215.112' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1060-aws x86_64)

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

  System information as of Tue Mar 31 03:24:36 UTC 2020

  System load:  0.57               Processes:           227
  Usage of /:   55.4% of 15.45GB   Users logged in:     0
  Memory usage: 1%                 IP address for ens5: 172.31.9.154
  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
export openblas_num_threads=1; export OMP_NUM_THREADS=1ubuntu@ip-172-31-9-154:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1
ubuntu@ip-172-31-9-154:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1
ubuntu@ip-172-31-9-154:~$ python EC2-l.py
23:26, 03/30; num of cores:16

Traceback (most recent call last):
  File "EC2-l.py", line 42, in <module>
    shared_setting = "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_range] = " + str([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_range]) +
"\n"
NameError: name 'day_range' is not defined
ubuntu@ip-172-31-9-154:~$ python EC2-l.py
23:26, 03/30; 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_range] = [None, 10,
10, 5, 0.2, 0.5, 1, 0.0001, False, True, 5, False, [3, 7, 14], [80, 90, 100, 110, 120, 130]]


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

max(u_O) =  155.7
O_threshold = 80
means of Order:

141.6 107.8 121.0

155.7 144.5 81.8

120.3 96.5 97.5

target policy:

1 1 1

1 1 1

1 1 1

number of reward locations:  9
O_threshold = 90
target policy:

1 1 1

1 1 0

1 1 1

number of reward locations:  8
```

```
O_threshold = 100
target policy:

1 1 1

1 1 0

1 0 0

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

1 0 1

1 1 0

1 0 0

number of reward locations:  5
O_threshold = 120
target policy:

1 0 1

1 1 0

1 0 0

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

1 0 0

1 1 0

0 0 0

number of reward locations:  3
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:79.0
O_threshold = 80
MC for this TARGET:[88.205, 0.132]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[6.04, 5.91, 3.81]][[6.53, 6.18, 6.01]][[-88.2, -88.2, -88.2]][-9.2]
std:[[0.27, 0.29, 0.38]][[0.27, 0.27, 0.26]][[0.0, 0.0, 0.0]][0.15]
MSE:[[6.05, 5.92, 3.83]][[6.54, 6.19, 6.02]][[88.2, 88.2, 88.2]][9.2]
MSE(-DR):[[0.0, -0.13, -2.22]][[0.49, 0.14, -0.03]][[82.15, 82.15, 82.15]][3.15]
**
==============


O_threshold = 90
MC for this TARGET:[90.911, 0.12]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[3.58, 3.43, 0.27]][[4.57, 4.21, 3.84]][[-90.91, -90.91, -90.91]][-11.91]
std:[[0.26, 0.29, 0.34]][[0.39, 0.35, 0.39]][[0.0, 0.0, 0.0]][0.15]
MSE:[[3.59, 3.44, 0.43]][[4.59, 4.22, 3.86]][[90.91, 90.91, 90.91]][11.91]
MSE(-DR):[[0.0, -0.15, -3.16]][[1.0, 0.63, 0.27]][[87.32, 87.32, 87.32]][8.32]
**
MC-based ATE = 2.71
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.46, -2.48, -3.54]][[-1.95, -1.96, -2.17]][[-2.71, -2.71, -2.71]][-2.71]
std:[[0.09, 0.06, 0.12]][[0.14, 0.14, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.46, 2.48, 3.54]][[1.96, 1.96, 2.17]][[2.71, 2.71, 2.71]][2.71]
MSE(-DR):[[0.0, 0.02, 1.08]][[-0.5, -0.5, -0.29]][[0.25, 0.25, 0.25]][0.25]
==============


O_threshold = 100
MC for this TARGET:[88.572, 0.111]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.2, 1.06, -1.27]][[2.52, 2.24, 1.79]][[-88.57, -88.57, -88.57]][-9.57]
std:[[0.11, 0.11, 0.15]][[0.45, 0.43, 0.33]][[0.0, 0.0, 0.0]][0.15]
MSE:[[1.21, 1.07, 1.28]][[2.56, 2.28, 1.82]][[88.57, 88.57, 88.57]][9.57]
MSE(-DR):[[0.0, -0.14, 0.07]][[1.35, 1.07, 0.61]][[87.36, 87.36, 87.36]][8.36]
***
MC-based ATE = 0.37
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.83, -4.85, -5.09]][[-4.01, -3.94, -4.22]][[-0.37, -0.37, -0.37]][-0.37]
std:[[0.22, 0.24, 0.48]][[0.29, 0.32, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[4.84, 4.86, 5.11]][[4.02, 3.95, 4.22]][[0.37, 0.37, 0.37]][0.37]
MSE(-DR):[[0.0, 0.02, 0.27]][[-0.82, -0.89, -0.62]][[-4.47, -4.47, -4.47]][-4.47]
==============
```

```
O_threshold = 110
MC for this TARGET:[91.202, 0.107]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.9, -2.04, -4.81]][[-1.36, -1.69, -2.18]][[-91.2, -91.2, -91.2]][-12.2]
std:[[0.56, 0.56, 0.28]][[0.36, 0.32, 0.27]][[0.0, 0.0, 0.0]][0.15]
MSE:[[1.98, 2.12, 4.82]][[1.41, 1.72, 2.2]][[91.2, 91.2, 91.2]][12.2]
MSE(-DR):[[0.0, 0.14, 2.84]][[-0.57, -0.26, 0.22]][[89.22, 89.22, 89.22]][10.22]
MC-based ATE = 3.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.93, -7.95, -8.63]][[-7.89, -7.87, -8.18]][[-3.0, -3.0, -3.0]][-3.0]
std:[[0.82, 0.83, 0.49]][[0.22, 0.23, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.97, 7.99, 8.64]][[7.89, 7.87, 8.18]][[3.0, 3.0, 3.0]][3.0]
MSE(-DR):[[0.0, 0.02, 0.67]][[-0.08, -0.1, 0.21]][[-4.97, -4.97, -4.97]][-4.97]
==============


O_threshold = 120
MC for this TARGET:[91.202, 0.107]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.89, -2.04, -4.84]][[-1.37, -1.69, -2.18]][[-91.2, -91.2, -91.2]][-12.2]
std:[[0.54, 0.56, 0.18]][[0.4, 0.32, 0.33]][[0.0, 0.0, 0.0]][0.15]
MSE:[[1.97, 2.12, 4.84]][[1.43, 1.72, 2.2]][[91.2, 91.2, 91.2]][12.2]
MSE(-DR):[[0.0, 0.15, 2.87]][[-0.54, -0.25, 0.23]][[89.23, 89.23, 89.23]][10.23]
MC-based ATE = 3.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.93, -7.95, -8.66]][[-7.9, -7.87, -8.19]][[-3.0, -3.0, -3.0]][-3.0]
std:[[0.8, 0.83, 0.51]][[0.24, 0.23, 0.12]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.97, 7.99, 8.68]][[7.9, 7.87, 8.19]][[3.0, 3.0, 3.0]][3.0]
MSE(-DR):[[0.0, 0.02, 0.71]][[-0.07, -0.1, 0.22]][[-4.97, -4.97, -4.97]][-4.97]
==============


O_threshold = 130
MC for this TARGET:[84.301, 0.115]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.56, -5.58, -6.43]][[-4.31, -4.45, -4.73]][[-84.3, -84.3, -84.3]][-5.3]
std:[[0.3, 0.31, 0.42]][[0.29, 0.29, 0.16]][[0.0, 0.0, 0.0]][0.15]
MSE:[[5.57, 5.59, 6.44]][[4.32, 4.46, 4.73]][[84.3, 84.3, 84.3]][5.3]
MSE(-DR):[[0.0, 0.02, 0.87]][[-1.25, -1.11, -0.84]][[78.73, 78.73, 78.73]][-0.27]
MC-based ATE = -3.9
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.6, -11.49, -10.24]][[-10.84, -10.63, -10.74]][[3.9, 3.9, 3.9]][3.9]
std:[[0.37, 0.39, 0.8]][[0.13, 0.13, 0.18]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.61, 11.5, 10.27]][[10.84, 10.63, 10.74]][[3.9, 3.9, 3.9]][3.9]
MSE(-DR):[[0.0, -0.11, -1.34]][[-0.77, -0.98, -0.87]][[-7.71, -7.71, -7.71]][-7.71]
==============


time spent until now: 5.4 mins


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

max(u_O) =  155.7
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

target policy:

1 1 1 1

1 1 1 1

1 1 1 1

1 1 1 1

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

1 1 1 1

1 0 1 1

1 1 1 1

1 1 1 1
```

```
number of reward locations:  15
O_threshold = 100
target policy:

1 1 1 1

1 0 1 0

0 1 1 1

1 1 1 1

number of reward locations:  13
O_threshold = 110
target policy:

1 0 1 1

1 0 1 0

0 0 0 1

1 0 0 0

number of reward locations:  7
O_threshold = 120
target policy:

1 0 1 1

1 0 1 0

0 0 0 1

0 0 0 0

number of reward locations:  6
O_threshold = 130
target policy:

1 0 0 1

1 0 0 0

0 0 0 1

0 0 0 0

number of reward locations:  4
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:74.156
O_threshold = 80
MC for this TARGET:[84.166, 0.093]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[3.9, 3.79, 2.07]][[4.24, 4.0, 3.88]][[-84.17, -84.17, -84.17]][-10.01]
std:[[0.29, 0.27, 0.45]][[0.27, 0.31, 0.09]][[0.0, 0.0, 0.0]][0.14]
MSE:[[3.91, 3.8, 2.12]][[4.25, 4.01, 3.88]][[84.17, 84.17, 84.17]][10.01]
MSE(-DR):[[0.0, -0.11, -1.79]][[0.34, 0.1, -0.03]][[80.26, 80.26, 80.26]][6.1]
**
==============


O_threshold = 90
MC for this TARGET:[86.963, 0.093]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.74, 1.62, -0.33]][[2.46, 2.21, 1.97]][[-86.96, -86.96, -86.96]][-12.81]
std:[[0.34, 0.34, 0.36]][[0.26, 0.29, 0.07]][[0.0, 0.0, 0.0]][0.14]
MSE:[[1.77, 1.66, 0.49]][[2.47, 2.23, 1.97]][[86.96, 86.96, 86.96]][12.81]
MSE(-DR):[[0.0, -0.11, -1.28]][[0.7, 0.46, 0.2]][[85.19, 85.19, 85.19]][11.04]
**
MC-based ATE = 2.8
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.16, -2.17, -2.41]][[-1.78, -1.79, -1.91]][[-2.8, -2.8, -2.8]][-2.8]
std:[[0.06, 0.08, 0.1]][[0.05, 0.05, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.16, 2.17, 2.41]][[1.78, 1.79, 1.91]][[2.8, 2.8, 2.8]][2.8]
MSE(-DR):[[0.0, 0.01, 0.25]][[-0.38, -0.37, -0.25]][[0.64, 0.64, 0.64]][0.64]
==============


O_threshold = 100
MC for this TARGET:[83.769, 0.093]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[2.99, 2.87, 0.11]][[3.52, 3.27, 3.06]][[-83.77, -83.77, -83.77]][-9.61]
std:[[0.41, 0.43, 0.22]][[0.11, 0.15, 0.02]][[0.0, 0.0, 0.0]][0.14]
```

```
MSE:[[3.02, 2.9, 0.25]][[3.52, 3.27, 3.06]][[83.77, 83.77, 83.77]][9.61]
MSE(-DR):[[0.0, -0.12, -2.77]][[0.5, 0.25, 0.04]][[80.75, 80.75, 80.75]][6.59]
**
MC-based ATE = -0.4
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.91, -0.92, -1.96]][[-0.72, -0.73, -0.82]][[0.4, 0.4, 0.4]][0.4]
std:[[0.17, 0.2, 0.25]][[0.16, 0.17, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.93, 0.94, 1.98]][[0.74, 0.75, 0.82]][[0.4, 0.4, 0.4]][0.4]
MSE(-DR):[[0.0, 0.01, 1.05]][[-0.19, -0.18, -0.11]][[-0.53, -0.53, -0.53]][-0.53]
==============


O_threshold = 110
MC for this TARGET:[87.942, 0.077]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.83, -5.95, -7.34]][[-6.68, -6.89, -7.29]][[-87.94, -87.94, -87.94]][-13.79]
std:[[0.27, 0.25, 0.08]][[0.12, 0.1, 0.18]][[0.0, 0.0, 0.0]][0.14]
MSE:[[5.84, 5.96, 7.34]][[6.68, 6.89, 7.29]][[87.94, 87.94, 87.94]][13.79]
MSE(-DR):[[0.0, 0.12, 1.5]][[0.84, 1.05, 1.45]][[82.1, 82.1, 82.1]][7.95]
***
MC-based ATE = 3.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.72, -9.74, -9.41]][[-10.92, -10.9, -11.17]][[-3.78, -3.78, -3.78]][-3.78]
std:[[0.06, 0.06, 0.51]][[0.39, 0.41, 0.25]][[0.0, 0.0, 0.0]][0.0]
MSE:[[9.72, 9.74, 9.42]][[10.93, 10.91, 11.17]][[3.78, 3.78, 3.78]][3.78]
MSE(-DR):[[0.0, 0.02, -0.3]][[1.21, 1.19, 1.45]][[-5.94, -5.94, -5.94]][-5.94]
**
==============


O_threshold = 120
MC for this TARGET:[85.233, 0.081]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.03, -5.14, -6.44]][[-6.44, -6.62, -7.0]][[-85.23, -85.23, -85.23]][-11.08]
std:[[0.46, 0.44, 0.28]][[0.07, 0.03, 0.15]][[0.0, 0.0, 0.0]][0.14]
MSE:[[5.05, 5.16, 6.45]][[6.44, 6.62, 7.0]][[85.23, 85.23, 85.23]][11.08]
MSE(-DR):[[0.0, 0.11, 1.4]][[1.39, 1.57, 1.95]][[80.18, 80.18, 80.18]][6.03]
***
MC-based ATE = 1.07
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-8.93, -8.93, -8.52]][[-10.68, -10.62, -10.88]][[-1.07, -1.07, -1.07]][-1.07]
std:[[0.28, 0.25, 0.54]][[0.34, 0.34, 0.18]][[0.0, 0.0, 0.0]][0.0]
MSE:[[8.93, 8.93, 8.54]][[10.69, 10.63, 10.88]][[1.07, 1.07, 1.07]][1.07]
MSE(-DR):[[0.0, 0.0, -0.39]][[1.76, 1.7, 1.95]][[-7.86, -7.86, -7.86]][-7.86]
**
==============


O_threshold = 130
MC for this TARGET:[90.882, 0.087]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.36, -10.47, -10.6]][[-14.53, -14.68, -15.14]][[-90.88, -90.88, -90.88]][-16.73]
std:[[0.55, 0.57, 0.41]][[0.17, 0.14, 0.13]][[0.0, 0.0, 0.0]][0.14]
MSE:[[10.37, 10.49, 10.61]][[14.53, 14.68, 15.14]][[90.88, 90.88, 90.88]][16.73]
MSE(-DR):[[0.0, 0.12, 0.24]][[4.16, 4.31, 4.77]][[80.51, 80.51, 80.51]][6.36]
***
MC-based ATE = 6.72
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-14.26, -14.26, -12.68]][[-18.77, -18.68, -19.01]][[-6.72, -6.72, -6.72]][-6.72]
std:[[0.46, 0.46, 0.56]][[0.43, 0.44, 0.21]][[0.0, 0.0, 0.0]][0.0]
MSE:[[14.27, 14.27, 12.69]][[18.77, 18.69, 19.01]][[6.72, 6.72, 6.72]][6.72]
MSE(-DR):[[0.0, 0.0, -1.58]][[4.5, 4.42, 4.74]][[-7.55, -7.55, -7.55]][-7.55]
**
==============


time spent until now: 13.7 mins


----------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 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
O_threshold = 130
target policy:

1 0 0 1 1

0 0 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  6
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

--------------------------------------
Value of Behaviour policy:72.847
O_threshold = 80
MC for this TARGET:[83.948, 0.075]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

```
bias:[[1.09, 0.99, 0.37]][[2.12, 1.92, 1.72]][[-83.95, -83.95, -83.95]][-11.1]
std:[[0.33, 0.33, 0.21]][[0.11, 0.09, 0.05]][[0.0, 0.0, 0.0]][0.05]
MSE:[[1.14, 1.04, 0.43]][[2.12, 1.92, 1.72]][[83.95, 83.95, 83.95]][11.1]
MSE(-DR):[[0.0, -0.1, -0.71]][[0.98, 0.78, 0.58]][[82.81, 82.81, 82.81]][9.96]
**
==============


O_threshold = 90
MC for this TARGET:[81.134, 0.067]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[2.17, 2.07, 0.66]][[3.6, 3.41, 3.15]][[-81.13, -81.13, -81.13]][-8.29]
std:[[0.21, 0.19, 0.22]][[0.16, 0.15, 0.16]][[0.0, 0.0, 0.0]][0.05]
MSE:[[2.18, 2.08, 0.7]][[3.6, 3.41, 3.15]][[81.13, 81.13, 81.13]][8.29]
MSE(-DR):[[0.0, -0.1, -1.48]][[1.42, 1.23, 0.97]][[78.95, 78.95, 78.95]][6.11]
**
MC-based ATE = -2.81
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.09, 1.08, 0.3]][[1.48, 1.49, 1.43]][[2.81, 2.81, 2.81]][2.81]
std:[[0.16, 0.17, 0.15]][[0.1, 0.1, 0.11]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.1, 1.09, 0.34]][[1.48, 1.49, 1.43]][[2.81, 2.81, 2.81]][2.81]
MSE(-DR):[[0.0, -0.01, -0.76]][[0.38, 0.39, 0.33]][[1.71, 1.71, 1.71]][1.71]
**
==============


O_threshold = 100
MC for this TARGET:[84.549, 0.072]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.06, -0.19, -2.48]][[1.27, 1.04, 0.53]][[-84.55, -84.55, -84.55]][-11.7]
std:[[0.18, 0.19, 0.13]][[0.15, 0.11, 0.2]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.19, 0.27, 2.48]][[1.28, 1.05, 0.57]][[84.55, 84.55, 84.55]][11.7]
MSE(-DR):[[0.0, 0.08, 2.29]][[1.09, 0.86, 0.38]][[84.36, 84.36, 84.36]][11.51]
***
MC-based ATE = 0.6
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.15, -1.18, -2.85]][[-0.84, -0.88, -1.18]][[-0.6, -0.6, -0.6]][-0.6]
std:[[0.51, 0.52, 0.18]][[0.08, 0.06, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.26, 1.29, 2.86]][[0.84, 0.88, 1.19]][[0.6, 0.6, 0.6]][0.6]
MSE(-DR):[[0.0, 0.03, 1.6]][[-0.42, -0.38, -0.07]][[-0.66, -0.66, -0.66]][-0.66]
==============


O_threshold = 110
MC for this TARGET:[80.45, 0.059]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.85, -0.97, -1.21]][[-1.57, -1.76, -2.2]][[-80.45, -80.45, -80.45]][-7.6]
std:[[0.14, 0.12, 0.09]][[0.15, 0.12, 0.21]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.86, 0.98, 1.21]][[1.58, 1.76, 2.21]][[80.45, 80.45, 80.45]][7.6]
MSE(-DR):[[0.0, 0.12, 0.35]][[0.72, 0.9, 1.35]][[79.59, 79.59, 79.59]][6.74]
***
MC-based ATE = -3.5
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.94, -1.95, -1.58]][[-3.69, -3.67, -3.92]][[3.5, 3.5, 3.5]][3.5]
std:[[0.29, 0.28, 0.13]][[0.04, 0.04, 0.16]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.96, 1.97, 1.59]][[3.69, 3.67, 3.92]][[3.5, 3.5, 3.5]][3.5]
MSE(-DR):[[0.0, 0.01, -0.37]][[1.73, 1.71, 1.96]][[1.54, 1.54, 1.54]][1.54]
**
==============


O_threshold = 120
MC for this TARGET:[82.255, 0.058]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.09, -6.13, -5.97]][[-7.26, -7.4, -7.82]][[-82.26, -82.26, -82.26]][-9.41]
std:[[0.37, 0.37, 0.14]][[0.19, 0.14, 0.23]][[0.0, 0.0, 0.0]][0.05]
MSE:[[6.1, 6.14, 5.97]][[7.26, 7.4, 7.82]][[82.26, 82.26, 82.26]][9.41]
MSE(-DR):[[0.0, 0.04, -0.13]][[1.16, 1.3, 1.72]][[76.16, 76.16, 76.16]][3.31]
**
MC-based ATE = -1.69
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.17, -7.12, -6.33]][[-9.38, -9.31, -9.54]][[1.69, 1.69, 1.69]][1.69]
std:[[0.38, 0.39, 0.08]][[0.08, 0.05, 0.19]][[0.0, 0.0, 0.0]][0.0]
MSE:[[7.18, 7.13, 6.33]][[9.38, 9.31, 9.54]][[1.69, 1.69, 1.69]][1.69]
MSE(-DR):[[0.0, -0.05, -0.85]][[2.2, 2.13, 2.36]][[-5.49, -5.49, -5.49]][-5.49]
**
==============


O_threshold = 130
MC for this TARGET:[86.646, 0.06]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.39, -11.4, -10.24]][[-13.99, -14.07, -14.51]][[-86.65, -86.65, -86.65]][-13.8]
std:[[0.18, 0.18, 0.09]][[0.13, 0.08, 0.19]][[0.0, 0.0, 0.0]][0.05]
MSE:[[11.39, 11.4, 10.24]][[13.99, 14.07, 14.51]][[86.65, 86.65, 86.65]][13.8]
MSE(-DR):[[0.0, 0.01, -1.15]][[2.6, 2.68, 3.12]][[75.26, 75.26, 75.26]][2.41]
**
MC-based ATE = 2.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

bias:[[-12.48, -12.39, -10.61]][[-16.1, -15.99, -16.22]][[-2.7, -2.7, -2.7]][-2.7]
std:[[0.38, 0.4, 0.18]][[0.02, 0.01, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[12.49, 12.4, 10.61]][[16.1, 15.99, 16.22]][[2.7, 2.7, 2.7]][2.7]
MSE(-DR):[[0.0, -0.09, -1.88]][[3.61, 3.5, 3.73]][[-9.79, -9.79, -9.79]][-9.79]
**
==========

time spent until now: 26.8 mins

[pattern_seed, T, sd_R] = [0, 672, 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 74.4 100.4 95.8 135.2 133.5

102.6 107.3 83.3 66.9 92.8 102.6

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 0 1 1 1 1

1 1 1 0 1 1

number of reward locations: 33
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 0 1 1 1 1

1 1 0 0 1 1

number of reward locations: 29
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 0 1 0 1 1

1 1 0 0 0 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 0 0 0 1 1

```
0 0 0 0 0 0

number of reward locations:  13
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 0 0 0 1 1

0 0 0 0 0 0

number of reward locations:  10
O_threshold = 130
target policy:

1 0 0 1 1 0

0 0 0 0 0 1

0 0 0 0 1 0

0 0 0 0 0 0

1 0 0 0 1 1

0 0 0 0 0 0

number of reward locations:  8
1 -th target; 2 -th target; 3 -th target; ^[[C^[[C^[[C^[[C^[[C^[[C^[[C4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:70.415
O_threshold = 80
MC for this TARGET:[82.795, 0.064]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.43, 0.33, -1.27]][[2.54, 2.23, 2.09]][[-82.8, -82.8, -82.8]][-12.38]
std:[[0.51, 0.5, 0.29]][[0.1, 0.07, 0.08]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.67, 0.6, 1.3]][[2.54, 2.23, 2.09]][[82.8, 82.8, 82.8]][12.38]
MSE(-DR):[[0.0, -0.07, 0.63]][[1.87, 1.56, 1.42]][[82.13, 82.13, 82.13]][11.71]
***
==============


O_threshold = 90
MC for this TARGET:[80.739, 0.062]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.95, 0.86, -0.88]][[2.74, 2.47, 2.26]][[-80.74, -80.74, -80.74]][-10.32]
std:[[0.47, 0.48, 0.22]][[0.06, 0.04, 0.02]][[0.0, 0.0, 0.0]][0.06]
MSE:[[1.06, 0.98, 0.91]][[2.74, 2.47, 2.26]][[80.74, 80.74, 80.74]][10.32]
MSE(-DR):[[0.0, -0.08, -0.15]][[1.68, 1.41, 1.2]][[79.68, 79.68, 79.68]][9.26]
**
MC-based ATE = -2.06
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.52, 0.53, 0.39]][[0.2, 0.25, 0.18]][[2.06, 2.06, 2.06]][2.06]
std:[[0.06, 0.04, 0.16]][[0.04, 0.04, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.52, 0.53, 0.42]][[0.2, 0.25, 0.19]][[2.06, 2.06, 2.06]][2.06]
MSE(-DR):[[0.0, 0.01, -0.1]][[-0.32, -0.27, -0.33]][[1.54, 1.54, 1.54]][1.54]
==============


O_threshold = 100
MC for this TARGET:[81.322, 0.063]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.79, -0.92, -3.04]][[0.64, 0.36, 0.04]][[-81.32, -81.32, -81.32]][-10.91]
std:[[0.28, 0.29, 0.15]][[0.08, 0.06, 0.03]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.84, 0.96, 3.04]][[0.64, 0.36, 0.05]][[81.32, 81.32, 81.32]][10.91]
MSE(-DR):[[0.0, 0.12, 2.2]][[-0.2, -0.48, -0.79]][[80.48, 80.48, 80.48]][10.07]
MC-based ATE = -1.47
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-1.22, -1.25, -1.77]][[-1.9, -1.86, -2.04]][[1.47, 1.47, 1.47]][1.47]
std:[[0.35, 0.34, 0.24]][[0.01, 0.02, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.27, 1.3, 1.79]][[1.9, 1.86, 2.04]][[1.47, 1.47, 1.47]][1.47]
MSE(-DR):[[0.0, 0.03, 0.52]][[0.63, 0.59, 0.77]][[0.2, 0.2, 0.2]][0.2]
*
==============


O_threshold = 110
MC for this TARGET:[82.04, 0.065]
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.74, -5.84, -6.62]][[-6.58, -6.77, -7.25]][[-82.04, -82.04, -82.04]][-11.62]
std:[[0.29, 0.27, 0.16]][[0.06, 0.06, 0.07]][[0.0, 0.0, 0.0]][0.06]
MSE:[[5.75, 5.85, 6.62]][[6.58, 6.77, 7.25]][[82.04, 82.04, 82.04]][11.62]
MSE(-DR):[[0.0, 0.1, 0.87]][[0.83, 1.02, 1.5]][[76.29, 76.29, 76.29]][5.87]
***
MC-based ATE = -0.75
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.17, -6.17, -5.35]][[-9.12, -9.0, -9.33]][[0.75, 0.75, 0.75]][0.75]
std:[[0.79, 0.77, 0.43]][[0.15, 0.14, 0.15]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.22, 6.22, 5.37]][[9.12, 9.0, 9.33]][[0.75, 0.75, 0.75]][0.75]
MSE(-DR):[[0.0, 0.0, -0.85]][[2.9, 2.78, 3.11]][[-5.47, -5.47, -5.47]][-5.47]
**
==============


O_threshold = 120
MC for this TARGET:[85.14, 0.064]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-10.87, -10.95, -11.12]][[-12.02, -12.19, -12.66]][[-85.14, -85.14, -85.14]][-14.72]
std:[[0.47, 0.45, 0.24]][[0.13, 0.13, 0.14]][[0.0, 0.0, 0.0]][0.06]
MSE:[[10.88, 10.96, 11.12]][[12.02, 12.19, 12.66]][[85.14, 85.14, 85.14]][14.72]
MSE(-DR):[[0.0, 0.08, 0.24]][[1.14, 1.31, 1.78]][[74.26, 74.26, 74.26]][3.84]
***
MC-based ATE = 2.34
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.3, -11.28, -9.85]][[-14.55, -14.42, -14.74]][[-2.34, -2.34, -2.34]][-2.34]
std:[[0.97, 0.95, 0.52]][[0.23, 0.2, 0.22]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.34, 11.32, 9.86]][[14.55, 14.42, 14.74]][[2.34, 2.34, 2.34]][2.34]
MSE(-DR):[[0.0, -0.02, -1.48]][[3.21, 3.08, 3.4]][[-9.0, -9.0, -9.0]][-9.0]
**
==============


O_threshold = 130
MC for this TARGET:[83.783, 0.065]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.26, -11.3, -10.87]][[-13.21, -13.33, -13.79]][[-83.78, -83.78, -83.78]][-13.37]
std:[[0.34, 0.33, 0.18]][[0.1, 0.1, 0.1]][[0.0, 0.0, 0.0]][0.06]
MSE:[[11.27, 11.3, 10.87]][[13.21, 13.33, 13.79]][[83.78, 83.78, 83.78]][13.37]
MSE(-DR):[[0.0, 0.03, -0.4]][[1.94, 2.06, 2.52]][[72.51, 72.51, 72.51]][2.1]
**
MC-based ATE = 0.99
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.69, -11.63, -9.6]][[-15.75, -15.56, -15.87]][[-0.99, -0.99, -0.99]][-0.99]
std:[[0.84, 0.83, 0.43]][[0.19, 0.17, 0.18]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.72, 11.66, 9.61]][[15.75, 15.56, 15.87]][[0.99, 0.99, 0.99]][0.99]
MSE(-DR):[[0.0, -0.06, -2.11]][[4.03, 3.84, 4.15]][[-10.73, -10.73, -10.73]][-10.73]
**
==============


time spent until now: 45.7 mins


----------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 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 74.4 100.4 95.8

135.2 133.5 102.6 107.3 83.3 66.9 92.8

102.6 127.2 126.5 92.1 93.6 80.7 74.9

70.7 147.0 89.8 91.1 77.4 116.2 72.0

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 0 1 1

1 1 1 1 1 0 1

1 1 1 1 1 1 0
```

```
0 1 1 1 0 1 0

number of reward locations:  42
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 0 1 1

1 1 1 1 0 0 1

1 1 1 1 1 0 0

0 1 0 1 0 1 0

number of reward locations:  36
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 0 1 0

1 1 1 1 0 0 0

1 1 1 0 0 0 0

0 1 0 0 0 1 0

number of reward locations:  28
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 0 0 0

1 1 0 0 0 0 0

0 1 1 0 0 0 0

0 1 0 0 0 1 0

number of reward locations:  17
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 0 0 0

1 1 0 0 0 0 0

0 1 1 0 0 0 0

0 1 0 0 0 0 0

number of reward locations:  13
O_threshold = 130
target policy:

1 0 0 1 1 0 0

0 0 0 0 1 0 0

0 0 1 0 0 0 0

0 0 0 1 0 0 0

1 1 0 0 0 0 0
```

```
0 0 0 0 0 0 0

0 1 0 0 0 0 0

number of reward locations:  9
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

---------------------------------------
Value of Behaviour policy:69.224
O_threshold = 80
MC for this TARGET:[80.099, 0.05]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.1, 1.02, -0.46]][[2.5, 2.27, 1.92]][[-80.1, -80.1, -80.1]][-10.88]
std:[[0.57, 0.58, 0.07]][[0.04, 0.05, 0.06]][[0.0, 0.0, 0.0]][0.15]
MSE:[[1.24, 1.17, 0.47]][[2.5, 2.27, 1.92]][[80.1, 80.1, 80.1]][10.88]
MSE(-DR):[[0.0, -0.07, -0.77]][[1.26, 1.03, 0.68]][[78.86, 78.86, 78.86]][9.64]
**
==============


O_threshold = 90
MC for this TARGET:[79.639, 0.049]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[1.44, 1.31, -0.7]][[2.44, 2.21, 1.73]][[-79.64, -79.64, -79.64]][-10.42]
std:[[0.29, 0.31, 0.09]][[0.08, 0.09, 0.07]][[0.0, 0.0, 0.0]][0.15]
MSE:[[1.47, 1.35, 0.71]][[2.44, 2.21, 1.73]][[79.64, 79.64, 79.64]][10.42]
MSE(-DR):[[0.0, -0.12, -0.76]][[0.97, 0.74, 0.26]][[78.17, 78.17, 78.17]][8.95]
**
MC-based ATE = -0.46
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.34, 0.3, -0.24]][[-0.06, -0.06, -0.19]][[0.46, 0.46, 0.46]][0.46]
std:[[0.33, 0.32, 0.05]][[0.05, 0.04, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.47, 0.44, 0.25]][[0.08, 0.07, 0.19]][[0.46, 0.46, 0.46]][0.46]
MSE(-DR):[[0.0, -0.03, -0.22]][[-0.39, -0.4, -0.28]][[-0.01, -0.01, -0.01]][-0.01]
==============


O_threshold = 100
MC for this TARGET:[77.488, 0.05]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.8, 0.69, -0.78]][[1.62, 1.39, 0.9]][[-77.49, -77.49, -77.49]][-8.26]
std:[[0.21, 0.21, 0.17]][[0.15, 0.14, 0.14]][[0.0, 0.0, 0.0]][0.15]
MSE:[[0.83, 0.72, 0.8]][[1.63, 1.4, 0.91]][[77.49, 77.49, 77.49]][8.26]
MSE(-DR):[[0.0, -0.11, -0.03]][[0.8, 0.57, 0.08]][[76.66, 76.66, 76.66]][7.43]
**
MC-based ATE = -2.61
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-0.3, -0.32, -0.32]][[-0.89, -0.88, -1.02]][[2.61, 2.61, 2.61]][2.61]
std:[[0.44, 0.44, 0.11]][[0.11, 0.09, 0.09]][[0.0, 0.0, 0.0]][0.0]
MSE:[[0.53, 0.54, 0.34]][[0.9, 0.88, 1.02]][[2.61, 2.61, 2.61]][2.61]
MSE(-DR):[[0.0, 0.01, -0.19]][[0.37, 0.35, 0.49]][[2.08, 2.08, 2.08]][2.08]
**
==============


O_threshold = 110
MC for this TARGET:[78.922, 0.045]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.54, -4.64, -5.02]][[-5.01, -5.18, -5.69]][[-78.92, -78.92, -78.92]][-9.7]
std:[[0.3, 0.3, 0.03]][[0.13, 0.14, 0.12]][[0.0, 0.0, 0.0]][0.15]
MSE:[[4.55, 4.65, 5.02]][[5.01, 5.18, 5.69]][[78.92, 78.92, 78.92]][9.7]
MSE(-DR):[[0.0, 0.1, 0.47]][[0.46, 0.63, 1.14]][[74.37, 74.37, 74.37]][5.15]
***
MC-based ATE = -1.18
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-5.64, -5.66, -4.57]][[-7.51, -7.45, -7.61]][[1.18, 1.18, 1.18]][1.18]
std:[[0.87, 0.87, 0.08]][[0.08, 0.09, 0.09]][[0.0, 0.0, 0.0]][0.0]
MSE:[[5.71, 5.73, 4.57]][[7.51, 7.45, 7.61]][[1.18, 1.18, 1.18]][1.18]
MSE(-DR):[[0.0, 0.02, -1.14]][[1.8, 1.74, 1.9]][[-4.53, -4.53, -4.53]][-4.53]
**
==============


O_threshold = 120
MC for this TARGET:[80.153, 0.051]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.9, -7.96, -7.89]][[-9.01, -9.14, -9.64]][[-80.15, -80.15, -80.15]][-10.93]
std:[[0.38, 0.38, 0.15]][[0.08, 0.08, 0.12]][[0.0, 0.0, 0.0]][0.15]
MSE:[[7.91, 7.97, 7.89]][[9.01, 9.14, 9.64]][[80.15, 80.15, 80.15]][10.93]
MSE(-DR):[[0.0, 0.06, -0.02]][[1.1, 1.23, 1.73]][[72.24, 72.24, 72.24]][3.02]
**
MC-based ATE = 0.05
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-9.0, -8.97, -7.43]][[-11.51, -11.41, -11.56]][[-0.05, -0.05, -0.05]][-0.05]
std:[[0.93, 0.94, 0.12]][[0.04, 0.03, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[9.05, 9.02, 7.43]][[11.51, 11.41, 11.56]][[0.05, 0.05, 0.05]][0.05]
MSE(-DR):[[0.0, -0.03, -1.62]][[2.46, 2.36, 2.51]][[-9.0, -9.0, -9.0]][-9.0]
```

**
==========

O_threshold = 130
MC for this TARGET:[81.289, 0.047]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.08, -11.12, -10.41]][[-13.15, -13.23, -13.67]][[-81.29, -81.29, -81.29]][-12.07]
std:[[0.38, 0.38, 0.16]][[0.12, 0.11, 0.16]][[0.0, 0.0, 0.0]][0.15]
MSE:[[11.09, 11.13, 10.41]][[13.15, 13.23, 13.67]][[81.29, 81.29, 81.29]][12.07]
MSE(-DR):[[0.0, 0.04, -0.68]][[2.06, 2.14, 2.58]][[70.2, 70.2, 70.2]][0.98]
**
MC-based ATE = 1.19
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-12.18, -12.14, -9.95]][[-15.66, -15.5, -15.59]][[-1.19, -1.19, -1.19]][-1.19]
std:[[0.93, 0.94, 0.12]][[0.1, 0.08, 0.1]][[0.0, 0.0, 0.0]][0.0]
MSE:[[12.22, 12.18, 9.95]][[15.66, 15.5, 15.59]][[1.19, 1.19, 1.19]][1.19]
MSE(-DR):[[0.0, -0.04, -2.27]][[3.44, 3.28, 3.37]][[-11.03, -11.03, -11.03]][-11.03]
**
==========

time spent until now: 71.0 mins

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

target policy:

1 1 1

1 1 0

1 1 1

number of reward locations: 8
O_threshold = 90
target policy:

1 0 0

0 1 0

1 0 1

number of reward locations: 4
O_threshold = 100
target policy:

1 0 0

0 1 0

1 0 1

number of reward locations: 4
O_threshold = 110
target policy:

1 0 0

0 1 0

1 0 0

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

1 0 0

0 0 0

1 0 0

number of reward locations: 2
O_threshold = 130
target policy:

```
1 0 0

0 0 0

1 0 0

number of reward locations:  2
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:63.003
O_threshold = 80
MC for this TARGET:[69.635, 0.122]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[3.93, 3.78, 1.82]][[4.22, 3.95, 3.89]][[-69.64, -69.64, -69.64]][-6.63]
std:[[0.28, 0.32, 0.1]][[0.07, 0.09, 0.09]][[0.0, 0.0, 0.0]][0.17]
MSE:[[3.94, 3.79, 1.82]][[4.22, 3.95, 3.89]][[69.64, 69.64, 69.64]][6.63]
MSE(-DR):[[0.0, -0.15, -2.12]][[0.28, 0.01, -0.05]][[65.7, 65.7, 65.7]][2.69]
**
==============


O_threshold = 90
MC for this TARGET:[73.544, 0.107]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.91, -3.05, -4.11]][[-4.93, -5.16, -5.47]][[-73.54, -73.54, -73.54]][-10.54]
std:[[0.25, 0.23, 0.29]][[0.07, 0.06, 0.08]][[0.0, 0.0, 0.0]][0.17]
MSE:[[2.92, 3.06, 4.12]][[4.93, 5.16, 5.47]][[73.54, 73.54, 73.54]][10.54]
MSE(-DR):[[0.0, 0.14, 1.2]][[2.01, 2.24, 2.55]][[70.62, 70.62, 70.62]][7.62]
***
MC-based ATE = 3.91
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.84, -6.83, -5.93]][[-9.15, -9.11, -9.36]][[-3.91, -3.91, -3.91]][-3.91]
std:[[0.5, 0.51, 0.3]][[0.03, 0.04, 0.01]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.86, 6.85, 5.94]][[9.15, 9.11, 9.36]][[3.91, 3.91, 3.91]][3.91]
MSE(-DR):[[0.0, -0.01, -0.92]][[2.29, 2.25, 2.5]][[-2.95, -2.95, -2.95]][-2.95]
**
==============


O_threshold = 100
MC for this TARGET:[73.544, 0.107]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-2.88, -3.05, -4.11]][[-4.95, -5.16, -5.48]][[-73.54, -73.54, -73.54]][-10.54]
std:[[0.25, 0.23, 0.29]][[0.07, 0.06, 0.02]][[0.0, 0.0, 0.0]][0.17]
MSE:[[2.89, 3.06, 4.12]][[4.95, 5.16, 5.48]][[73.54, 73.54, 73.54]][10.54]
MSE(-DR):[[0.0, 0.17, 1.23]][[2.06, 2.27, 2.59]][[70.65, 70.65, 70.65]][7.65]
***
MC-based ATE = 3.91
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-6.82, -6.83, -5.93]][[-9.18, -9.11, -9.37]][[-3.91, -3.91, -3.91]][-3.91]
std:[[0.48, 0.51, 0.32]][[0.01, 0.04, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[6.84, 6.85, 5.94]][[9.18, 9.11, 9.37]][[3.91, 3.91, 3.91]][3.91]
MSE(-DR):[[0.0, 0.01, -0.9]][[2.34, 2.27, 2.53]][[-2.93, -2.93, -2.93]][-2.93]
**
==============


O_threshold = 110
MC for this TARGET:[72.14, 0.105]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-7.24, -7.33, -7.46]][[-7.77, -7.9, -8.15]][[-72.14, -72.14, -72.14]][-9.14]
std:[[0.67, 0.68, 0.26]][[0.08, 0.09, 0.14]][[0.0, 0.0, 0.0]][0.17]
MSE:[[7.27, 7.36, 7.46]][[7.77, 7.9, 8.15]][[72.14, 72.14, 72.14]][9.14]
MSE(-DR):[[0.0, 0.09, 0.19]][[0.5, 0.63, 0.88]][[64.87, 64.87, 64.87]][1.87]
***
MC-based ATE = 2.5
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.17, -11.11, -9.28]][[-11.99, -11.85, -12.03]][[-2.5, -2.5, -2.5]][-2.5]
std:[[0.9, 0.94, 0.23]][[0.15, 0.18, 0.13]][[0.0, 0.0, 0.0]][0.0]
MSE:[[11.21, 11.15, 9.28]][[11.99, 11.85, 12.03]][[2.5, 2.5, 2.5]][2.5]
MSE(-DR):[[0.0, -0.06, -1.93]][[0.78, 0.64, 0.82]][[-8.71, -8.71, -8.71]][-8.71]
**
==============


O_threshold = 120
MC for this TARGET:[82.508, 0.117]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.32, -11.43, -10.03]][[-17.41, -17.59, -18.17]][[-82.51, -82.51, -82.51]][-19.5]
std:[[0.62, 0.62, 0.12]][[0.05, 0.03, 0.02]][[0.0, 0.0, 0.0]][0.17]
MSE:[[11.34, 11.45, 10.03]][[17.41, 17.59, 18.17]][[82.51, 82.51, 82.51]][19.5]
MSE(-DR):[[0.0, 0.11, -1.31]][[6.07, 6.25, 6.83]][[71.17, 71.17, 71.17]][8.16]
**
MC-based ATE = 12.87
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
```

```
bias:[[-15.25, -15.21, -11.85]][[-21.63, -21.54, -22.06]][[-12.87, -12.87, -12.87]][-12.87]
std:[[0.85, 0.9, 0.16]][[0.08, 0.1, 0.08]][[0.0, 0.0, 0.0]][0.0]
MSE:[[15.27, 15.24, 11.85]][[21.63, 21.54, 22.06]][[12.87, 12.87, 12.87]][12.87]
MSE(-DR):[[0.0, -0.03, -3.42]][[6.36, 6.27, 6.79]][[-2.4, -2.4, -2.4]][-2.4]
**
==============


O_threshold = 130
MC for this TARGET:[82.508, 0.117]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-11.33, -11.43, -10.05]][[-17.42, -17.59, -18.22]][[-82.51, -82.51, -82.51]][-19.5]
std:[[0.6, 0.62, 0.11]][[0.06, 0.03, 0.04]][[0.0, 0.0, 0.0]][0.17]
MSE:[[11.35, 11.45, 10.05]][[17.42, 17.59, 18.22]][[82.51, 82.51, 82.51]][19.5]
MSE(-DR):[[0.0, 0.1, -1.3]][[6.07, 6.24, 6.87]][[71.16, 71.16, 71.16]][8.15]
**
MC-based ATE = 12.87
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-15.27, -15.21, -11.87]][[-21.65, -21.54, -22.1]][[-12.87, -12.87, -12.87]][-12.87]
std:[[0.83, 0.9, 0.14]][[0.1, 0.1, 0.07]][[0.0, 0.0, 0.0]][0.0]
MSE:[[15.29, 15.24, 11.87]][[21.65, 21.54, 22.1]][[12.87, 12.87, 12.87]][12.87]
MSE(-DR):[[0.0, -0.05, -3.42]][[6.36, 6.25, 6.81]][[-2.42, -2.42, -2.42]][-2.42]
**
==============


time spent until now: 76.0 mins


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

target policy:

1 1 1 1

1 0 1 1

1 1 1 0

1 1 1 0

number of reward locations:  13
O_threshold = 90
target policy:

1 0 0 0

1 0 1 0

1 1 1 0

1 1 1 0

number of reward locations:  9
O_threshold = 100
target policy:

1 0 0 0

1 0 1 0

1 0 1 0

0 0 1 0

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

1 0 0 0

1 0 1 0

0 0 1 0

0 0 1 0
```

```
number of reward locations:  5
O_threshold = 120
target policy:

1 0 0 0

0 0 1 0

0 0 1 0

0 0 1 0

number of reward locations:  4
O_threshold = 130
target policy:

1 0 0 0

0 0 1 0

0 0 1 0

0 0 0 0

number of reward locations:  3
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE

---------------------------------------
Value of Behaviour policy:65.286
O_threshold = 80
MC for this TARGET:[78.036, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[0.02, -0.11, -1.83]][[1.14, 0.87, 0.51]][[-78.04, -78.04, -78.04]][-12.75]
std:[[0.37, 0.36, 0.14]][[0.17, 0.15, 0.1]][[0.0, 0.0, 0.0]][0.17]
MSE:[[0.37, 0.38, 1.84]][[1.15, 0.88, 0.52]][[78.04, 78.04, 78.04]][12.75]
MSE(-DR):[[0.0, 0.01, 1.47]][[0.78, 0.51, 0.15]][[77.67, 77.67, 77.67]][12.38]
***
==============


O_threshold = 90
MC for this TARGET:[71.497, 0.08]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[2.47, 2.33, 1.23]][[2.17, 1.95, 1.63]][[-71.5, -71.5, -71.5]][-6.21]
std:[[0.07, 0.03, 0.09]][[0.19, 0.16, 0.08]][[0.0, 0.0, 0.0]][0.17]
MSE:[[2.47, 2.33, 1.23]][[2.18, 1.96, 1.63]][[71.5, 71.5, 71.5]][6.21]
MSE(-DR):[[0.0, -0.14, -1.24]][[-0.29, -0.51, -0.84]][[69.03, 69.03, 69.03]][3.74]
MC-based ATE = -6.54
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[2.45, 2.44, 3.05]][[1.03, 1.07, 1.11]][[6.54, 6.54, 6.54]][6.54]
std:[[0.36, 0.36, 0.17]][[0.02, 0.01, 0.06]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.48, 2.47, 3.05]][[1.03, 1.07, 1.11]][[6.54, 6.54, 6.54]][6.54]
MSE(-DR):[[0.0, -0.01, 0.57]][[-1.45, -1.41, -1.37]][[4.06, 4.06, 4.06]][4.06]
==============


O_threshold = 100
MC for this TARGET:[74.634, 0.083]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.8, -3.92, -4.73]][[-4.11, -4.29, -4.71]][[-74.63, -74.63, -74.63]][-9.35]
std:[[0.22, 0.24, 0.17]][[0.26, 0.23, 0.13]][[0.0, 0.0, 0.0]][0.17]
MSE:[[3.81, 3.93, 4.73]][[4.12, 4.3, 4.71]][[74.63, 74.63, 74.63]][9.35]
MSE(-DR):[[0.0, 0.12, 0.92]][[0.31, 0.49, 0.9]][[70.82, 70.82, 70.82]][5.54]
***
MC-based ATE = -3.4
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.82, -3.81, -2.91]][[-5.25, -5.16, -5.22]][[3.4, 3.4, 3.4]][3.4]
std:[[0.15, 0.13, 0.04]][[0.15, 0.16, 0.18]][[0.0, 0.0, 0.0]][0.0]
MSE:[[3.82, 3.81, 2.91]][[5.25, 5.16, 5.22]][[3.4, 3.4, 3.4]][3.4]
MSE(-DR):[[0.0, -0.01, -0.91]][[1.43, 1.34, 1.4]][[-0.42, -0.42, -0.42]][-0.42]
**
==============


O_threshold = 110
MC for this TARGET:[73.075, 0.083]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.91, -4.0, -4.52]][[-4.9, -5.05, -5.42]][[-73.08, -73.08, -73.08]][-7.79]
std:[[0.17, 0.18, 0.16]][[0.28, 0.24, 0.17]][[0.0, 0.0, 0.0]][0.17]
MSE:[[3.91, 4.0, 4.52]][[4.91, 5.06, 5.42]][[73.08, 73.08, 73.08]][7.79]
MSE(-DR):[[0.0, 0.09, 0.61]][[1.0, 1.15, 1.51]][[69.17, 69.17, 69.17]][3.88]
***
MC-based ATE = -4.96
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.92, -3.9, -2.7]][[-6.04, -5.92, -5.93]][[4.96, 4.96, 4.96]][4.96]
std:[[0.37, 0.37, 0.08]][[0.12, 0.11, 0.14]][[0.0, 0.0, 0.0]][0.0]
```

```
MSE:[[3.94, 3.92, 2.7]][[6.04, 5.92, 5.93]][[4.96, 4.96, 4.96]][4.96]
MSE(-DR):[[0.0, -0.02, -1.24]][[2.1, 1.98, 1.99]][[1.02, 1.02, 1.02]][1.02]
**
==============


O_threshold = 120
MC for this TARGET:[70.12, 0.079]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.1, -3.21, -3.01]][[-4.74, -4.82, -5.16]][[-70.12, -70.12, -70.12]][-4.83]
std:[[0.05, 0.03, 0.19]][[0.27, 0.23, 0.15]][[0.0, 0.0, 0.0]][0.17]
MSE:[[3.1, 3.21, 3.02]][[4.75, 4.83, 5.16]][[70.12, 70.12, 70.12]][4.83]
MSE(-DR):[[0.0, 0.11, -0.08]][[1.65, 1.73, 2.06]][[67.02, 67.02, 67.02]][1.73]
**
MC-based ATE = -7.92
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-3.12, -3.11, -1.18]][[-5.88, -5.69, -5.68]][[7.92, 7.92, 7.92]][7.92]
std:[[0.33, 0.32, 0.07]][[0.11, 0.09, 0.14]][[0.0, 0.0, 0.0]][0.0]
MSE:[[3.14, 3.13, 1.18]][[5.88, 5.69, 5.68]][[7.92, 7.92, 7.92]][7.92]
MSE(-DR):[[0.0, -0.01, -1.96]][[2.74, 2.55, 2.54]][[4.78, 4.78, 4.78]][4.78]
**
==============


O_threshold = 130
MC for this TARGET:[68.947, 0.073]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.24, -4.31, -3.57]][[-6.58, -6.6, -6.84]][[-68.95, -68.95, -68.95]][-3.66]
std:[[0.17, 0.17, 0.25]][[0.26, 0.22, 0.16]][[0.0, 0.0, 0.0]][0.17]
MSE:[[4.24, 4.31, 3.58]][[6.59, 6.6, 6.84]][[68.95, 68.95, 68.95]][3.66]
MSE(-DR):[[0.0, 0.07, -0.66]][[2.35, 2.36, 2.6]][[64.71, 64.71, 64.71]][-0.58]
**
MC-based ATE = -9.09
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [V_behav]
bias:[[-4.26, -4.21, -1.74]][[-7.72, -7.47, -7.35]][[9.09, 9.09, 9.09]][9.09]
std:[[0.2, 0.2, 0.11]][[0.13, 0.13, 0.18]][[0.0, 0.0, 0.0]][0.0]
MSE:[[4.26, 4.21, 1.74]][[7.72, 7.47, 7.35]][[9.09, 9.09, 9.09]][9.09]
MSE(-DR):[[0.0, -0.05, -2.52]][[3.46, 3.21, 3.09]][[4.83, 4.83, 4.83]][4.83]
**
==============


time spent until now: 83.8 mins


---------------------------------------
[pattern_seed, T, sd_R] = [1, 672, 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
O_threshold = 130
target policy:

1 0 0 0 0

0 1 0 0 0

1 0 0 0 0

0 0 0 0 0

0 0 0 0 0

number of reward locations:  3
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; 6 -th target; one rep DONE
1 -th target; 2 -th target; packet_write_wait: Connection to 3.219.215.112 port 22: Broken pipe
Run-Mac:.ssh mac$
```