```
Last login: Mon Apr 13 15:34:11 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-235-59-105.compute-1.amazonaws.com
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1060-aws x86_64)

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

 System information disabled due to load higher than 96.0

 * 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

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


*** System restart required ***
Last login: Mon Apr 13 19:34:41 2020 from 107.13.161.147
ubuntu@ip-172-31-6-10:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
16:16, 04/13; num of cores:96
sd_u_O_30_uo_ud_40

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 40, 0.5, 1.5, 40, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[85, 95, 100, 110, 120, 125], [0, 20, 40], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]
]

--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 0]

max(u_O) =  191.7
O_threshold = 85
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  13
O_threshold = 95
number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

--------------------------------------
Value of Behaviour policy:31.726
```

```
O_threshold = 85
MC for this TARGET:[36.67, 0.032]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.84, -1.01, -1.37]][[-0.64, -36.67, -4.94]]
std:[[0.34, 0.33, 0.25]][[0.27, 0.0, 0.14]]
MSE:[[0.91, 1.06, 1.39]][[0.69, 36.67, 4.94]]
MSE(-DR):[[0.0, 0.15, 0.48]][[-0.22, 35.76, 4.03]]
==============
O_threshold = 95
MC for this TARGET:[36.73, 0.036]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.83, -1.97, -1.76]][[-2.55, -36.73, -5.0]]
std:[[0.32, 0.34, 0.21]][[0.25, 0.0, 0.14]]
MSE:[[1.86, 2.0, 1.77]][[2.56, 36.73, 5.0]]
MSE(-DR):[[0.0, 0.14, -0.09]][[0.7, 34.87, 3.14]]
**
==============
O_threshold = 100
MC for this TARGET:[37.101, 0.028]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.67, -4.73, -4.46]][[-5.9, -37.1, -5.38]]
std:[[0.34, 0.35, 0.22]][[0.25, 0.0, 0.14]]
MSE:[[4.68, 4.74, 4.47]][[5.91, 37.1, 5.38]]
MSE(-DR):[[0.0, 0.06, -0.21]][[1.23, 32.42, 0.7]]
**
==============
O_threshold = 110
MC for this TARGET:[36.985, 0.026]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.48, -6.48, -5.81]][[-8.82, -36.98, -5.26]]
std:[[0.39, 0.41, 0.18]][[0.22, 0.0, 0.14]]
MSE:[[6.49, 6.49, 5.81]][[8.82, 36.98, 5.26]]
MSE(-DR):[[0.0, 0.0, -0.68]][[2.33, 30.49, -1.23]]
**
==============
O_threshold = 120
MC for this TARGET:[36.871, 0.03]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.84, -6.84, -6.46]][[-9.79, -36.87, -5.15]]
std:[[0.44, 0.45, 0.19]][[0.21, 0.0, 0.14]]
MSE:[[6.85, 6.85, 6.46]][[9.79, 36.87, 5.15]]
MSE(-DR):[[0.0, 0.0, -0.39]][[2.94, 30.02, -1.7]]
**
==============
O_threshold = 125
MC for this TARGET:[36.626, 0.028]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.45, -8.38, -8.89]][[-13.32, -36.63, -4.9]]
std:[[0.63, 0.66, 0.24]][[0.2, 0.0, 0.14]]
MSE:[[8.47, 8.41, 8.89]][[13.32, 36.63, 4.9]]
MSE(-DR):[[0.0, -0.06, 0.42]][[4.85, 28.16, -3.57]]
***
==============
[[ 0.91  1.06  1.39  0.69 36.67  4.94]
 [ 1.86  2.    1.77  2.56 36.73  5.  ]
 [ 4.68  4.74  4.47  5.91 37.1   5.38]
 [ 6.49  6.49  5.81  8.82 36.98  5.26]
 [ 6.85  6.85  6.46  9.79 36.87  5.15]
 [ 8.47  8.41  8.89 13.32 36.63  4.9 ]]


time spent until now: 17.3 mins

16:33, 04/13

---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 20]

max(u_O) =  191.7
O_threshold = 85
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1
```

```
0 1 1 0 1

0 1 1 0 1

number of reward locations:  13
O_threshold = 95
number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

--------------------------------------
Value of Behaviour policy:31.697
O_threshold = 85
MC for this TARGET:[36.702, 0.236]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.83, -0.94, -1.32]][[-0.86, -36.7, -5.0]]
std:[[1.18, 1.17, 0.81]][[0.65, 0.0, 0.26]]
MSE:[[1.44, 1.5, 1.55]][[1.08, 36.7, 5.01]]
MSE(-DR):[[0.0, 0.06, 0.11]][[-0.36, 35.26, 3.57]]
==============
O_threshold = 95
MC for this TARGET:[36.762, 0.237]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.7, -1.87, -1.69]][[-2.74, -36.76, -5.06]]
std:[[0.87, 0.9, 0.56]][[0.57, 0.0, 0.26]]
MSE:[[1.91, 2.08, 1.78]][[2.8, 36.76, 5.07]]
MSE(-DR):[[0.0, 0.17, -0.13]][[0.89, 34.85, 3.16]]
**
==============
O_threshold = 100
MC for this TARGET:[37.133, 0.236]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.41, -4.48, -4.34]][[-6.02, -37.13, -5.44]]
std:[[0.86, 0.89, 0.5]][[0.58, 0.0, 0.26]]
MSE:[[4.49, 4.57, 4.37]][[6.05, 37.13, 5.45]]
MSE(-DR):[[0.0, 0.08, -0.12]][[1.56, 32.64, 0.96]]
**
==============
O_threshold = 110
MC for this TARGET:[37.016, 0.235]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.59, -6.57, -5.73]][[-8.85, -37.02, -5.32]]
std:[[1.04, 1.09, 0.52]][[0.5, 0.0, 0.26]]
MSE:[[6.67, 6.66, 5.75]][[8.86, 37.02, 5.33]]
MSE(-DR):[[0.0, -0.01, -0.92]][[2.19, 30.35, -1.34]]
**
==============
O_threshold = 120
MC for this TARGET:[36.903, 0.233]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.02, -6.98, -6.55]][[-9.89, -36.9, -5.21]]
std:[[1.04, 1.08, 0.61]][[0.45, 0.0, 0.26]]
MSE:[[7.1, 7.06, 6.58]][[9.9, 36.9, 5.22]]
MSE(-DR):[[0.0, -0.04, -0.52]][[2.8, 29.8, -1.88]]
**
==============
O_threshold = 125
MC for this TARGET:[36.657, 0.236]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.94, -8.84, -9.15]][[-13.41, -36.66, -4.96]]
std:[[1.22, 1.31, 0.7]][[0.51, 0.0, 0.26]]
MSE:[[9.02, 8.94, 9.18]][[13.42, 36.66, 4.97]]
MSE(-DR):[[0.0, -0.08, 0.16]][[4.4, 27.64, -4.05]]
***
==============
[[ 0.91  1.06  1.39  0.69 36.67  4.94]
 [ 1.86  2.    1.77  2.56 36.73  5.  ]
 [ 4.68  4.74  4.47  5.91 37.1   5.38]
 [ 6.49  6.49  5.81  8.82 36.98  5.26]
 [ 6.85  6.85  6.46  9.79 36.87  5.15]
 [ 8.47  8.41  8.89 13.32 36.63  4.9 ]]


[[ 1.44  1.5   1.55  1.08 36.7   5.01]
 [ 1.91  2.08  1.78  2.8  36.76  5.07]
 [ 4.49  4.57  4.37  6.05 37.13  5.45]
```

```
 [ 6.67  6.66  5.75  8.86 37.02  5.33]
 [ 7.1   7.06  6.58  9.9  36.9   5.22]
 [ 9.02  8.94  9.18 13.42 36.66  4.97]]


time spent until now: 34.6 mins

16:50, 04/13

--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 40]

max(u_O) =  191.7
O_threshold = 85
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  13
O_threshold = 95
number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

--------------------------------------
Value of Behaviour policy:31.668
O_threshold = 85
MC for this TARGET:[36.733, 0.466]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.77, -0.89, -1.17]][[-1.09, -36.73, -5.06]]
std:[[2.15, 2.14, 1.57]][[1.11, 0.0, 0.47]]
MSE:[[2.28, 2.32, 1.96]][[1.56, 36.73, 5.08]]
MSE(-DR):[[0.0, 0.04, -0.32]][[-0.72, 34.45, 2.8]]
==============
O_threshold = 95
MC for this TARGET:[36.793, 0.466]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.65, -1.78, -1.7]][[-2.92, -36.79, -5.12]]
std:[[1.67, 1.69, 1.14]][[0.99, 0.0, 0.47]]
MSE:[[2.35, 2.45, 2.05]][[3.08, 36.79, 5.14]]
MSE(-DR):[[0.0, 0.1, -0.3]][[0.73, 34.44, 2.79]]
**
==============
O_threshold = 100
MC for this TARGET:[37.165, 0.466]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.17, -4.22, -4.23]][[-6.17, -37.16, -5.5]]
std:[[1.51, 1.58, 0.93]][[0.97, 0.0, 0.47]]
MSE:[[4.43, 4.51, 4.33]][[6.25, 37.16, 5.52]]
MSE(-DR):[[0.0, 0.08, -0.1]][[1.82, 32.73, 1.09]]
**
==============
O_threshold = 110
MC for this TARGET:[37.048, 0.465]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.67, -6.67, -5.67]][[-8.9, -37.05, -5.38]]
std:[[1.93, 2.01, 0.99]][[0.85, 0.0, 0.47]]
```

```
MSE:[[6.94, 6.97, 5.76]][[8.94, 37.05, 5.4]]
MSE(-DR):[[0.0, 0.03, -1.18]][[2.0, 30.11, -1.54]]
**
==============
O_threshold = 120
MC for this TARGET:[36.934, 0.463]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.13, -7.1, -6.56]][[-9.98, -36.93, -5.27]]
std:[[1.9, 1.91, 1.2]][[0.76, 0.0, 0.47]]
MSE:[[7.38, 7.35, 6.67]][[10.01, 36.93, 5.29]]
MSE(-DR):[[0.0, -0.03, -0.71]][[2.63, 29.55, -2.09]]
**
==============
O_threshold = 125
MC for this TARGET:[36.689, 0.466]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.37, -9.27, -9.33]][[-13.51, -36.69, -5.02]]
std:[[2.04, 2.16, 1.48]][[0.92, 0.0, 0.47]]
MSE:[[9.59, 9.52, 9.45]][[13.54, 36.69, 5.04]]
MSE(-DR):[[0.0, -0.07, -0.14]][[3.95, 27.1, -4.55]]
**
==============
[[ 0.91  1.06  1.39  0.69 36.67  4.94]
 [ 1.86  2.    1.77  2.56 36.73  5.  ]
 [ 4.68  4.74  4.47  5.91 37.1   5.38]
 [ 6.49  6.49  5.81  8.82 36.98  5.26]
 [ 6.85  6.85  6.46  9.79 36.87  5.15]
 [ 8.47  8.41  8.89 13.32 36.63  4.9 ]]


[[ 1.44  1.5   1.55  1.08 36.7   5.01]
 [ 1.91  2.08  1.78  2.8  36.76  5.07]
 [ 4.49  4.57  4.37  6.05 37.13  5.45]
 [ 6.67  6.66  5.75  8.86 37.02  5.33]
 [ 7.1   7.06  6.58  9.9  36.9   5.22]
 [ 9.02  8.94  9.18 13.42 36.66  4.97]]


[[ 2.28  2.32  1.96  1.56 36.73  5.08]
 [ 2.35  2.45  2.05  3.08 36.79  5.14]
 [ 4.43  4.51  4.33  6.25 37.16  5.52]
 [ 6.94  6.97  5.76  8.94 37.05  5.4 ]
 [ 7.38  7.35  6.67 10.01 36.93  5.29]
 [ 9.59  9.52  9.45 13.54 36.69  5.04]]


time spent until now: 51.9 mins

17:08, 04/13
ubuntu@ip-172-31-6-10:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
17:09, 04/13; num of cores:96
sd_u_O_30_uo_ud_40

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 40, 0.5, 1.5, 30, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[85, 95, 100, 110, 120, 125], [0, 20, 40], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]
]

--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 0]

max(u_O) =  191.7
O_threshold = 85
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  13
O_threshold = 95
```

```
number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

----------------------------------------
Value of Behaviour policy:38.251
O_threshold = 85
MC for this TARGET:[46.873, 0.047]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.39, -0.65, -1.39]][[-0.43, -46.87, -8.62]]
std:[[0.42, 0.42, 0.22]][[0.36, 0.0, 0.19]]
MSE:[[0.57, 0.77, 1.41]][[0.56, 46.87, 8.62]]
MSE(-DR):[[0.0, 0.2, 0.84]][[-0.01, 46.3, 8.05]]
==============
O_threshold = 95
MC for this TARGET:[46.11, 0.051]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.28, -1.49, -1.59]][[-2.2, -46.11, -7.86]]
std:[[0.38, 0.39, 0.24]][[0.33, 0.0, 0.19]]
MSE:[[1.34, 1.54, 1.61]][[2.22, 46.11, 7.86]]
MSE(-DR):[[0.0, 0.2, 0.27]][[0.88, 44.77, 6.52]]
***
==============
O_threshold = 100
MC for this TARGET:[47.187, 0.047]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.13, -5.26, -5.19]][[-6.72, -47.19, -8.94]]
std:[[0.44, 0.46, 0.25]][[0.33, 0.0, 0.19]]
MSE:[[5.15, 5.28, 5.2]][[6.73, 47.19, 8.94]]
MSE(-DR):[[0.0, 0.13, 0.05]][[1.58, 42.04, 3.79]]
***
==============
O_threshold = 110
MC for this TARGET:[47.605, 0.043]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.16, -8.21, -7.82]][[-10.9, -47.6, -9.35]]
std:[[0.47, 0.5, 0.19]][[0.28, 0.0, 0.19]]
MSE:[[8.17, 8.23, 7.82]][[10.9, 47.6, 9.35]]
MSE(-DR):[[0.0, 0.06, -0.35]][[2.73, 39.43, 1.18]]
**
==============
O_threshold = 120
MC for this TARGET:[46.829, 0.043]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.28, -8.3, -8.16]][[-11.65, -46.83, -8.58]]
std:[[0.49, 0.51, 0.25]][[0.27, 0.0, 0.19]]
MSE:[[8.29, 8.32, 8.16]][[11.65, 46.83, 8.58]]
MSE(-DR):[[0.0, 0.03, -0.13]][[3.36, 38.54, 0.29]]
**
==============
O_threshold = 125
MC for this TARGET:[45.542, 0.047]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.86, -9.81, -10.07]][[-15.53, -45.54, -7.29]]
std:[[0.79, 0.82, 0.23]][[0.26, 0.0, 0.19]]
MSE:[[9.89, 9.84, 10.07]][[15.53, 45.54, 7.29]]
MSE(-DR):[[0.0, -0.05, 0.18]][[5.64, 35.65, -2.6]]
***
==============
[[ 0.57  0.77  1.41  0.56 46.87  8.62]
 [ 1.34  1.54  1.61  2.22 46.11  7.86]
 [ 5.15  5.28  5.2   6.73 47.19  8.94]
 [ 8.17  8.23  7.82 10.9  47.6   9.35]
 [ 8.29  8.32  8.16 11.65 46.83  8.58]
 [ 9.89  9.84 10.07 15.53 45.54  7.29]]


time spent until now: 17.3 mins

17:26, 04/13

----------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 20]

max(u_O) =  191.7
O_threshold = 85
means of Order:
```

```
83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  13
O_threshold = 95
number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

---------------------------------------
Value of Behaviour policy:38.222
O_threshold = 85
MC for this TARGET:[46.905, 0.237]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.34, -0.58, -1.35]][[-0.67, -46.9, -8.68]]
std:[[1.24, 1.28, 0.72]][[0.74, 0.0, 0.26]]
MSE:[[1.29, 1.41, 1.53]][[1.0, 46.9, 8.68]]
MSE(-DR):[[0.0, 0.12, 0.24]][[-0.29, 45.61, 7.39]]
==============
O_threshold = 95
MC for this TARGET:[46.141, 0.234]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.21, -1.39, -1.56]][[-2.39, -46.14, -7.92]]
std:[[0.93, 0.98, 0.55]][[0.67, 0.0, 0.26]]
MSE:[[1.53, 1.7, 1.65]][[2.48, 46.14, 7.92]]
MSE(-DR):[[0.0, 0.17, 0.12]][[0.95, 44.61, 6.39]]
***
==============
O_threshold = 100
MC for this TARGET:[47.219, 0.23]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.91, -5.01, -5.16]][[-6.86, -47.22, -9.0]]
std:[[0.98, 1.02, 0.49]][[0.66, 0.0, 0.26]]
MSE:[[5.01, 5.11, 5.18]][[6.89, 47.22, 9.0]]
MSE(-DR):[[0.0, 0.1, 0.17]][[1.88, 42.21, 3.99]]
***
==============
O_threshold = 110
MC for this TARGET:[47.637, 0.233]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.26, -8.3, -7.79]][[-10.93, -47.64, -9.41]]
std:[[1.07, 1.12, 0.55]][[0.56, 0.0, 0.26]]
MSE:[[8.33, 8.38, 7.81]][[10.94, 47.64, 9.41]]
MSE(-DR):[[0.0, 0.05, -0.52]][[2.61, 39.31, 1.08]]
**
==============
O_threshold = 120
MC for this TARGET:[46.861, 0.236]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.4, -8.43, -8.17]][[-11.76, -46.86, -8.64]]
std:[[1.09, 1.11, 0.67]][[0.51, 0.0, 0.26]]
MSE:[[8.47, 8.5, 8.2]][[11.77, 46.86, 8.64]]
MSE(-DR):[[0.0, 0.03, -0.27]][[3.3, 38.39, 0.17]]
**
==============
O_threshold = 125
MC for this TARGET:[45.573, 0.239]
```

```
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-10.43, -10.36, -10.35]][[-15.63, -45.57, -7.35]]
std:[[1.34, 1.41, 0.68]][[0.55, 0.0, 0.26]]
MSE:[[10.52, 10.46, 10.37]][[15.64, 45.57, 7.35]]
MSE(-DR):[[0.0, -0.06, -0.15]][[5.12, 35.05, -3.17]]
**
==============
[[ 0.57  0.77  1.41  0.56 46.87  8.62]
 [ 1.34  1.54  1.61  2.22 46.11  7.86]
 [ 5.15  5.28  5.2   6.73 47.19  8.94]
 [ 8.17  8.23  7.82 10.9  47.6   9.35]
 [ 8.29  8.32  8.16 11.65 46.83  8.58]
 [ 9.89  9.84 10.07 15.53 45.54  7.29]]


[[ 1.29  1.41  1.53  1.   46.9   8.68]
 [ 1.53  1.7   1.65  2.48 46.14  7.92]
 [ 5.01  5.11  5.18  6.89 47.22  9.  ]
 [ 8.33  8.38  7.81 10.94 47.64  9.41]
 [ 8.47  8.5   8.2  11.77 46.86  8.64]
 [10.52 10.46 10.37 15.64 45.57  7.35]]


time spent until now: 34.5 mins

17:44, 04/13

---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 40]

max(u_O) =  191.7
O_threshold = 85
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  13
O_threshold = 95
number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
number of reward locations:  3
```