```
Last login: Mon Mar 30 17:29:56 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-18-232-178-254.compute-1.amazonaws.com
The authenticity of host 'ec2-18-232-178-254.compute-1.amazonaws.com (18.232.178.254)' can't be established.
ECDSA key fingerprint is SHA256:dAthr73DXTNWRKG02wk3Ou5JId+UDaG1HzYTS+EjwOA.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-18-232-178-254.compute-1.amazonaws.com,18.232.178.254' (ECDSA) to the list of known hos
ts.
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 01:14:25 UTC 2020

  System load:  0.77               Processes:           221
  Usage of /:   55.4% of 15.45GB   Users logged in:     0
  Memory usage: 1%                 IP address for ens5: 172.31.8.160
  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-172-31-8-160:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1
ubuntu@ip-172-31-8-160:~$ python EC2.py
21:16, 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] = [None, 5, 5, 10,
0.2, 1, 1, 1e-05, False, True, 5, False]


---------------------------------------
[pattern_seed, T, sd_R] = [0, 144, 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
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:77.041
O_threshold = 80
MC for this TARGET:[86.048, 0.239]
```

```
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.84, 0.8, -0.1]][[2.95, 2.81, 2.11]][[-86.05, -86.05]][[-86.05, -0.14, -9.01]]
std:[[0.94, 0.94, 0.32]][[0.44, 0.43, 0.43]][[0.0, 0.0]][[0.0, 0.33, 0.2]]
MSE:[[1.26, 1.23, 0.34]][[2.98, 2.84, 2.15]][[86.05, 86.05]][[86.05, 0.36, 9.01]]
MSE(-DR):[[0.0, -0.03, -0.92]][[1.72, 1.58, 0.89]][[84.79, 84.79]][[84.79, -0.9, 7.75]]
**
==============


O_threshold = 90
MC for this TARGET:[84.263, 0.216]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.22, 1.18, -0.22]][[3.8, 3.67, 2.83]][[-84.26, -84.26]][[-84.26, -0.26, -7.22]]
std:[[0.72, 0.72, 0.31]][[0.26, 0.26, 0.29]][[0.0, 0.0]][[0.0, 0.32, 0.2]]
MSE:[[1.42, 1.38, 0.38]][[3.81, 3.68, 2.84]][[84.26, 84.26]][[84.26, 0.41, 7.22]]
MSE(-DR):[[0.0, -0.04, -1.04]][[2.39, 2.26, 1.42]][[82.84, 82.84]][[82.84, -1.01, 5.8]]
**
MC-based ATE = -1.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.38, 0.38, -0.11]][[0.85, 0.85, 0.72]][[1.78, 1.78]][-0.12]
std:[[0.28, 0.29, 0.01]][[0.21, 0.19, 0.14]][[0.0, 0.0]][0.01]
MSE:[[0.47, 0.48, 0.11]][[0.88, 0.87, 0.73]][[1.78, 1.78]][0.12]
MSE(-DR):[[0.0, 0.01, -0.36]][[0.41, 0.4, 0.26]][[1.31, 1.31]][-0.35]
**
==============


O_threshold = 100
MC for this TARGET:[87.926, 0.209]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.02, -1.08, -3.15]][[0.88, 0.71, -0.36]][[-87.93, -87.93]][[-87.93, -3.22, -10.89]]
std:[[0.23, 0.22, 0.32]][[0.29, 0.26, 0.3]][[0.0, 0.0]][[0.0, 0.31, 0.2]]
MSE:[[1.05, 1.1, 3.17]][[0.93, 0.76, 0.47]][[87.93, 87.93]][[87.93, 3.23, 10.89]]
MSE(-DR):[[0.0, 0.05, 2.12]][[-0.12, -0.29, -0.58]][[86.88, 86.88]][[86.88, 2.18, 9.84]]
MC-based ATE = 1.88
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.86, -1.88, -3.05]][[-2.07, -2.1, -2.47]][[-1.88, -1.88]][-3.07]
std:[[0.71, 0.72, 0.63]][[0.29, 0.25, 0.13]][[0.0, 0.0]][0.63]
MSE:[[1.99, 2.01, 3.11]][[2.09, 2.11, 2.47]][[1.88, 1.88]][3.13]
MSE(-DR):[[0.0, 0.02, 1.12]][[0.1, 0.12, 0.48]][[-0.11, -0.11]][1.14]
**
==============


O_threshold = 110
MC for this TARGET:[85.587, 0.211]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.65, -2.72, -3.45]][[-2.53, -2.66, -3.67]][[-85.59, -85.59]][[-85.59, -3.51, -8.55]]
std:[[0.84, 0.83, 0.19]][[0.34, 0.32, 0.31]][[0.0, 0.0]][[0.0, 0.19, 0.2]]
MSE:[[2.78, 2.84, 3.46]][[2.55, 2.68, 3.68]][[85.59, 85.59]][[85.59, 3.52, 8.55]]
MSE(-DR):[[0.0, 0.06, 0.68]][[-0.23, -0.1, 0.9]][[82.81, 82.81]][[82.81, 0.74, 5.77]]
MC-based ATE = -0.46
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.5, -3.52, -3.34]][[-5.48, -5.47, -5.78]][[0.46, 0.46]][-3.36]
std:[[0.99, 0.98, 0.35]][[0.55, 0.54, 0.28]][[0.0, 0.0]][0.35]
MSE:[[3.64, 3.65, 3.36]][[5.51, 5.5, 5.79]][[0.46, 0.46]][3.38]
MSE(-DR):[[0.0, 0.01, -0.28]][[1.87, 1.86, 2.15]][[-3.18, -3.18]][-0.26]
**
==============


O_threshold = 120
MC for this TARGET:[86.747, 0.215]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-7.58, -7.61, -7.25]][[-6.87, -6.98, -8.0]][[-86.75, -86.75]][[-86.75, -7.27, -9.71]]
std:[[0.91, 0.93, 0.72]][[0.32, 0.33, 0.21]][[0.0, 0.0]][[0.0, 0.73, 0.2]]
MSE:[[7.63, 7.67, 7.29]][[6.88, 6.99, 8.0]][[86.75, 86.75]][[86.75, 7.31, 9.71]]
MSE(-DR):[[0.0, 0.04, -0.34]][[-0.75, -0.64, 0.37]][[79.12, 79.12]][[79.12, -0.32, 2.08]]
MC-based ATE = 0.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-8.42, -8.41, -7.14]][[-9.82, -9.8, -10.11]][[-0.7, -0.7]][-7.13]
std:[[1.21, 1.23, 1.0]][[0.56, 0.55, 0.37]][[0.0, 0.0]][1.03]
MSE:[[8.51, 8.5, 7.21]][[9.84, 9.82, 10.12]][[0.7, 0.7]][7.2]
MSE(-DR):[[0.0, -0.01, -1.3]][[1.33, 1.31, 1.61]][[-7.81, -7.81]][-1.31]
**
==============
```

```
O_threshold = 130
MC for this TARGET:[89.34, 0.223]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-8.74, -8.77, -9.29]][[-11.03, -11.16, -12.3]][[-89.34, -89.34]][[-89.34, -9.32, -12.3]]
std:[[1.98, 1.99, 0.6]][[0.61, 0.6, 0.34]][[0.0, 0.0]][[0.0, 0.62, 0.2]]
MSE:[[8.96, 8.99, 9.31]][[11.05, 11.18, 12.3]][[89.34, 89.34]][[89.34, 9.34, 12.3]]
MSE(-DR):[[0.0, 0.03, 0.35]][[2.09, 2.22, 3.34]][[80.38, 80.38]][[80.38, 0.38, 3.34]]
******
MC-based ATE = 3.29
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-9.58, -9.57, -9.19]][[-13.98, -13.98, -14.42]][[-3.29, -3.29]][-9.18]
std:[[1.77, 1.81, 0.89]][[0.78, 0.77, 0.46]][[0.0, 0.0]][0.92]
MSE:[[9.74, 9.74, 9.23]][[14.0, 14.0, 14.43]][[3.29, 3.29]][9.23]
MSE(-DR):[[0.0, 0.0, -0.51]][[4.26, 4.26, 4.69]][[-6.45, -6.45]][-0.51]
**
==============


time spent until now: 11.9 mins


---------------------------------------
[pattern_seed, T, sd_R] = [0, 336, 10]

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

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1
```

```
number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 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:76.562
O_threshold = 80
MC for this TARGET:[86.044, 0.142]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.55, -0.58, -0.18]][[2.2, 2.07, 1.31]][[-86.04, -86.04]][[-86.04, -0.22, -9.48]]
std:[[0.7, 0.69, 0.17]][[0.21, 0.23, 0.24]][[0.0, 0.0]][[0.0, 0.15, 0.2]]
MSE:[[0.89, 0.9, 0.25]][[2.21, 2.08, 1.33]][[86.04, 86.04]][[86.04, 0.27, 9.48]]
MSE(-DR):[[0.0, 0.01, -0.64]][[1.32, 1.19, 0.44]][[85.15, 85.15]][[85.15, -0.62, 8.59]]
**
==============


O_threshold = 90
MC for this TARGET:[84.276, 0.139]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.22, -0.25, -0.39]][[2.97, 2.82, 1.99]][[-84.28, -84.28]][[-84.28, -0.42, -7.71]]
std:[[0.5, 0.5, 0.16]][[0.27, 0.28, 0.24]][[0.0, 0.0]][[0.0, 0.15, 0.2]]
MSE:[[0.55, 0.56, 0.42]][[2.98, 2.83, 2.0]][[84.28, 84.28]][[84.28, 0.45, 7.71]]
MSE(-DR):[[0.0, 0.01, -0.13]][[2.43, 2.28, 1.45]][[83.73, 83.73]][[83.73, -0.1, 7.16]]
**
MC-based ATE = -1.77
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.33, 0.34, -0.21]][[0.77, 0.74, 0.68]][[1.77, 1.77]][-0.2]
std:[[0.22, 0.21, 0.01]][[0.07, 0.07, 0.07]][[0.0, 0.0]][0.0]
MSE:[[0.4, 0.4, 0.21]][[0.77, 0.74, 0.68]][[1.77, 1.77]][0.2]
MSE(-DR):[[0.0, 0.0, -0.19]][[0.37, 0.34, 0.28]][[1.37, 1.37]][-0.2]
**
==============
```

```
O_threshold = 100
MC for this TARGET:[87.954, 0.138]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.5, -1.54, -3.41]][[0.14, -0.04, -1.2]][[-87.95, -87.95]][[-87.95, -3.45, -11.39]]
std:[[0.52, 0.52, 0.12]][[0.33, 0.33, 0.32]][[0.0, 0.0]][[0.0, 0.13, 0.2]]
MSE:[[1.59, 1.63, 3.41]][[0.36, 0.33, 1.24]][[87.95, 87.95]][[87.95, 3.45, 11.39]]
MSE(-DR):[[0.0, 0.04, 1.82]][[-1.23, -1.26, -0.35]][[86.36, 86.36]][[86.36, 1.86, 9.8]]
MC-based ATE = 1.91
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.95, -0.96, -3.22]][[-2.06, -2.11, -2.51]][[-1.91, -1.91]][-3.23]
std:[[0.82, 0.8, 0.06]][[0.16, 0.15, 0.11]][[0.0, 0.0]][0.04]
MSE:[[1.25, 1.25, 3.22]][[2.07, 2.12, 2.51]][[1.91, 1.91]][3.23]
MSE(-DR):[[0.0, 0.0, 1.97]][[0.82, 0.87, 1.26]][[0.66, 0.66]][1.98]
**
==============


O_threshold = 110
MC for this TARGET:[85.608, 0.14]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.34, -2.4, -3.64]][[-2.88, -3.02, -4.22]][[-85.61, -85.61]][[-85.61, -3.71, -9.05]]
std:[[0.46, 0.48, 0.4]][[0.25, 0.26, 0.29]][[0.0, 0.0]][[0.0, 0.4, 0.2]]
MSE:[[2.38, 2.45, 3.66]][[2.89, 3.03, 4.23]][[85.61, 85.61]][[85.61, 3.73, 9.05]]
MSE(-DR):[[0.0, 0.07, 1.28]][[0.51, 0.65, 1.85]][[83.23, 83.23]][[83.23, 1.35, 6.67]]
******
MC-based ATE = -0.44
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.79, -1.82, -3.46]][[-5.08, -5.09, -5.53]][[0.44, 0.44]][-3.49]
std:[[0.42, 0.43, 0.43]][[0.09, 0.08, 0.09]][[0.0, 0.0]][0.41]
MSE:[[1.84, 1.87, 3.49]][[5.08, 5.09, 5.53]][[0.44, 0.44]][3.51]
MSE(-DR):[[0.0, 0.03, 1.65]][[3.24, 3.25, 3.69]][[-1.4, -1.4]][1.67]
**
==============


O_threshold = 120
MC for this TARGET:[86.751, 0.137]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-6.54, -6.61, -7.39]][[-7.16, -7.26, -8.36]][[-86.75, -86.75]][[-86.75, -7.45, -10.19]]
std:[[1.1, 1.1, 0.59]][[0.29, 0.3, 0.27]][[0.0, 0.0]][[0.0, 0.55, 0.2]]
MSE:[[6.63, 6.7, 7.41]][[7.17, 7.27, 8.36]][[86.75, 86.75]][[86.75, 7.47, 10.19]]
MSE(-DR):[[0.0, 0.07, 0.78]][[0.54, 0.64, 1.73]][[80.12, 80.12]][[80.12, 0.84, 3.56]]
******
MC-based ATE = 0.71
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-6.0, -6.02, -7.2]][[-9.36, -9.34, -9.67]][[-0.71, -0.71]][-7.23]
std:[[0.51, 0.55, 0.54]][[0.11, 0.11, 0.07]][[0.0, 0.0]][0.48]
MSE:[[6.02, 6.05, 7.22]][[9.36, 9.34, 9.67]][[0.71, 0.71]][7.25]
MSE(-DR):[[0.0, 0.03, 1.2]][[3.34, 3.32, 3.65]][[-5.31, -5.31]][1.23]
**
==============


O_threshold = 130
MC for this TARGET:[89.341, 0.139]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-9.88, -9.93, -9.92]][[-11.56, -11.62, -12.67]][[-89.34, -89.34]][[-89.34, -9.97, -12.78]]
std:[[0.65, 0.68, 0.57]][[0.22, 0.22, 0.26]][[0.0, 0.0]][[0.0, 0.58, 0.2]]
MSE:[[9.9, 9.95, 9.94]][[11.56, 11.62, 12.67]][[89.34, 89.34]][[89.34, 9.99, 12.78]]
MSE(-DR):[[0.0, 0.05, 0.04]][[1.66, 1.72, 2.77]][[79.44, 79.44]][[79.44, 0.09, 2.88]]
******
MC-based ATE = 3.3
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-9.33, -9.35, -9.74]][[-13.76, -13.7, -13.98]][[-3.3, -3.3]][-9.75]
std:[[0.36, 0.4, 0.45]][[0.12, 0.14, 0.1]][[0.0, 0.0]][0.46]
MSE:[[9.34, 9.36, 9.75]][[13.76, 13.7, 13.98]][[3.3, 3.3]][9.76]
MSE(-DR):[[0.0, 0.02, 0.41]][[4.42, 4.36, 4.64]][[-6.04, -6.04]][0.42]
**
==============


time spent until now: 23.8 mins


--------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 10]

max(u_O) =  156.6
```

```
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 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; ^CProcess Process-832:
Process Process-826:
Process Process-819:
Process Process-831:
Process Process-818:
Traceback (most recent call last):
  File "EC2.py", line 71, in <module>
Process Process-821:
    print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 62, in simu
Process Process-827:
Process Process-829:
    value_reps = rep_seeds(once, OPE_rep_times)
  File "/home/ubuntu/_uti_basic.py", line 119, in rep_seeds
    return list(map(fun, range(rep_times)))
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 187, in simu_once
Process Process-817:
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 130, in V_DR
Process Process-825:
    r = arr(parmap(getOneRegionValue, range(N), n_cores))
  File "/home/ubuntu/_uti_basic.py", line 74, in parmap
    sent = [q_in.put((i, x)) for i, x in enumerate(X)]
  File "/home/ubuntu/_uti_basic.py", line 74, in <listcomp>
    sent = [q_in.put((i, x)) for i, x in enumerate(X)]
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
Process Process-822:
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 283, in train
    subsamples = np.random.choice(N, batch_size)
KeyboardInterrupt
Process Process-823:
Process Process-830:
Process Process-820:
Traceback (most recent call last):
Traceback (most recent call last):
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
Traceback (most recent call last):
Traceback (most recent call last):
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
```

```
Process Process-824:
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
```

```
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
```

```
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
Traceback (most recent call last):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
KeyboardInterrupt
KeyboardInterrupt
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
Traceback (most recent call last):
Process Process-828:
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
```

```
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
Traceback (most recent call last):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1145, in _run
    not subfeed_t.get_shape().is_compatible_with(np_val.shape)):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/tensor_shape.py", line 1081, in is_
compatible_with
    other = as_shape(other)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/tensor_shape.py", line 1204, in as_
shape
    return TensorShape(shape)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/tensor_shape.py", line 755, in __in
it__
    elif isinstance(dims, tensor_shape_pb2.TensorShapeProto):
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
```

```
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 283, in train
    subsamples = np.random.choice(N, batch_size)
  File "mtrand.pyx", line 1152, in mtrand.RandomState.choice
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/numpy/core/fromnumeric.py", line 2772, in prod
    initial=initial)
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/numpy/core/fromnumeric.py", line 86, in _wrapreduction
    return ufunc.reduce(obj, axis, dtype, out, **passkwargs)
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/main.py", line 58, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 236, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_se
ssionrun
    run_metadata)
KeyboardInterrupt
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
ubuntu@ip-172-31-8-160:~$ python EC2.py
21:42, 03/30; num of cores:16

Basic setting:[T, sd_0, sd_D, sd_R, sd_u_0, w_0, w_A, lam, simple, M_in_R, u_0_u_D, mean_reversion] = [None, 5, 5, 10,
0.2, 1, 1, 0.001, False, True, 5, False]


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

```
[pattern_seed, T, sd_R] = [0, 144, 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
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:77.041
O_threshold = 80
MC for this TARGET:[86.048, 0.239]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.36, -2.68, -0.17]][[2.72, 2.46, 2.13]][[-86.05, -86.05]][[-86.05, -0.5, -9.01]]
std:[[0.52, 0.45, 0.31]][[0.49, 0.46, 0.44]][[0.0, 0.0]][[0.0, 0.34, 0.2]]
MSE:[[2.42, 2.72, 0.35]][[2.76, 2.5, 2.17]][[86.05, 86.05]][[86.05, 0.6, 9.01]]
MSE(-DR):[[0.0, 0.3, -2.07]][[0.34, 0.08, -0.25]][[83.63, 83.63]][[83.63, -1.82, 6.59]]
==============


O_threshold = 90
MC for this TARGET:[84.263, 0.216]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.76, -2.07, -0.14]][[3.52, 3.27, 2.84]][[-84.26, -84.26]][[-84.26, -0.44, -7.22]]
std:[[0.38, 0.29, 0.27]][[0.29, 0.28, 0.27]][[0.0, 0.0]][[0.0, 0.31, 0.2]]
MSE:[[1.8, 2.09, 0.3]][[3.53, 3.28, 2.85]][[84.26, 84.26]][[84.26, 0.54, 7.22]]
MSE(-DR):[[0.0, 0.29, -1.5]][[1.73, 1.48, 1.05]][[82.46, 82.46]][[82.46, -1.26, 5.42]]
**
MC-based ATE = -1.78
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.59, 0.61, 0.03]][[0.8, 0.81, 0.71]][[1.78, 1.78]][0.05]
std:[[0.21, 0.24, 0.06]][[0.23, 0.21, 0.18]][[0.0, 0.0]][0.03]
MSE:[[0.63, 0.66, 0.07]][[0.83, 0.84, 0.73]][[1.78, 1.78]][0.06]
MSE(-DR):[[0.0, 0.03, -0.56]][[0.2, 0.21, 0.1]][[1.15, 1.15]][-0.57]
**
==============


O_threshold = 100
MC for this TARGET:[87.926, 0.209]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.81, -3.13, -3.25]][[0.62, 0.32, -0.41]][[-87.93, -87.93]][[-87.93, -3.57, -10.89]]
std:[[0.53, 0.49, 0.38]][[0.38, 0.35, 0.31]][[0.0, 0.0]][[0.0, 0.27, 0.2]]
MSE:[[2.86, 3.17, 3.27]][[0.73, 0.47, 0.51]][[87.93, 87.93]][[87.93, 3.58, 10.89]]
MSE(-DR):[[0.0, 0.31, 0.41]][[-2.13, -2.39, -2.35]][[85.07, 85.07]][[85.07, 0.72, 8.03]]
MC-based ATE = 1.88
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.46, -0.45, -3.08]][[-2.1, -2.14, -2.54]][[-1.88, -1.88]][-3.07]
std:[[0.66, 0.7, 0.68]][[0.18, 0.17, 0.14]][[0.0, 0.0]][0.6]
MSE:[[0.8, 0.83, 3.15]][[2.11, 2.15, 2.54]][[1.88, 1.88]][3.13]
MSE(-DR):[[0.0, 0.03, 2.35]][[1.31, 1.35, 1.74]][[1.08, 1.08]][2.33]
**
==============


O_threshold = 110
MC for this TARGET:[85.587, 0.211]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.11, -3.4, -3.43]][[-2.6, -2.85, -3.64]][[-85.59, -85.59]][[-85.59, -3.72, -8.55]]
std:[[0.56, 0.5, 0.34]][[0.27, 0.24, 0.3]][[0.0, 0.0]][[0.0, 0.26, 0.2]]
MSE:[[3.16, 3.44, 3.45]][[2.61, 2.86, 3.65]][[85.59, 85.59]][[85.59, 3.73, 8.55]]
```

```
MSE(-DR):[[0.0, 0.28, 0.29]][[-0.55, -0.3, 0.49]][[82.43, 82.43]][[82.43, 0.57, 5.39]]
MC-based ATE = -0.46
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.75, -0.71, -3.26]][[-5.32, -5.31, -5.78]][[0.46, 0.46]][-3.23]
std:[[0.15, 0.19, 0.36]][[0.45, 0.46, 0.25]][[0.0, 0.0]][0.37]
MSE:[[0.76, 0.73, 3.28]][[5.34, 5.33, 5.79]][[0.46, 0.46]][3.25]
MSE(-DR):[[0.0, -0.03, 2.52]][[4.58, 4.57, 5.03]][[-0.3, -0.3]][2.49]
**
==============


O_threshold = 120
MC for this TARGET:[86.747, 0.215]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-7.79, -8.01, -7.25]][[-6.85, -7.08, -7.97]][[-86.75, -86.75]][[-86.75, -7.47, -9.71]]
std:[[1.15, 1.1, 0.85]][[0.26, 0.24, 0.24]][[0.0, 0.0]][[0.0, 0.78, 0.2]]
MSE:[[7.87, 8.09, 7.3]][[6.85, 7.08, 7.97]][[86.75, 86.75]][[86.75, 7.51, 9.71]]
MSE(-DR):[[0.0, 0.22, -0.57]][[-1.02, -0.79, 0.1]][[78.88, 78.88]][[78.88, -0.36, 1.84]]
MC-based ATE = 0.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-5.44, -5.33, -7.09]][[-9.58, -9.54, -10.1]][[-0.7, -0.7]][-6.98]
std:[[0.99, 0.99, 1.14]][[0.5, 0.47, 0.35]][[0.0, 0.0]][1.12]
MSE:[[5.53, 5.42, 7.18]][[9.59, 9.55, 10.11]][[0.7, 0.7]][7.07]
MSE(-DR):[[0.0, -0.11, 1.65]][[4.06, 4.02, 4.58]][[-4.83, -4.83]][1.54]
**
==============


O_threshold = 130
MC for this TARGET:[89.34, 0.223]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-10.79, -10.96, -9.26]][[-11.04, -11.25, -12.28]][[-89.34, -89.34]][[-89.34, -9.44, -12.3]]
std:[[0.96, 0.93, 0.68]][[0.56, 0.52, 0.4]][[0.0, 0.0]][[0.0, 0.68, 0.2]]
MSE:[[10.83, 11.0, 9.28]][[11.05, 11.26, 12.29]][[89.34, 89.34]][[89.34, 9.46, 12.3]]
MSE(-DR):[[0.0, 0.17, -1.55]][[0.22, 0.43, 1.46]][[78.51, 78.51]][[78.51, -1.37, 1.47]]
**
MC-based ATE = 3.29
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-8.44, -8.28, -9.09]][[-13.77, -13.71, -14.41]][[-3.29, -3.29]][-8.94]
std:[[0.56, 0.6, 0.97]][[0.7, 0.69, 0.46]][[0.0, 0.0]][1.01]
MSE:[[8.46, 8.3, 9.14]][[13.79, 13.73, 14.42]][[3.29, 3.29]][9.0]
MSE(-DR):[[0.0, -0.16, 0.68]][[5.33, 5.27, 5.96]][[-5.17, -5.17]][0.54]
**
==============


time spent until now: 11.7 mins


---------------------------------------
[pattern_seed, T, sd_R] = [0, 336, 10]

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

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1
```

```
number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 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:76.562
O_threshold = 80
```

```
MC for this TARGET:[86.044, 0.142]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.45, -2.73, -0.18]][[2.01, 1.75, 1.29]][[-86.04, -86.04]][[-86.04, -0.47, -9.48]]
std:[[0.17, 0.21, 0.19]][[0.27, 0.3, 0.25]][[0.0, 0.0]][[0.0, 0.15, 0.2]]
MSE:[[2.46, 2.74, 0.26]][[2.03, 1.78, 1.31]][[86.04, 86.04]][[86.04, 0.49, 9.48]]
MSE(-DR):[[0.0, 0.28, -2.2]][[-0.43, -0.68, -1.15]][[83.58, 83.58]][[83.58, -1.97, 7.02]]
==============


O_threshold = 90
MC for this TARGET:[84.276, 0.139]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.78, -2.04, -0.34]][[2.76, 2.49, 1.97]][[-84.28, -84.28]][[-84.28, -0.6, -7.71]]
std:[[0.36, 0.34, 0.1]][[0.3, 0.33, 0.26]][[0.0, 0.0]][[0.0, 0.11, 0.2]]
MSE:[[1.82, 2.07, 0.35]][[2.78, 2.51, 1.99]][[84.28, 84.28]][[84.28, 0.61, 7.71]]
MSE(-DR):[[0.0, 0.25, -1.47]][[0.96, 0.69, 0.17]][[82.46, 82.46]][[82.46, -1.21, 5.89]]
**
MC-based ATE = -1.77
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.66, 0.68, -0.16]][[0.75, 0.74, 0.68]][[1.77, 1.77]][-0.14]
std:[[0.21, 0.15, 0.1]][[0.04, 0.06, 0.05]][[0.0, 0.0]][0.05]
MSE:[[0.69, 0.7, 0.19]][[0.75, 0.74, 0.68]][[1.77, 1.77]][0.15]
MSE(-DR):[[0.0, 0.01, -0.5]][[0.06, 0.05, -0.01]][[1.08, 1.08]][-0.54]
==============


O_threshold = 100
MC for this TARGET:[87.954, 0.138]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.7, -3.0, -3.39]][[-0.1, -0.43, -1.24]][[-87.95, -87.95]][[-87.95, -3.69, -11.39]]
std:[[0.51, 0.52, 0.11]][[0.35, 0.37, 0.32]][[0.0, 0.0]][[0.0, 0.11, 0.2]]
MSE:[[2.75, 3.04, 3.39]][[0.36, 0.57, 1.28]][[87.95, 87.95]][[87.95, 3.69, 11.39]]
MSE(-DR):[[0.0, 0.29, 0.64]][[-2.39, -2.18, -1.47]][[85.2, 85.2]][[85.2, 0.94, 8.64]]
MC-based ATE = 1.91
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.25, -0.27, -3.21]][[-2.11, -2.18, -2.54]][[-1.91, -1.91]][-3.23]
std:[[0.53, 0.52, 0.13]][[0.12, 0.11, 0.07]][[0.0, 0.0]][0.09]
MSE:[[0.59, 0.59, 3.21]][[2.11, 2.18, 2.54]][[1.91, 1.91]][3.23]
MSE(-DR):[[0.0, 0.0, 2.62]][[1.52, 1.59, 1.95]][[1.32, 1.32]][2.64]
**
==============


O_threshold = 110
MC for this TARGET:[85.608, 0.14]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.3, -3.58, -3.64]][[-3.0, -3.27, -4.19]][[-85.61, -85.61]][[-85.61, -3.93, -9.05]]
std:[[0.42, 0.38, 0.39]][[0.29, 0.3, 0.28]][[0.0, 0.0]][[0.0, 0.33, 0.2]]
MSE:[[3.33, 3.6, 3.66]][[3.01, 3.28, 4.2]][[85.61, 85.61]][[85.61, 3.94, 9.05]]
MSE(-DR):[[0.0, 0.27, 0.33]][[-0.32, -0.05, 0.87]][[82.28, 82.28]][[82.28, 0.61, 5.72]]
MC-based ATE = -0.44
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.85, -0.86, -3.46]][[-5.01, -5.02, -5.48]][[0.44, 0.44]][-3.46]
std:[[0.56, 0.55, 0.42]][[0.02, 0.0, 0.1]][[0.0, 0.0]][0.33]
MSE:[[1.02, 1.02, 3.49]][[5.01, 5.02, 5.48]][[0.44, 0.44]][3.48]
MSE(-DR):[[0.0, 0.0, 2.47]][[3.99, 4.0, 4.46]][[-0.58, -0.58]][2.46]
**
==============


O_threshold = 120
MC for this TARGET:[86.751, 0.137]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-7.69, -7.91, -7.41]][[-7.2, -7.4, -8.37]][[-86.75, -86.75]][[-86.75, -7.62, -10.19]]
std:[[0.8, 0.75, 0.59]][[0.31, 0.31, 0.28]][[0.0, 0.0]][[0.0, 0.49, 0.2]]
MSE:[[7.73, 7.95, 7.43]][[7.21, 7.41, 8.37]][[86.75, 86.75]][[86.75, 7.64, 10.19]]
MSE(-DR):[[0.0, 0.22, -0.3]][[-0.52, -0.32, 0.64]][[79.02, 79.02]][[79.02, -0.09, 2.46]]
MC-based ATE = 0.71
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-5.25, -5.18, -7.22]][[-9.21, -9.15, -9.66]][[-0.71, -0.71]][-7.16]
std:[[0.95, 0.94, 0.52]][[0.05, 0.03, 0.07]][[0.0, 0.0]][0.39]
MSE:[[5.34, 5.26, 7.24]][[9.21, 9.15, 9.66]][[0.71, 0.71]][7.17]
MSE(-DR):[[0.0, -0.08, 1.9]][[3.87, 3.81, 4.32]][[-4.63, -4.63]][1.83]
**
==============


O_threshold = 130
```

```
MC for this TARGET:[89.341, 0.139]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-11.61, -11.86, -9.97]][[-11.52, -11.69, -12.63]][[-89.34, -89.34]][[-89.34, -10.22, -12.78]]
std:[[0.68, 0.61, 0.6]][[0.2, 0.2, 0.26]][[0.0, 0.0]][[0.0, 0.52, 0.2]]
MSE:[[11.63, 11.88, 9.99]][[11.52, 11.69, 12.63]][[89.34, 89.34]][[89.34, 10.23, 12.78]]
MSE(-DR):[[0.0, 0.25, -1.64]][[-0.11, 0.06, 1.0]][[77.71, 77.71]][[77.71, -1.4, 1.15]]
MC-based ATE = 3.3
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-9.16, -9.13, -9.78]][[-13.53, -13.44, -13.92]][[-3.3, -3.3]][-9.75]
std:[[0.84, 0.8, 0.44]][[0.07, 0.1, 0.08]][[0.0, 0.0]][0.37]
MSE:[[9.2, 9.16, 9.79]][[13.53, 13.44, 13.92]][[3.3, 3.3]][9.76]
MSE(-DR):[[0.0, -0.04, 0.59]][[4.33, 4.24, 4.72]][[-5.9, -5.9]][0.56]
**
==============


time spent until now: 23.4 mins


--------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 10]

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

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
```

```
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 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;
```