```
Last login: Wed Apr  1 13:19:12 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-18-204-44-50.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 36.0

 * 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.


*** System restart required ***
Last login: Wed Apr  1 17:19:16 2020 from 107.13.161.147
^[[Aubuntu@ip-172-31-9-82:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
13:43, 04/01; num of cores:36

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A,  simple, M_in_R, u_O_u_D, mean_reversion, poisO] = [672, 10, 10, None, 0.3, 0.5, 1
, False, True, 10, False, True]


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

max(u_O) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 105
target policy:

0 0 0 1 0

0 1 0 0 0
```

```
1 1 0 0 1

0 0 1 0 0

0 0 1 0 0

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

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 135
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

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

--------------------------------------
Value of Behaviour policy:60.758
O_threshold = 80
MC for this TARGET:[70.898, 0.05]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.78, 0.65, -1.32]][[1.09, -70.9, -10.14]]
std:[[0.07, 0.05, 0.14]][[0.04, 0.0, 0.03]]
MSE:[[0.78, 0.65, 1.33]][[1.09, 70.9, 10.14]]
MSE(-DR):[[0.0, -0.13, 0.55]][[0.31, 70.12, 9.36]]
***
==============


O_threshold = 90
MC for this TARGET:[69.38, 0.056]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.33, 0.23, -1.17]][[-0.73, -69.38, -8.62]]
std:[[0.28, 0.25, 0.1]][[0.02, 0.0, 0.03]]
MSE:[[0.43, 0.34, 1.17]][[0.73, 69.38, 8.62]]
MSE(-DR):[[0.0, -0.09, 0.74]][[0.3, 68.95, 8.19]]
***
==============


O_threshold = 105
MC for this TARGET:[71.388, 0.056]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.21, -6.3, -6.8]][[-8.16, -71.39, -10.63]]
std:[[0.04, 0.05, 0.24]][[0.02, 0.0, 0.03]]
MSE:[[6.21, 6.3, 6.8]][[8.16, 71.39, 10.63]]
MSE(-DR):[[0.0, 0.09, 0.59]][[1.95, 65.18, 4.42]]
***
==============


O_threshold = 120
MC for this TARGET:[70.557, 0.05]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.06, -9.08, -8.42]][[-13.57, -70.56, -9.8]]
std:[[0.29, 0.28, 0.08]][[0.02, 0.0, 0.03]]
MSE:[[9.06, 9.08, 8.42]][[13.57, 70.56, 9.8]]
MSE(-DR):[[0.0, 0.02, -0.64]][[4.51, 61.5, 0.74]]
**
==============


O_threshold = 135
MC for this TARGET:[70.557, 0.05]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.07, -9.08, -8.4]][[-13.59, -70.56, -9.8]]
std:[[0.3, 0.28, 0.08]][[0.02, 0.0, 0.03]]
```

```
MSE:[[9.07, 9.08, 8.4]][[13.59, 70.56, 9.8]]
MSE(-DR):[[0.0, 0.01, -0.67]][[4.52, 61.49, 0.73]]
**
==============


[[ 0.78  0.65  1.33  1.09 70.9  10.14]
 [ 0.43  0.34  1.17  0.73 69.38  8.62]
 [ 6.21  6.3   6.8   8.16 71.39 10.63]
 [ 9.06  9.08  8.42 13.57 70.56  9.8 ]
 [ 9.07  9.08  8.4  13.59 70.56  9.8 ]]


time spent until now: 3.2 mins


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

max(u_O) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 105
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 1 0 0

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

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 135
```

```
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

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

--------------------------------------
Value of Behaviour policy:60.786
O_threshold = 80
MC for this TARGET:[70.89, 0.157]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.76, 0.69, -1.24]][[1.26, -70.89, -10.1]]
std:[[0.11, 0.11, 0.1]][[0.07, 0.0, 0.03]]
MSE:[[0.77, 0.7, 1.24]][[1.26, 70.89, 10.1]]
MSE(-DR):[[0.0, -0.07, 0.47]][[0.49, 70.12, 9.33]]
***
==============


O_threshold = 90
MC for this TARGET:[69.373, 0.161]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.17, 0.08, -1.3]][[-0.57, -69.37, -8.59]]
std:[[0.12, 0.13, 0.06]][[0.0, 0.0, 0.03]]
MSE:[[0.21, 0.15, 1.3]][[0.57, 69.37, 8.59]]
MSE(-DR):[[0.0, -0.06, 1.09]][[0.36, 69.16, 8.38]]
***
==============


O_threshold = 105
MC for this TARGET:[71.38, 0.149]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.24, -6.34, -7.04]][[-7.85, -71.38, -10.59]]
std:[[0.51, 0.5, 0.3]][[0.23, 0.0, 0.03]]
MSE:[[6.26, 6.36, 7.05]][[7.85, 71.38, 10.59]]
MSE(-DR):[[0.0, 0.1, 0.79]][[1.59, 65.12, 4.33]]
***
==============


O_threshold = 120
MC for this TARGET:[70.549, 0.15]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.03, -9.0, -8.4]][[-13.45, -70.55, -9.76]]
std:[[0.35, 0.4, 0.08]][[0.04, 0.0, 0.03]]
MSE:[[9.04, 9.01, 8.4]][[13.45, 70.55, 9.76]]
MSE(-DR):[[0.0, -0.03, -0.64]][[4.41, 61.51, 0.72]]
**
==============


O_threshold = 135
MC for this TARGET:[70.549, 0.15]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.0, -9.0, -8.32]][[-13.44, -70.55, -9.76]]
std:[[0.37, 0.4, 0.13]][[0.07, 0.0, 0.03]]
MSE:[[9.01, 9.01, 8.32]][[13.44, 70.55, 9.76]]
MSE(-DR):[[0.0, 0.0, -0.69]][[4.43, 61.54, 0.75]]
**
==============


[[ 0.78  0.65  1.33  1.09 70.9  10.14]
 [ 0.43  0.34  1.17  0.73 69.38  8.62]
 [ 6.21  6.3   6.8   8.16 71.39 10.63]
 [ 9.06  9.08  8.42 13.57 70.56  9.8 ]
 [ 9.07  9.08  8.4  13.59 70.56  9.8 ]]


[[ 0.77  0.7   1.24  1.26 70.89 10.1 ]
 [ 0.21  0.15  1.3   0.57 69.37  8.59]
 [ 6.26  6.36  7.05  7.85 71.38 10.59]
 [ 9.04  9.01  8.4  13.45 70.55  9.76]
 [ 9.01  9.01  8.32 13.44 70.55  9.76]]


time spent until now: 6.5 mins
```

```
--------------------------------------
[pattern_seed, sd_R] = [2, 100]

max(u_O) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 105
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 1 0 0

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

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 135
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

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

```
--------------------------------------
Value of Behaviour policy:60.903
O_threshold = 80
MC for this TARGET:[70.86, 0.725]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.93, 0.83, -0.81]][[2.04, -70.86, -9.96]]
std:[[0.3, 0.39, 0.19]][[0.19, 0.0, 0.01]]
MSE:[[0.98, 0.92, 0.83]][[2.05, 70.86, 9.96]]
MSE(-DR):[[0.0, -0.06, -0.15]][[1.07, 69.88, 8.98]]
**
==============


O_threshold = 90
MC for this TARGET:[69.342, 0.728]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.6, -0.54, -1.99]][[-0.07, -69.34, -8.44]]
std:[[0.49, 0.35, 0.42]][[0.17, 0.0, 0.01]]
MSE:[[0.77, 0.64, 2.03]][[0.18, 69.34, 8.44]]
MSE(-DR):[[0.0, -0.13, 1.26]][[-0.59, 68.57, 7.67]]
==============


O_threshold = 105
MC for this TARGET:[71.35, 0.715]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.38, -6.53, -7.69]][[-6.64, -71.35, -10.45]]
std:[[2.29, 2.37, 0.52]][[1.05, 0.0, 0.01]]
MSE:[[6.78, 6.95, 7.71]][[6.72, 71.35, 10.45]]
MSE(-DR):[[0.0, 0.17, 0.93]][[-0.06, 64.57, 3.67]]
==============


O_threshold = 120
MC for this TARGET:[70.519, 0.718]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.76, -8.71, -8.15]][[-12.82, -70.52, -9.62]]
std:[[3.2, 3.2, 0.66]][[0.35, 0.0, 0.01]]
MSE:[[9.33, 9.28, 8.18]][[12.82, 70.52, 9.62]]
MSE(-DR):[[0.0, -0.05, -1.15]][[3.49, 61.19, 0.29]]
**
==============


O_threshold = 135
MC for this TARGET:[70.519, 0.718]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.79, -8.71, -8.07]][[-12.8, -70.52, -9.62]]
std:[[3.07, 3.2, 0.57]][[0.37, 0.0, 0.01]]
MSE:[[9.31, 9.28, 8.09]][[12.81, 70.52, 9.62]]
MSE(-DR):[[0.0, -0.03, -1.22]][[3.5, 61.21, 0.31]]
**
==============


[[ 0.78  0.65  1.33  1.09 70.9  10.14]
 [ 0.43  0.34  1.17  0.73 69.38  8.62]
 [ 6.21  6.3   6.8   8.16 71.39 10.63]
 [ 9.06  9.08  8.42 13.57 70.56  9.8 ]
 [ 9.07  9.08  8.4  13.59 70.56  9.8 ]]


[[ 0.77  0.7   1.24  1.26 70.89 10.1 ]
 [ 0.21  0.15  1.3   0.57 69.37  8.59]
 [ 6.26  6.36  7.05  7.85 71.38 10.59]
 [ 9.04  9.01  8.4  13.45 70.55  9.76]
 [ 9.01  9.01  8.32 13.44 70.55  9.76]]


[[ 0.98  0.92  0.83  2.05 70.86  9.96]
 [ 0.77  0.64  2.03  0.18 69.34  8.44]
 [ 6.78  6.95  7.71  6.72 71.35 10.45]
 [ 9.33  9.28  8.18 12.82 70.52  9.62]
 [ 9.31  9.28  8.09 12.81 70.52  9.62]]


time spent until now: 9.7 mins


--------------------------------------
[pattern_seed, sd_R] = [3, 0.5]

max(u_O) =  170.1
O_threshold = 80
means of Order:

170.1 113.4 102.4 56.9 91.5
```

```
89.4 97.0 82.4 98.2 86.2

67.1 129.7 129.6 166.1 101.0

88.1 84.5 62.6 133.6 71.5

69.7 93.5 155.4 106.8 73.2

target policy:

1 1 1 0 1

1 1 1 1 1

0 1 1 1 1

1 1 0 1 0

0 1 1 1 0

number of reward locations:  19
O_threshold = 90
target policy:

1 1 1 0 1

0 1 0 1 0

0 1 1 1 1

0 0 0 1 0

0 1 1 1 0

number of reward locations:  14
O_threshold = 105
target policy:

1 1 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 1 0

number of reward locations:  8
O_threshold = 120
target policy:

1 0 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 0 0

number of reward locations:  6
O_threshold = 135
target policy:

1 0 0 0 0

0 0 0 0 0

0 0 0 1 0

0 0 0 0 0

0 0 1 0 0

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

--------------------------------------
Value of Behaviour policy:63.696
O_threshold = 80
MC for this TARGET:[73.338, 0.052]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.08, 1.01, -1.32]][[2.93, -73.34, -9.64]]
std:[[0.4, 0.38, 0.04]][[0.06, 0.0, 0.07]]
MSE:[[1.15, 1.08, 1.32]][[2.93, 73.34, 9.64]]
MSE(-DR):[[0.0, -0.07, 0.17]][[1.78, 72.19, 8.49]]
```

```
***
==============


O_threshold = 90
MC for this TARGET:[73.443, 0.051]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.49, -0.61, -3.11]][[-0.02, -73.44, -9.75]]
std:[[0.14, 0.14, 0.08]][[0.01, 0.0, 0.07]]
MSE:[[0.51, 0.63, 3.11]][[0.02, 73.44, 9.75]]
MSE(-DR):[[0.0, 0.12, 2.6]][[-0.49, 72.93, 9.24]]
==============


O_threshold = 105
MC for this TARGET:[71.833, 0.056]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.48, -2.56, -3.54]][[-4.54, -71.83, -8.14]]
std:[[0.07, 0.07, 0.02]][[0.05, 0.0, 0.07]]
MSE:[[2.48, 2.56, 3.54]][[4.54, 71.83, 8.14]]
MSE(-DR):[[0.0, 0.08, 1.06]][[2.06, 69.35, 5.66]]
***
==============


O_threshold = 120
MC for this TARGET:[69.164, 0.052]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.48, -2.53, -2.47]][[-5.15, -69.16, -5.47]]
std:[[0.12, 0.13, 0.06]][[0.03, 0.0, 0.07]]
MSE:[[2.48, 2.53, 2.47]][[5.15, 69.16, 5.47]]
MSE(-DR):[[0.0, 0.05, -0.01]][[2.67, 66.68, 2.99]]
**
==============


O_threshold = 135
MC for this TARGET:[76.028, 0.055]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-12.36, -12.36, -11.68]][[-17.5, -76.03, -12.33]]
std:[[0.14, 0.14, 0.16]][[0.1, 0.0, 0.07]]
MSE:[[12.36, 12.36, 11.68]][[17.5, 76.03, 12.33]]
MSE(-DR):[[0.0, 0.0, -0.68]][[5.14, 63.67, -0.03]]
**
==============


[[ 0.78  0.65  1.33  1.09 70.9  10.14]
 [ 0.43  0.34  1.17  0.73 69.38  8.62]
 [ 6.21  6.3   6.8   8.16 71.39 10.63]
 [ 9.06  9.08  8.42 13.57 70.56  9.8 ]
 [ 9.07  9.08  8.4  13.59 70.56  9.8 ]]


[[ 0.77  0.7   1.24  1.26 70.89 10.1 ]
 [ 0.21  0.15  1.3   0.57 69.37  8.59]
 [ 6.26  6.36  7.05  7.85 71.38 10.59]
 [ 9.04  9.01  8.4  13.45 70.55  9.76]
 [ 9.01  9.01  8.32 13.44 70.55  9.76]]


[[ 0.98  0.92  0.83  2.05 70.86  9.96]
 [ 0.77  0.64  2.03  0.18 69.34  8.44]
 [ 6.78  6.95  7.71  6.72 71.35 10.45]
 [ 9.33  9.28  8.18 12.82 70.52  9.62]
 [ 9.31  9.28  8.09 12.81 70.52  9.62]]


[[1.150e+00 1.080e+00 1.320e+00 2.930e+00 7.334e+01 9.640e+00]
 [5.100e-01 6.300e-01 3.110e+00 2.000e-02 7.344e+01 9.750e+00]
 [2.480e+00 2.560e+00 3.540e+00 4.540e+00 7.183e+01 8.140e+00]
 [2.480e+00 2.530e+00 2.470e+00 5.150e+00 6.916e+01 5.470e+00]
 [1.236e+01 1.236e+01 1.168e+01 1.750e+01 7.603e+01 1.233e+01]]


time spent until now: 12.9 mins


--------------------------------------
[pattern_seed, sd_R] = [3, 20]

max(u_O) =  170.1
O_threshold = 80
means of Order:

170.1 113.4 102.4 56.9 91.5

89.4 97.0 82.4 98.2 86.2
```

```
67.1 129.7 129.6 166.1 101.0

88.1 84.5 62.6 133.6 71.5

69.7 93.5 155.4 106.8 73.2

target policy:

1 1 1 0 1

1 1 1 1 1

0 1 1 1 1

1 1 0 1 0

0 1 1 1 0

number of reward locations:  19
O_threshold = 90
target policy:

1 1 1 0 1

0 1 0 1 0

0 1 1 1 1

0 0 0 1 0

0 1 1 1 0

number of reward locations:  14
O_threshold = 105
target policy:

1 1 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 1 0

number of reward locations:  8
O_threshold = 120
target policy:

1 0 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 0 0

number of reward locations:  6
O_threshold = 135
target policy:

1 0 0 0 0

0 0 0 0 0

0 0 0 1 0

0 0 0 0 0

0 0 1 0 0

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

--------------------------------------
Value of Behaviour policy:63.724
O_threshold = 80
MC for this TARGET:[73.33, 0.15]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.43, 1.39, -1.04]][[3.03, -73.33, -9.61]]
std:[[1.04, 1.06, 0.02]][[0.01, 0.0, 0.08]]
MSE:[[1.77, 1.75, 1.04]][[3.03, 73.33, 9.61]]
MSE(-DR):[[0.0, -0.02, -0.73]][[1.26, 71.56, 7.84]]
**
==============
O_threshold = 90
MC for this TARGET:[73.436, 0.151]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.75, -0.82, -3.24]][[0.16, -73.44, -9.71]]
std:[[0.25, 0.27, 0.15]][[0.2, 0.0, 0.08]]
MSE:[[0.79, 0.86, 3.24]][[0.26, 73.44, 9.71]]
MSE(-DR):[[0.0, 0.07, 2.45]][[-0.53, 72.65, 8.92]]
==============


O_threshold = 105
MC for this TARGET:[71.826, 0.148]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.74, -1.83, -3.12]][[-4.48, -71.83, -8.1]]
std:[[0.6, 0.61, 0.53]][[0.05, 0.0, 0.08]]
MSE:[[1.84, 1.93, 3.16]][[4.48, 71.83, 8.1]]
MSE(-DR):[[0.0, 0.09, 1.32]][[2.64, 69.99, 6.26]]
***
==============


O_threshold = 120
MC for this TARGET:[69.157, 0.15]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.19, -2.29, -1.88]][[-5.17, -69.16, -5.43]]
std:[[0.08, 0.08, 0.23]][[0.21, 0.0, 0.08]]
MSE:[[2.19, 2.29, 1.89]][[5.17, 69.16, 5.43]]
MSE(-DR):[[0.0, 0.1, -0.3]][[2.98, 66.97, 3.24]]
**
==============


O_threshold = 135
MC for this TARGET:[76.021, 0.155]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-11.71, -11.76, -11.49]][[-17.61, -76.02, -12.3]]
std:[[0.49, 0.45, 0.44]][[0.33, 0.0, 0.08]]
MSE:[[11.72, 11.77, 11.5]][[17.61, 76.02, 12.3]]
MSE(-DR):[[0.0, 0.05, -0.22]][[5.89, 64.3, 0.58]]
**
==============


[[ 0.78  0.65  1.33  1.09 70.9  10.14]
 [ 0.43  0.34  1.17  0.73 69.38  8.62]
 [ 6.21  6.3   6.8   8.16 71.39 10.63]
 [ 9.06  9.08  8.42 13.57 70.56  9.8 ]
 [ 9.07  9.08  8.4  13.59 70.56  9.8 ]]


[[ 0.77  0.7   1.24  1.26 70.89 10.1 ]
 [ 0.21  0.15  1.3   0.57 69.37  8.59]
 [ 6.26  6.36  7.05  7.85 71.38 10.59]
 [ 9.04  9.01  8.4  13.45 70.55  9.76]
 [ 9.01  9.01  8.32 13.44 70.55  9.76]]


[[ 0.98  0.92  0.83  2.05 70.86  9.96]
 [ 0.77  0.64  2.03  0.18 69.34  8.44]
 [ 6.78  6.95  7.71  6.72 71.35 10.45]
 [ 9.33  9.28  8.18 12.82 70.52  9.62]
 [ 9.31  9.28  8.09 12.81 70.52  9.62]]


[[1.150e+00 1.080e+00 1.320e+00 2.930e+00 7.334e+01 9.640e+00]
 [5.100e-01 6.300e-01 3.110e+00 2.000e-02 7.344e+01 9.750e+00]
 [2.480e+00 2.560e+00 3.540e+00 4.540e+00 7.183e+01 8.140e+00]
 [2.480e+00 2.530e+00 2.470e+00 5.150e+00 6.916e+01 5.470e+00]
 [1.236e+01 1.236e+01 1.168e+01 1.750e+01 7.603e+01 1.233e+01]]


[[ 1.77  1.75  1.04  3.03 73.33  9.61]
 [ 0.79  0.86  3.24  0.26 73.44  9.71]
 [ 1.84  1.93  3.16  4.48 71.83  8.1 ]
 [ 2.19  2.29  1.89  5.17 69.16  5.43]
 [11.72 11.77 11.5  17.61 76.02 12.3 ]]


time spent until now: 16.1 mins


--------------------------------------
[pattern_seed, sd_R] = [3, 100]

max(u_O) =  170.1
O_threshold = 80
means of Order:
```

```
170.1 113.4 102.4 56.9 91.5

89.4 97.0 82.4 98.2 86.2

67.1 129.7 129.6 166.1 101.0

88.1 84.5 62.6 133.6 71.5

69.7 93.5 155.4 106.8 73.2

target policy:

1 1 1 0 1

1 1 1 1 1

0 1 1 1 1

1 1 0 1 0

0 1 1 1 0

number of reward locations:  19
O_threshold = 90
target policy:

1 1 1 0 1

0 1 0 1 0

0 1 1 1 1

0 0 0 1 0

0 1 1 1 0

number of reward locations:  14
O_threshold = 105
target policy:

1 1 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 1 0

number of reward locations:  8
O_threshold = 120
target policy:

1 0 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 0 0

number of reward locations:  6
O_threshold = 135
target policy:

1 0 0 0 0

0 0 0 0 0

0 0 0 1 0

0 0 0 0 0

0 0 1 0 0

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

-------------------------------------
Value of Behaviour policy:63.842
O_threshold = 80
MC for this TARGET:[73.3, 0.717]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[3.02, 2.96, 0.26]][[3.58, -73.3, -9.46]]
```

```
std:[[3.87, 3.86, 0.15]][[0.19, 0.0, 0.09]]
MSE:[[4.91, 4.86, 0.3]][[3.59, 73.3, 9.46]]
MSE(-DR):[[0.0, -0.05, -4.61]][[-1.32, 68.39, 4.55]]
==============


O_threshold = 90
MC for this TARGET:[73.405, 0.718]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.71, -1.68, -3.42]][[1.03, -73.4, -9.56]]
std:[[0.75, 0.78, 0.5]][[0.9, 0.0, 0.09]]
MSE:[[1.87, 1.85, 3.46]][[1.37, 73.4, 9.56]]
MSE(-DR):[[0.0, -0.02, 1.59]][[-0.5, 71.53, 7.69]]
==============


O_threshold = 105
MC for this TARGET:[71.795, 0.714]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.26, 1.16, -1.29]][[-4.25, -71.8, -7.95]]
std:[[3.38, 3.43, 2.79]][[0.2, 0.0, 0.09]]
MSE:[[3.61, 3.62, 3.07]][[4.25, 71.8, 7.95]]
MSE(-DR):[[0.0, 0.01, -0.54]][[0.64, 68.19, 4.34]]
**
==============


O_threshold = 120
MC for this TARGET:[69.126, 0.717]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.09, -1.32, 0.19]][[-5.36, -69.13, -5.28]]
std:[[0.93, 0.96, 1.47]][[0.89, 0.0, 0.09]]
MSE:[[1.43, 1.63, 1.48]][[5.43, 69.13, 5.28]]
MSE(-DR):[[0.0, 0.2, 0.05]][[4.0, 67.7, 3.85]]
***
==============


O_threshold = 135
MC for this TARGET:[75.99, 0.721]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.11, -9.31, -10.69]][[-18.01, -75.99, -12.15]]
std:[[2.75, 2.91, 2.56]][[1.19, 0.0, 0.09]]
MSE:[[9.52, 9.75, 10.99]][[18.05, 75.99, 12.15]]
MSE(-DR):[[0.0, 0.23, 1.47]][[8.53, 66.47, 2.63]]
***
==============


[[ 0.78  0.65  1.33  1.09 70.9  10.14]
 [ 0.43  0.34  1.17  0.73 69.38  8.62]
 [ 6.21  6.3   6.8   8.16 71.39 10.63]
 [ 9.06  9.08  8.42 13.57 70.56  9.8 ]
 [ 9.07  9.08  8.4  13.59 70.56  9.8 ]]


[[ 0.77  0.7   1.24  1.26 70.89 10.1 ]
 [ 0.21  0.15  1.3   0.57 69.37  8.59]
 [ 6.26  6.36  7.05  7.85 71.38 10.59]
 [ 9.04  9.01  8.4  13.45 70.55  9.76]
 [ 9.01  9.01  8.32 13.44 70.55  9.76]]


[[ 0.98  0.92  0.83  2.05 70.86  9.96]
 [ 0.77  0.64  2.03  0.18 69.34  8.44]
 [ 6.78  6.95  7.71  6.72 71.35 10.45]
 [ 9.33  9.28  8.18 12.82 70.52  9.62]
 [ 9.31  9.28  8.09 12.81 70.52  9.62]]


[[1.150e+00 1.080e+00 1.320e+00 2.930e+00 7.334e+01 9.640e+00]
 [5.100e-01 6.300e-01 3.110e+00 2.000e-02 7.344e+01 9.750e+00]
 [2.480e+00 2.560e+00 3.540e+00 4.540e+00 7.183e+01 8.140e+00]
 [2.480e+00 2.530e+00 2.470e+00 5.150e+00 6.916e+01 5.470e+00]
 [1.236e+01 1.236e+01 1.168e+01 1.750e+01 7.603e+01 1.233e+01]]


[[ 1.77  1.75  1.04  3.03 73.33  9.61]
 [ 0.79  0.86  3.24  0.26 73.44  9.71]
 [ 1.84  1.93  3.16  4.48 71.83  8.1 ]
 [ 2.19  2.29  1.89  5.17 69.16  5.43]
 [11.72 11.77 11.5  17.61 76.02 12.3 ]]


[[ 4.91  4.86  0.3   3.59 73.3   9.46]
 [ 1.87  1.85  3.46  1.37 73.4   9.56]
 [ 3.61  3.62  3.07  4.25 71.8   7.95]
 [ 1.43  1.63  1.48  5.43 69.13  5.28]
 [ 9.52  9.75 10.99 18.05 75.99 12.15]]
```

```
time spent until now: 19.4 mins


--------------------------------------
[pattern_seed, sd_R] = [4, 0.5]

max(u_O) =  193.8
O_threshold = 80
means of Order:

101.0 115.6 73.8 122.5 87.8

61.8 81.9 119.1 109.9 70.5

119.8 96.9 113.0 109.9 70.3

110.5 82.9 158.2 123.6 100.9

74.1 101.1 104.4 69.2 193.8

target policy:

1 1 0 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

0 1 1 0 1

number of reward locations:  19
O_threshold = 90
target policy:

1 1 0 1 0

0 0 1 1 0

1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 105
target policy:

0 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 0

0 0 0 0 1

number of reward locations:  11
O_threshold = 120
target policy:

0 0 0 1 0

0 0 0 0 0

0 0 0 0 0

0 0 1 1 0

0 0 0 0 1

number of reward locations:  4
O_threshold = 135
target policy:

0 0 0 0 0

0 0 0 0 0

0 0 0 0 0

0 0 1 0 0

0 0 0 0 1
```

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

--------------------------------------
Value of Behaviour policy:65.176
O_threshold = 80
MC for this TARGET:[72.841, 0.051]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.93, 2.81, 0.44]][[4.2, -72.84, -7.67]]
std:[[0.21, 0.21, 0.23]][[0.28, 0.0, 0.04]]
MSE:[[2.94, 2.82, 0.5]][[4.21, 72.84, 7.67]]
MSE(-DR):[[0.0, -0.12, -2.44]][[1.27, 69.9, 4.73]]
**
==============


O_threshold = 90
MC for this TARGET:[74.177, 0.054]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.58, 1.47, -0.96]][[2.2, -74.18, -9.0]]
std:[[0.23, 0.23, 0.17]][[0.24, 0.0, 0.04]]
MSE:[[1.6, 1.49, 0.97]][[2.21, 74.18, 9.0]]
MSE(-DR):[[0.0, -0.11, -0.63]][[0.61, 72.58, 7.4]]
**
==============


O_threshold = 105
MC for this TARGET:[69.993, 0.05]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.59, 0.52, -0.39]][[0.37, -69.99, -4.82]]
std:[[0.36, 0.36, 0.24]][[0.27, 0.0, 0.04]]
MSE:[[0.69, 0.63, 0.46]][[0.46, 69.99, 4.82]]
MSE(-DR):[[0.0, -0.06, -0.23]][[-0.23, 69.3, 4.13]]
==============


O_threshold = 120
MC for this TARGET:[73.761, 0.048]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.82, -8.84, -8.99]][[-12.19, -73.76, -8.59]]
std:[[0.25, 0.25, 0.1]][[0.1, 0.0, 0.04]]
MSE:[[8.82, 8.84, 8.99]][[12.19, 73.76, 8.59]]
MSE(-DR):[[0.0, 0.02, 0.17]][[3.37, 64.94, -0.23]]
***
==============


O_threshold = 135
MC for this TARGET:[76.678, 0.042]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-13.26, -13.27, -13.11]][[-18.09, -76.68, -11.5]]
std:[[0.34, 0.36, 0.01]][[0.09, 0.0, 0.04]]
MSE:[[13.26, 13.27, 13.11]][[18.09, 76.68, 11.5]]
MSE(-DR):[[0.0, 0.01, -0.15]][[4.83, 63.42, -1.76]]
**
==============


[[ 0.78  0.65  1.33  1.09 70.9  10.14]
 [ 0.43  0.34  1.17  0.73 69.38  8.62]
 [ 6.21  6.3   6.8   8.16 71.39 10.63]
 [ 9.06  9.08  8.42 13.57 70.56  9.8 ]
 [ 9.07  9.08  8.4  13.59 70.56  9.8 ]]


[[ 0.77  0.7   1.24  1.26 70.89 10.1 ]
 [ 0.21  0.15  1.3   0.57 69.37  8.59]
 [ 6.26  6.36  7.05  7.85 71.38 10.59]
 [ 9.04  9.01  8.4  13.45 70.55  9.76]
 [ 9.01  9.01  8.32 13.44 70.55  9.76]]


[[ 0.98  0.92  0.83  2.05 70.86  9.96]
 [ 0.77  0.64  2.03  0.18 69.34  8.44]
 [ 6.78  6.95  7.71  6.72 71.35 10.45]
 [ 9.33  9.28  8.18 12.82 70.52  9.62]
 [ 9.31  9.28  8.09 12.81 70.52  9.62]]


[[1.150e+00 1.080e+00 1.320e+00 2.930e+00 7.334e+01 9.640e+00]
 [5.100e-01 6.300e-01 3.110e+00 2.000e-02 7.344e+01 9.750e+00]
 [2.480e+00 2.560e+00 3.540e+00 4.540e+00 7.183e+01 8.140e+00]
 [2.480e+00 2.530e+00 2.470e+00 5.150e+00 6.916e+01 5.470e+00]
 [1.236e+01 1.236e+01 1.168e+01 1.750e+01 7.603e+01 1.233e+01]]
```

```
[[ 1.77  1.75  1.04  3.03 73.33  9.61]
 [ 0.79  0.86  3.24  0.26 73.44  9.71]
 [ 1.84  1.93  3.16  4.48 71.83  8.1 ]
 [ 2.19  2.29  1.89  5.17 69.16  5.43]
 [11.72 11.77 11.5  17.61 76.02 12.3 ]]


[[ 4.91  4.86  0.3   3.59 73.3   9.46]
 [ 1.87  1.85  3.46  1.37 73.4   9.56]
 [ 3.61  3.62  3.07  4.25 71.8   7.95]
 [ 1.43  1.63  1.48  5.43 69.13  5.28]
 [ 9.52  9.75 10.99 18.05 75.99 12.15]]


[[ 2.94  2.82  0.5   4.21 72.84  7.67]
 [ 1.6   1.49  0.97  2.21 74.18  9.  ]
 [ 0.69  0.63  0.46  0.46 69.99  4.82]
 [ 8.82  8.84  8.99 12.19 73.76  8.59]
 [13.26 13.27 13.11 18.09 76.68 11.5 ]]


time spent until now: 22.6 mins


--------------------------------------
[pattern_seed, sd_R] = [4, 20]

max(u_O) =  193.8
O_threshold = 80
means of Order:

101.0 115.6 73.8 122.5 87.8

61.8 81.9 119.1 109.9 70.5

119.8 96.9 113.0 109.9 70.3

110.5 82.9 158.2 123.6 100.9

74.1 101.1 104.4 69.2 193.8

target policy:

1 1 0 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

0 1 1 0 1

number of reward locations:  19
O_threshold = 90
target policy:

1 1 0 1 0

0 0 1 1 0

1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 105
target policy:

0 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 0

0 0 0 0 1

number of reward locations:  11
O_threshold = 120
target policy:

0 0 0 1 0

0 0 0 0 0
```

```
0 0 0 0 0

0 0 1 1 0

0 0 0 0 1

number of reward locations:  4
O_threshold = 135
target policy:

0 0 0 0 0

0 0 0 0 0

0 0 0 0 0

0 0 1 0 0

0 0 0 0 1

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

--------------------------------------
Value of Behaviour policy:65.204
O_threshold = 80
MC for this TARGET:[72.833, 0.15]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.89, 2.75, 0.21]][[4.22, -72.83, -7.63]]
std:[[0.68, 0.68, 0.46]][[0.01, 0.0, 0.04]]
MSE:[[2.97, 2.83, 0.51]][[4.22, 72.83, 7.63]]
MSE(-DR):[[0.0, -0.14, -2.46]][[1.25, 69.86, 4.66]]
**
==============


O_threshold = 90
MC for this TARGET:[74.17, 0.149]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.88, 0.74, -1.31]][[2.0, -74.17, -8.97]]
std:[[0.03, 0.02, 0.24]][[0.11, 0.0, 0.04]]
MSE:[[0.88, 0.74, 1.33]][[2.0, 74.17, 8.97]]
MSE(-DR):[[0.0, -0.14, 0.45]][[1.12, 73.29, 8.09]]
***
==============


O_threshold = 105
MC for this TARGET:[69.986, 0.15]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.6, 0.54, -0.15]][[0.23, -69.99, -4.78]]
std:[[0.43, 0.45, 0.41]][[0.25, 0.0, 0.04]]
MSE:[[0.74, 0.7, 0.44]][[0.34, 69.99, 4.78]]
MSE(-DR):[[0.0, -0.04, -0.3]][[-0.4, 69.25, 4.04]]
==============


O_threshold = 120
MC for this TARGET:[73.754, 0.152]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.62, -8.62, -8.62]][[-12.22, -73.75, -8.55]]
std:[[0.35, 0.34, 0.21]][[0.13, 0.0, 0.04]]
MSE:[[8.63, 8.63, 8.62]][[12.22, 73.75, 8.55]]
MSE(-DR):[[0.0, 0.0, -0.01]][[3.59, 65.12, -0.08]]
**
==============


O_threshold = 135
MC for this TARGET:[76.671, 0.151]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-13.23, -13.23, -12.89]][[-18.19, -76.67, -11.47]]
std:[[0.1, 0.08, 0.84]][[0.12, 0.0, 0.04]]
MSE:[[13.23, 13.23, 12.92]][[18.19, 76.67, 11.47]]
MSE(-DR):[[0.0, 0.0, -0.31]][[4.96, 63.44, -1.76]]
**
==============


[[ 0.78  0.65  1.33  1.09 70.9  10.14]
 [ 0.43  0.34  1.17  0.73 69.38  8.62]
 [ 6.21  6.3   6.8   8.16 71.39 10.63]
 [ 9.06  9.08  8.42 13.57 70.56  9.8 ]
 [ 9.07  9.08  8.4  13.59 70.56  9.8 ]]


[[ 0.77  0.7   1.24  1.26 70.89 10.1 ]
 [ 0.21  0.15  1.3   0.57 69.37  8.59]
 [ 6.26  6.36  7.05  7.85 71.38 10.59]
 [ 9.04  9.01  8.4  13.45 70.55  9.76]
 [ 9.01  9.01  8.32 13.44 70.55  9.76]]


[[ 0.98  0.92  0.83  2.05 70.86  9.96]
 [ 0.77  0.64  2.03  0.18 69.34  8.44]
 [ 6.78  6.95  7.71  6.72 71.35 10.45]
 [ 9.33  9.28  8.18 12.82 70.52  9.62]
 [ 9.31  9.28  8.09 12.81 70.52  9.62]]


[[1.150e+00 1.080e+00 1.320e+00 2.930e+00 7.334e+01 9.640e+00]
 [5.100e-01 6.300e-01 3.110e+00 2.000e-02 7.344e+01 9.750e+00]
 [2.480e+00 2.560e+00 3.540e+00 4.540e+00 7.183e+01 8.140e+00]
 [2.480e+00 2.530e+00 2.470e+00 5.150e+00 6.916e+01 5.470e+00]
 [1.236e+01 1.236e+01 1.168e+01 1.750e+01 7.603e+01 1.233e+01]]


[[ 1.77  1.75  1.04  3.03 73.33  9.61]
 [ 0.79  0.86  3.24  0.26 73.44  9.71]
 [ 1.84  1.93  3.16  4.48 71.83  8.1 ]
 [ 2.19  2.29  1.89  5.17 69.16  5.43]
 [11.72 11.77 11.5  17.61 76.02 12.3 ]]


[[ 4.91  4.86  0.3   3.59 73.3   9.46]
 [ 1.87  1.85  3.46  1.37 73.4   9.56]
 [ 3.61  3.62  3.07  4.25 71.8   7.95]
 [ 1.43  1.63  1.48  5.43 69.13  5.28]
 [ 9.52  9.75 10.99 18.05 75.99 12.15]]


[[ 2.94  2.82  0.5   4.21 72.84  7.67]
 [ 1.6   1.49  0.97  2.21 74.18  9.  ]
 [ 0.69  0.63  0.46  0.46 69.99  4.82]
 [ 8.82  8.84  8.99 12.19 73.76  8.59]
 [13.26 13.27 13.11 18.09 76.68 11.5 ]]


[[ 2.97  2.83  0.51  4.22 72.83  7.63]
 [ 0.88  0.74  1.33  2.   74.17  8.97]
 [ 0.74  0.7   0.44  0.34 69.99  4.78]
 [ 8.63  8.63  8.62 12.22 73.75  8.55]
 [13.23 13.23 12.92 18.19 76.67 11.47]]


time spent until now: 25.8 mins


--------------------------------------
[pattern_seed, sd_R] = [4, 100]

max(u_O) =  193.8
O_threshold = 80
means of Order:

101.0 115.6 73.8 122.5 87.8

61.8 81.9 119.1 109.9 70.5

119.8 96.9 113.0 109.9 70.3

110.5 82.9 158.2 123.6 100.9

74.1 101.1 104.4 69.2 193.8

target policy:

1 1 0 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

0 1 1 0 1

number of reward locations:  19
O_threshold = 90
target policy:

1 1 0 1 0

0 0 1 1 0

1 1 1 1 0
```

```
1 0 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 105
target policy:

0 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 0

0 0 0 0 1

number of reward locations:  11
O_threshold = 120
target policy:

0 0 0 1 0

0 0 0 0 0

0 0 0 0 0

0 0 1 1 0

0 0 0 0 1

number of reward locations:  4
O_threshold = 135
target policy:

0 0 0 0 0

0 0 0 0 0

0 0 0 0 0

0 0 1 0 0

0 0 0 0 1

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

--------------------------------------
Value of Behaviour policy:65.322
O_threshold = 80
MC for this TARGET:[72.803, 0.717]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.69, 2.48, -0.91]][[4.38, -72.8, -7.48]]
std:[[4.38, 4.36, 1.32]][[1.07, 0.0, 0.02]]
MSE:[[5.14, 5.02, 1.6]][[4.51, 72.8, 7.48]]
MSE(-DR):[[0.0, -0.12, -3.54]][[-0.63, 67.66, 2.34]]
==============


O_threshold = 90
MC for this TARGET:[74.139, 0.715]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.96, -2.23, -2.48]][[1.38, -74.14, -8.82]]
std:[[1.11, 1.07, 0.17]][[0.48, 0.0, 0.02]]
MSE:[[2.25, 2.47, 2.49]][[1.46, 74.14, 8.82]]
MSE(-DR):[[0.0, 0.22, 0.24]][[-0.79, 71.89, 6.57]]
==============


O_threshold = 105
MC for this TARGET:[69.955, 0.718]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.67, 0.62, 1.04]][[-0.29, -69.96, -4.63]]
std:[[0.8, 0.84, 1.11]][[0.2, 0.0, 0.02]]
MSE:[[1.04, 1.04, 1.52]][[0.35, 69.96, 4.63]]
MSE(-DR):[[0.0, 0.0, 0.48]][[-0.69, 68.92, 3.59]]
==============


O_threshold = 120
MC for this TARGET:[73.723, 0.721]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.76, -7.73, -6.96]][[-12.35, -73.72, -8.4]]
std:[[0.7, 0.7, 1.65]][[0.15, 0.0, 0.02]]
MSE:[[7.79, 7.76, 7.15]][[12.35, 73.72, 8.4]]
MSE(-DR):[[0.0, -0.03, -0.64]][[4.56, 65.93, 0.61]]
```

```
**
==============


O_threshold = 135
MC for this TARGET:[76.64, 0.72]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-13.0, -13.04, -11.88]][[-18.64, -76.64, -11.32]]
std:[[1.87, 1.87, 4.13]][[0.3, 0.0, 0.02]]
MSE:[[13.13, 13.17, 12.58]][[18.64, 76.64, 11.32]]
MSE(-DR):[[0.0, 0.04, -0.55]][[5.51, 63.51, -1.81]]
**
==============


[[ 0.78  0.65  1.33  1.09 70.9  10.14]
 [ 0.43  0.34  1.17  0.73 69.38  8.62]
 [ 6.21  6.3   6.8   8.16 71.39 10.63]
 [ 9.06  9.08  8.42 13.57 70.56  9.8 ]
 [ 9.07  9.08  8.4  13.59 70.56  9.8 ]]


[[ 0.77  0.7   1.24  1.26 70.89 10.1 ]
 [ 0.21  0.15  1.3   0.57 69.37  8.59]
 [ 6.26  6.36  7.05  7.85 71.38 10.59]
 [ 9.04  9.01  8.4  13.45 70.55  9.76]
 [ 9.01  9.01  8.32 13.44 70.55  9.76]]


[[ 0.98  0.92  0.83  2.05 70.86  9.96]
 [ 0.77  0.64  2.03  0.18 69.34  8.44]
 [ 6.78  6.95  7.71  6.72 71.35 10.45]
 [ 9.33  9.28  8.18 12.82 70.52  9.62]
 [ 9.31  9.28  8.09 12.81 70.52  9.62]]


[[1.150e+00 1.080e+00 1.320e+00 2.930e+00 7.334e+01 9.640e+00]
 [5.100e-01 6.300e-01 3.110e+00 2.000e-02 7.344e+01 9.750e+00]
 [2.480e+00 2.560e+00 3.540e+00 4.540e+00 7.183e+01 8.140e+00]
 [2.480e+00 2.530e+00 2.470e+00 5.150e+00 6.916e+01 5.470e+00]
 [1.236e+01 1.236e+01 1.168e+01 1.750e+01 7.603e+01 1.233e+01]]


[[ 1.77  1.75  1.04  3.03 73.33  9.61]
 [ 0.79  0.86  3.24  0.26 73.44  9.71]
 [ 1.84  1.93  3.16  4.48 71.83  8.1 ]
 [ 2.19  2.29  1.89  5.17 69.16  5.43]
 [11.72 11.77 11.5  17.61 76.02 12.3 ]]


[[ 4.91  4.86  0.3   3.59 73.3   9.46]
 [ 1.87  1.85  3.46  1.37 73.4   9.56]
 [ 3.61  3.62  3.07  4.25 71.8   7.95]
 [ 1.43  1.63  1.48  5.43 69.13  5.28]
 [ 9.52  9.75 10.99 18.05 75.99 12.15]]


[[ 2.94  2.82  0.5   4.21 72.84  7.67]
 [ 1.6   1.49  0.97  2.21 74.18  9.  ]
 [ 0.69  0.63  0.46  0.46 69.99  4.82]
 [ 8.82  8.84  8.99 12.19 73.76  8.59]
 [13.26 13.27 13.11 18.09 76.68 11.5 ]]


[[ 2.97  2.83  0.51  4.22 72.83  7.63]
 [ 0.88  0.74  1.33  2.   74.17  8.97]
 [ 0.74  0.7   0.44  0.34 69.99  4.78]
 [ 8.63  8.63  8.62 12.22 73.75  8.55]
 [13.23 13.23 12.92 18.19 76.67 11.47]]


[[ 5.14  5.02  1.6   4.51 72.8   7.48]
 [ 2.25  2.47  2.49  1.46 74.14  8.82]
 [ 1.04  1.04  1.52  0.35 69.96  4.63]
 [ 7.79  7.76  7.15 12.35 73.72  8.4 ]
 [13.13 13.17 12.58 18.64 76.64 11.32]]


time spent until now: 29.1 mins

ubuntu@ip-172-31-9-82:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
14:19, 04/01; num of cores:36

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, simple, u_O_u_D]] = [672, 10, 10, None, 0.3, 0.5,
1, [True, False, True, False, 10]]

-------------------------------------
[pattern_seed, sd_R] = [2, 0.5]
```

```
Traceback (most recent call last):
  File "EC2.py", line 70, in <module>
    print_flag_target = False
TypeError: simu() got an unexpected keyword argument 'DGP_choice'
ubuntu@ip-172-31-9-82:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
14:19, 04/01; num of cores:36

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, simple, u_O_u_D]] = [672, 10, 10, None, 0.3, 0.5,
1, [True, False, True, False, 10]]


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

max(u_O) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 105
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 1 0 0

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

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 135
target policy:

0 0 0 1 0
```

```
0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
Traceback (most recent call last):
  File "EC2.py", line 70, in <module>
    print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 46, in simu
    neigh = adj2neigh(getAdjGrid(l, simple = simple))
NameError: name 'simple' is not defined
ubuntu@ip-172-31-9-82:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
14:20, 04/01; num of cores:36

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, simple, u_O_u_D]] = [672, 10, 10, None, 0.3, 0.5,
1, [True, False, True, False, 10]]


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

max(u_O) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 105
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 1 0 0

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

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0
```

```
0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 135
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
Process Process-1:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-5:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-2:
Process Process-4:
Process Process-3:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
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()
  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/_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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
  File "/home/ubuntu/main.py", line 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
NameError: name 'simple' is not defined
Process Process-10:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-16:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
Process Process-7:
NameError: name 'simple' is not defined
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-17:
Process Process-18:
Process Process-13:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-6:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-21:
Process Process-25:
Process Process-11:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-15:
Process Process-24:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
Process Process-20:
NameError: name 'simple' is not defined
Process Process-8:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-23:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-22:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-12:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-9:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-14:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
Process Process-19:
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 46, in getOneRegionValue
    Ta_i = Ta_disc(np.mean([pi[j](s = None, random_choose = True) for j in neigh[i]]), simple = simple)
NameError: name 'simple' is not defined
^[[A^CTraceback (most recent call last):
  File "EC2.py", line 70, in <module>
    print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 62, in simu
    value_reps = rep_seeds(once, OPE_rep_times)
  File "/home/ubuntu/_uti_basic.py", line 119, in rep_seeds
    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 202, in simu_once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 131, in V_DR
    r = arr(parmap(getOneRegionValue, range(N), n_cores))
  File "/home/ubuntu/_uti_basic.py", line 75, in parmap
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/_uti_basic.py", line 75, in <listcomp>
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
ubuntu@ip-172-31-9-82:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
Traceback (most recent call last):
  File "EC2.py", line 5, in <module>
    from simu_funs import *
  File "/home/ubuntu/simu_funs.py", line 198
    Ts = Ts, Ta = Ta, penalty = penalty, penalty_NMF = penalty_NMF,
     ^
SyntaxError: invalid syntax
ubuntu@ip-172-31-9-82:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
14:21, 04/01; num of cores:36

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, simple, u_O_u_D]] = [672, 10, 10, None, 0.3, 0.5,
1, [True, False, True, False, 10]]


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

max(u_O) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1
```

```
0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 105
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 1 0 0

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

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 135
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

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

--------------------------------------
Value of Behaviour policy:60.758
O_threshold = 80
MC for this TARGET:[70.898, 0.05]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.86, 0.65, -1.05]][[1.2, -70.9, -10.14]]
std:[[0.07, 0.05, 0.11]][[0.03, 0.0, 0.03]]
MSE:[[0.86, 0.65, 1.06]][[1.2, 70.9, 10.14]]
MSE(-DR):[[0.0, -0.21, 0.2]][[0.34, 70.04, 9.28]]
***
==============


O_threshold = 90
MC for this TARGET:[69.38, 0.056]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.4, 0.23, -0.78]][[-0.59, -69.38, -8.62]]
std:[[0.28, 0.25, 0.12]][[0.02, 0.0, 0.03]]
MSE:[[0.49, 0.34, 0.79]][[0.59, 69.38, 8.62]]
MSE(-DR):[[0.0, -0.15, 0.3]][[0.1, 68.89, 8.13]]
***
==============


O_threshold = 105
MC for this TARGET:[71.388, 0.056]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[-6.18, -6.3, -6.59]][[-8.07, -71.39, -10.63]]
std:[[0.05, 0.05, 0.17]][[0.03, 0.0, 0.03]]
MSE:[[6.18, 6.3, 6.59]][[8.07, 71.39, 10.63]]
MSE(-DR):[[0.0, 0.12, 0.41]][[1.89, 65.21, 4.45]]
***
==============


O_threshold = 120
MC for this TARGET:[70.557, 0.05]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.09, -9.08, -8.45]][[-13.52, -70.56, -9.8]]
std:[[0.3, 0.28, 0.13]][[0.02, 0.0, 0.03]]
MSE:[[9.09, 9.08, 8.45]][[13.52, 70.56, 9.8]]
MSE(-DR):[[0.0, -0.01, -0.64]][[4.43, 61.47, 0.71]]
**
==============


O_threshold = 135
MC for this TARGET:[70.557, 0.05]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.1, -9.08, -8.48]][[-13.54, -70.56, -9.8]]
std:[[0.3, 0.28, 0.15]][[0.04, 0.0, 0.03]]
MSE:[[9.1, 9.08, 8.48]][[13.54, 70.56, 9.8]]
MSE(-DR):[[0.0, -0.02, -0.62]][[4.44, 61.46, 0.7]]
**
==============


[[ 0.86  0.65  1.06  1.2  70.9  10.14]
 [ 0.49  0.34  0.79  0.59 69.38  8.62]
 [ 6.18  6.3   6.59  8.07 71.39 10.63]
 [ 9.09  9.08  8.45 13.52 70.56  9.8 ]
 [ 9.1   9.08  8.48 13.54 70.56  9.8 ]]


time spent until now: 5.2 mins


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

max(u_O) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 105
target policy:

0 0 0 1 0
```

```
0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 1 0 0

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

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 135
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

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

--------------------------------------
Value of Behaviour policy:60.786
O_threshold = 80
MC for this TARGET:[70.89, 0.157]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.91, 0.69, -0.86]][[1.4, -70.89, -10.1]]
std:[[0.12, 0.11, 0.09]][[0.07, 0.0, 0.03]]
MSE:[[0.92, 0.7, 0.86]][[1.4, 70.89, 10.1]]
MSE(-DR):[[0.0, -0.22, -0.06]][[0.48, 69.97, 9.18]]
**
==============


O_threshold = 90
MC for this TARGET:[69.373, 0.161]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.24, 0.08, -0.98]][[-0.49, -69.37, -8.59]]
std:[[0.13, 0.13, 0.02]][[0.04, 0.0, 0.03]]
MSE:[[0.27, 0.15, 0.98]][[0.49, 69.37, 8.59]]
MSE(-DR):[[0.0, -0.12, 0.71]][[0.22, 69.1, 8.32]]
***
==============


O_threshold = 105
MC for this TARGET:[71.38, 0.149]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.23, -6.34, -6.74]][[-7.75, -71.38, -10.59]]
std:[[0.5, 0.5, 0.26]][[0.22, 0.0, 0.03]]
MSE:[[6.25, 6.36, 6.75]][[7.75, 71.38, 10.59]]
MSE(-DR):[[0.0, 0.11, 0.5]][[1.5, 65.13, 4.34]]
***
==============


O_threshold = 120
MC for this TARGET:[70.549, 0.15]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.01, -9.0, -8.33]][[-13.38, -70.55, -9.76]]
std:[[0.37, 0.4, 0.02]][[0.04, 0.0, 0.03]]
MSE:[[9.02, 9.01, 8.33]][[13.38, 70.55, 9.76]]
MSE(-DR):[[0.0, -0.01, -0.69]][[4.36, 61.53, 0.74]]
**
==============


O_threshold = 135
MC for this TARGET:[70.549, 0.15]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[-9.02, -9.0, -8.38]][[-13.4, -70.55, -9.76]]
std:[[0.34, 0.4, 0.01]][[0.07, 0.0, 0.03]]
MSE:[[9.03, 9.01, 8.38]][[13.4, 70.55, 9.76]]
MSE(-DR):[[0.0, -0.02, -0.65]][[4.37, 61.52, 0.73]]
**
==============


[[ 0.86  0.65  1.06  1.2  70.9  10.14]
 [ 0.49  0.34  0.79  0.59 69.38  8.62]
 [ 6.18  6.3   6.59  8.07 71.39 10.63]
 [ 9.09  9.08  8.45 13.52 70.56  9.8 ]
 [ 9.1   9.08  8.48 13.54 70.56  9.8 ]]


[[ 0.92  0.7   0.86  1.4  70.89 10.1 ]
 [ 0.27  0.15  0.98  0.49 69.37  8.59]
 [ 6.25  6.36  6.75  7.75 71.38 10.59]
 [ 9.02  9.01  8.33 13.38 70.55  9.76]
 [ 9.03  9.01  8.38 13.4  70.55  9.76]]


time spent until now: 10.4 mins


______________________________________
[pattern_seed, sd_R] = [2, 100]

max(u_O) =  197.9
O_threshold = 80
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 105
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 1 0 0

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

0 0 0 1 0

0 0 0 0 0
```

```
0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

number of reward locations:  3
O_threshold = 135
target policy:

0 0 0 1 0

0 0 0 0 0

0 1 0 0 0

0 0 1 0 0

0 0 0 0 0

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

-------------------------------------
Value of Behaviour policy:60.903
O_threshold = 80
MC for this TARGET:[70.86, 0.725]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.89, 0.83, -0.52]][[2.17, -70.86, -9.96]]
std:[[0.23, 0.39, 0.27]][[0.09, 0.0, 0.01]]
MSE:[[0.92, 0.92, 0.59]][[2.17, 70.86, 9.96]]
MSE(-DR):[[0.0, 0.0, -0.33]][[1.25, 69.94, 9.04]]
**
==============


O_threshold = 90
MC for this TARGET:[69.342, 0.728]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.36, -0.54, -1.46]][[-0.0, -69.34, -8.44]]
std:[[0.38, 0.35, 0.05]][[0.03, 0.0, 0.01]]
MSE:[[0.52, 0.64, 1.46]][[0.03, 69.34, 8.44]]
MSE(-DR):[[0.0, 0.12, 0.94]][[-0.49, 68.82, 7.92]]
==============


O_threshold = 105
MC for this TARGET:[71.35, 0.715]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.46, -6.53, -7.54]][[-6.52, -71.35, -10.45]]
std:[[2.26, 2.37, 0.57]][[1.22, 0.0, 0.01]]
MSE:[[6.84, 6.95, 7.56]][[6.63, 71.35, 10.45]]
MSE(-DR):[[0.0, 0.11, 0.72]][[-0.21, 64.51, 3.61]]
==============


O_threshold = 120
MC for this TARGET:[70.519, 0.718]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.86, -8.71, -8.16]][[-12.82, -70.52, -9.62]]
std:[[3.12, 3.2, 0.68]][[0.43, 0.0, 0.01]]
MSE:[[9.39, 9.28, 8.19]][[12.83, 70.52, 9.62]]
MSE(-DR):[[0.0, -0.11, -1.2]][[3.44, 61.13, 0.23]]
**
==============


O_threshold = 135
MC for this TARGET:[70.519, 0.718]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.63, -8.71, -7.86]][[-12.75, -70.52, -9.62]]
std:[[3.26, 3.2, 0.79]][[0.41, 0.0, 0.01]]
MSE:[[9.23, 9.28, 7.9]][[12.76, 70.52, 9.62]]
MSE(-DR):[[0.0, 0.05, -1.33]][[3.53, 61.29, 0.39]]
**
==============


[[ 0.86  0.65  1.06  1.2  70.9  10.14]
 [ 0.49  0.34  0.79  0.59 69.38  8.62]
 [ 6.18  6.3   6.59  8.07 71.39 10.63]
 [ 9.09  9.08  8.45 13.52 70.56  9.8 ]
 [ 9.1   9.08  8.48 13.54 70.56  9.8 ]]


[[ 0.92  0.7   0.86  1.4  70.89 10.1 ]
 [ 0.27  0.15  0.98  0.49 69.37  8.59]
 [ 6.25  6.36  6.75  7.75 71.38 10.59]
 [ 9.02  9.01  8.33 13.38 70.55  9.76]
 [ 9.03  9.01  8.38 13.4  70.55  9.76]]


[[9.200e-01 9.200e-01 5.900e-01 2.170e+00 7.086e+01 9.960e+00]
 [5.200e-01 6.400e-01 1.460e+00 3.000e-02 6.934e+01 8.440e+00]
 [6.840e+00 6.950e+00 7.560e+00 6.630e+00 7.135e+01 1.045e+01]
 [9.390e+00 9.280e+00 8.190e+00 1.283e+01 7.052e+01 9.620e+00]
 [9.230e+00 9.280e+00 7.900e+00 1.276e+01 7.052e+01 9.620e+00]]


time spent until now: 15.6 mins


--------------------------------------
[pattern_seed, sd_R] = [3, 0.5]

max(u_O) =  170.1
O_threshold = 80
means of Order:

170.1 113.4 102.4 56.9 91.5

89.4 97.0 82.4 98.2 86.2

67.1 129.7 129.6 166.1 101.0

88.1 84.5 62.6 133.6 71.5

69.7 93.5 155.4 106.8 73.2

target policy:

1 1 1 0 1

1 1 1 1 1

0 1 1 1 1

1 1 0 1 0

0 1 1 1 0

number of reward locations:  19
O_threshold = 90
target policy:

1 1 1 0 1

0 1 0 1 0

0 1 1 1 1

0 0 0 1 0

0 1 1 1 0

number of reward locations:  14
O_threshold = 105
target policy:

1 1 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 1 0

number of reward locations:  8
O_threshold = 120
target policy:

1 0 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 0 0

number of reward locations:  6
O_threshold = 135
target policy:
```