```
Last login: Thu Apr  2 16:25:00 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-34-226-244-218.compute-1.amazonaws.com
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1063-aws x86_64)

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

 System information disabled due to load higher than 96.0

 * Kubernetes 1.18 GA is now available! See https://microk8s.io for docs or
   install it with:

     sudo snap install microk8s --channel=1.18 --classic

 * Multipass 1.1 adds proxy support for developers behind enterprise
   firewalls. Rapid prototyping for cloud operations just got easier.

     https://multipass.run/

 * Canonical Livepatch is available for installation.
   - Reduce system reboots and improve kernel security. Activate at:
     https://ubuntu.com/livepatch

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


Last login: Thu Apr  2 20:25:51 2020 from 107.13.161.147
ubuntu@ip-172-31-79-148:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
18:40, 04/02; num of cores:96

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, simple, u_O_u_D]] = [N
one, 96, 10, 10, None, 0.3, 0.5, 1, [True, False, True, False, 10]]


---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 1]

max(u_O) =  197.9
O_threshold = 80
number of reward locations:  15
O_threshold = 90
number of reward locations:  12
O_threshold = 100
number of reward locations:  8
O_threshold = 110
number of reward locations:  6
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!

---------------------------------------
Value of Behaviour policy:60.785
O_threshold = 80
MC for this TARGET:[70.881, 0.074]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.55, -0.9, -1.01]][[-70.88, -70.88, -10.1]]
std:[[0.44, 0.47, 0.26]][[0.0, 0.0, 0.19]]
MSE:[[0.7, 1.02, 1.04]][[70.88, 70.88, 10.1]]
```

```
MSE(-DR):[[0.0, 0.32, 0.34]][[70.18, 70.18, 9.4]]
***
==============


O_threshold = 90
MC for this TARGET:[69.368, 0.068]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.61, -0.94, -1.08]][[-69.37, -69.37, -8.58]]
std:[[0.38, 0.41, 0.27]][[0.0, 0.0, 0.19]]
MSE:[[0.72, 1.03, 1.11]][[69.37, 69.37, 8.58]]
MSE(-DR):[[0.0, 0.31, 0.39]][[68.65, 68.65, 7.86]]
***
MC-based ATE = -1.51
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.06, -0.04, -0.06]][[1.51, 1.51, 1.51]]
std:[[0.39, 0.43, 0.22]][[0.0, 0.0, 0.0]]
MSE:[[0.39, 0.43, 0.23]][[1.51, 1.51, 1.51]]
MSE(-DR):[[0.0, 0.04, -0.16]][[1.12, 1.12, 1.12]]
==============


O_threshold = 100
MC for this TARGET:[68.937, 0.074]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.1, -3.36, -3.35]][[-68.94, -68.94, -8.15]]
std:[[0.42, 0.43, 0.25]][[0.0, 0.0, 0.19]]
MSE:[[3.13, 3.39, 3.36]][[68.94, 68.94, 8.15]]
MSE(-DR):[[0.0, 0.26, 0.23]][[65.81, 65.81, 5.02]]
***
MC-based ATE = -1.94
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.54, -2.46, -2.33]][[1.94, 1.94, 1.94]]
std:[[0.44, 0.46, 0.25]][[0.0, 0.0, 0.0]]
MSE:[[2.58, 2.5, 2.34]][[1.94, 1.94, 1.94]]
MSE(-DR):[[0.0, -0.08, -0.24]][[-0.64, -0.64, -0.64]]
==============


O_threshold = 110
MC for this TARGET:[70.481, 0.081]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.97, -7.24, -6.74]][[-70.48, -70.48, -9.7]]
std:[[0.51, 0.53, 0.32]][[0.0, 0.0, 0.19]]
MSE:[[6.99, 7.26, 6.75]][[70.48, 70.48, 9.7]]
MSE(-DR):[[0.0, 0.27, -0.24]][[63.49, 63.49, 2.71]]
**
MC-based ATE = -0.4
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.41, -6.34, -5.73]][[0.4, 0.4, 0.4]]
std:[[0.6, 0.63, 0.32]][[0.0, 0.0, 0.0]]
MSE:[[6.44, 6.37, 5.74]][[0.4, 0.4, 0.4]]
MSE(-DR):[[0.0, -0.07, -0.7]][[-6.04, -6.04, -6.04]]
==============


[[ 0.7   1.02  1.04 70.88 70.88 10.1 ]
 [ 0.72  1.03  1.11 69.37 69.37  8.58]
 [ 3.13  3.39  3.36 68.94 68.94  8.15]
 [ 6.99  7.26  6.75 70.48 70.48  9.7 ]]


time spent until now: 54.0 mins


---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 5]

max(u_O) =  197.9
O_threshold = 80
number of reward locations:  15
O_threshold = 90
number of reward locations:  12
O_threshold = 100
number of reward locations:  8
O_threshold = 110
number of reward locations:  6
1 -th region DONE!
```

```
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!

---------------------------------------
Value of Behaviour policy:60.785
O_threshold = 80
MC for this TARGET:[70.882, 0.098]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.53, -0.87, -1.02]][[-70.88, -70.88, -10.1]]
std:[[0.54, 0.57, 0.3]][[0.0, 0.0, 0.2]]
MSE:[[0.76, 1.04, 1.06]][[70.88, 70.88, 10.1]]
MSE(-DR):[[0.0, 0.28, 0.3]][[70.12, 70.12, 9.34]]
***
==============


O_threshold = 90
MC for this TARGET:[69.37, 0.09]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.59, -0.91, -1.08]][[-69.37, -69.37, -8.58]]
std:[[0.49, 0.51, 0.31]][[0.0, 0.0, 0.2]]
MSE:[[0.77, 1.04, 1.12]][[69.37, 69.37, 8.58]]
MSE(-DR):[[0.0, 0.27, 0.35]][[68.6, 68.6, 7.81]]
***
MC-based ATE = -1.51
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.06, -0.04, -0.06]][[1.51, 1.51, 1.51]]
std:[[0.5, 0.53, 0.26]][[0.0, 0.0, 0.0]]
MSE:[[0.5, 0.53, 0.27]][[1.51, 1.51, 1.51]]
MSE(-DR):[[0.0, 0.03, -0.23]][[1.01, 1.01, 1.01]]
==============


O_threshold = 100
MC for this TARGET:[68.938, 0.092]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.07, -3.33, -3.34]][[-68.94, -68.94, -8.15]]
std:[[0.49, 0.5, 0.3]][[0.0, 0.0, 0.2]]
MSE:[[3.11, 3.37, 3.35]][[68.94, 68.94, 8.15]]
MSE(-DR):[[0.0, 0.26, 0.24]][[65.83, 65.83, 5.04]]
***
MC-based ATE = -1.94
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.54, -2.46, -2.32]][[1.94, 1.94, 1.94]]
std:[[0.55, 0.58, 0.29]][[0.0, 0.0, 0.0]]
MSE:[[2.6, 2.53, 2.34]][[1.94, 1.94, 1.94]]
MSE(-DR):[[0.0, -0.07, -0.26]][[-0.66, -0.66, -0.66]]
==============


O_threshold = 110
MC for this TARGET:[70.482, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.94, -7.22, -6.75]][[-70.48, -70.48, -9.7]]
std:[[0.57, 0.59, 0.35]][[0.0, 0.0, 0.2]]
MSE:[[6.96, 7.24, 6.76]][[70.48, 70.48, 9.7]]
MSE(-DR):[[0.0, 0.28, -0.2]][[63.52, 63.52, 2.74]]
**
MC-based ATE = -0.4
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.42, -6.35, -5.73]][[0.4, 0.4, 0.4]]
```

```
std:[[0.7, 0.74, 0.36]][[0.0, 0.0, 0.0]]
MSE:[[6.46, 6.39, 5.74]][[0.4, 0.4, 0.4]]
MSE(-DR):[[0.0, -0.07, -0.72]][[-6.06, -6.06, -6.06]]
==============


[[ 0.7   1.02  1.04 70.88 70.88 10.1 ]
 [ 0.72  1.03  1.11 69.37 69.37  8.58]
 [ 3.13  3.39  3.36 68.94 68.94  8.15]
 [ 6.99  7.26  6.75 70.48 70.48  9.7 ]]


[[ 0.76  1.04  1.06 70.88 70.88 10.1 ]
 [ 0.77  1.04  1.12 69.37 69.37  8.58]
 [ 3.11  3.37  3.35 68.94 68.94  8.15]
 [ 6.96  7.24  6.76 70.48 70.48  9.7 ]]


time spent until now: 110.8 mins


---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 10]

max(u_O) =  197.9
O_threshold = 80
number of reward locations:  15
O_threshold = 90
number of reward locations:  12
O_threshold = 100
number of reward locations:  8
O_threshold = 110
number of reward locations:  6
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!

---------------------------------------
Value of Behaviour policy:60.786
O_threshold = 80
MC for this TARGET:[70.884, 0.141]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.49, -0.82, -1.01]][[-70.88, -70.88, -10.1]]
std:[[0.66, 0.67, 0.37]][[0.0, 0.0, 0.23]]
MSE:[[0.82, 1.06, 1.08]][[70.88, 70.88, 10.1]]
MSE(-DR):[[0.0, 0.24, 0.26]][[70.06, 70.06, 9.28]]
***
==============


O_threshold = 90
MC for this TARGET:[69.371, 0.133]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.55, -0.85, -1.06]][[-69.37, -69.37, -8.59]]
std:[[0.68, 0.71, 0.41]][[0.0, 0.0, 0.23]]
MSE:[[0.87, 1.11, 1.14]][[69.37, 69.37, 8.59]]
MSE(-DR):[[0.0, 0.24, 0.27]][[68.5, 68.5, 7.72]]
***
MC-based ATE = -1.51
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.05, -0.03, -0.05]][[1.51, 1.51, 1.51]]
```

```
std:[[0.64, 0.66, 0.33]][[0.0, 0.0, 0.0]]
MSE:[[0.64, 0.66, 0.33]][[1.51, 1.51, 1.51]]
MSE(-DR):[[0.0, 0.02, -0.31]][[0.87, 0.87, 0.87]]
==============


O_threshold = 100
MC for this TARGET:[68.94, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.02, -3.26, -3.34]][[-68.94, -68.94, -8.15]]
std:[[0.63, 0.64, 0.38]][[0.0, 0.0, 0.23]]
MSE:[[3.09, 3.32, 3.36]][[68.94, 68.94, 8.15]]
MSE(-DR):[[0.0, 0.23, 0.27]][[65.85, 65.85, 5.06]]
***
MC-based ATE = -1.94
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.52, -2.45, -2.33]][[1.94, 1.94, 1.94]]
std:[[0.71, 0.72, 0.4]][[0.0, 0.0, 0.0]]
MSE:[[2.62, 2.55, 2.36]][[1.94, 1.94, 1.94]]
MSE(-DR):[[0.0, -0.07, -0.26]][[-0.68, -0.68, -0.68]]
==============


O_threshold = 110
MC for this TARGET:[70.484, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.89, -7.14, -6.75]][[-70.48, -70.48, -9.7]]
std:[[0.71, 0.73, 0.44]][[0.0, 0.0, 0.23]]
MSE:[[6.93, 7.18, 6.76]][[70.48, 70.48, 9.7]]
MSE(-DR):[[0.0, 0.25, -0.17]][[63.55, 63.55, 2.77]]
**
MC-based ATE = -0.4
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.39, -6.32, -5.74]][[0.4, 0.4, 0.4]]
std:[[0.89, 0.92, 0.46]][[0.0, 0.0, 0.0]]
MSE:[[6.45, 6.39, 5.76]][[0.4, 0.4, 0.4]]
MSE(-DR):[[0.0, -0.06, -0.69]][[-6.05, -6.05, -6.05]]
==============


[[ 0.7   1.02  1.04 70.88 70.88 10.1 ]
 [ 0.72  1.03  1.11 69.37 69.37  8.58]
 [ 3.13  3.39  3.36 68.94 68.94  8.15]
 [ 6.99  7.26  6.75 70.48 70.48  9.7 ]]


[[ 0.76  1.04  1.06 70.88 70.88 10.1 ]
 [ 0.77  1.04  1.12 69.37 69.37  8.58]
 [ 3.11  3.37  3.35 68.94 68.94  8.15]
 [ 6.96  7.24  6.76 70.48 70.48  9.7 ]]


[[ 0.82  1.06  1.08 70.88 70.88 10.1 ]
 [ 0.87  1.11  1.14 69.37 69.37  8.59]
 [ 3.09  3.32  3.36 68.94 68.94  8.15]
 [ 6.93  7.18  6.76 70.48 70.48  9.7 ]]


time spent until now: 165.3 mins


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

max(u_O) =  197.9
O_threshold = 80
number of reward locations:  15
O_threshold = 90
number of reward locations:  12
O_threshold = 100
number of reward locations:  8
O_threshold = 110
number of reward locations:  6
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
```

```
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!
1 -th region DONE!
6 -th region DONE!
11 -th region DONE!
16 -th region DONE!
21 -th region DONE!

--------------------------------------
Value of Behaviour policy:60.787
O_threshold = 80
MC for this TARGET:[70.887, 0.239]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.53, -0.85, -1.02]][[-70.89, -70.89, -10.1]]
std:[[1.03, 1.05, 0.61]][[0.0, 0.0, 0.31]]
MSE:[[1.16, 1.35, 1.19]][[70.89, 70.89, 10.1]]
MSE(-DR):[[0.0, 0.19, 0.03]][[69.73, 69.73, 8.94]]
***
==============


O_threshold = 90
MC for this TARGET:[69.375, 0.232]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.6, -0.89, -1.06]][[-69.38, -69.38, -8.59]]
std:[[1.13, 1.14, 0.68]][[0.0, 0.0, 0.31]]
MSE:[[1.28, 1.45, 1.26]][[69.38, 69.38, 8.6]]
MSE(-DR):[[0.0, 0.17, -0.02]][[68.1, 68.1, 7.32]]
**
MC-based ATE = -1.51
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.06, -0.04, -0.04]][[1.51, 1.51, 1.51]]
std:[[1.02, 1.03, 0.56]][[0.0, 0.0, 0.0]]
MSE:[[1.02, 1.03, 0.56]][[1.51, 1.51, 1.51]]
MSE(-DR):[[0.0, 0.01, -0.46]][[0.49, 0.49, 0.49]]
==============


O_threshold = 100
MC for this TARGET:[68.943, 0.229]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.97, -3.22, -3.32]][[-68.94, -68.94, -8.16]]
std:[[1.01, 1.02, 0.62]][[0.0, 0.0, 0.31]]
MSE:[[3.14, 3.38, 3.38]][[68.94, 68.94, 8.17]]
MSE(-DR):[[0.0, 0.24, 0.24]][[65.8, 65.8, 5.03]]
***
MC-based ATE = -1.94
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.44, -2.37, -2.3]][[1.94, 1.94, 1.94]]
std:[[1.14, 1.15, 0.69]][[0.0, 0.0, 0.0]]
MSE:[[2.69, 2.63, 2.4]][[1.94, 1.94, 1.94]]
MSE(-DR):[[0.0, -0.06, -0.29]][[-0.75, -0.75, -0.75]]
==============


O_threshold = 110
MC for this TARGET:[70.487, 0.229]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.94, -7.18, -6.75]][[-70.49, -70.49, -9.7]]
std:[[1.14, 1.15, 0.69]][[0.0, 0.0, 0.31]]
MSE:[[7.03, 7.27, 6.79]][[70.49, 70.49, 9.7]]
MSE(-DR):[[0.0, 0.24, -0.24]][[63.46, 63.46, 2.67]]
**
MC-based ATE = -0.4
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.41, -6.33, -5.73]][[0.4, 0.4, 0.4]]
std:[[1.44, 1.47, 0.76]][[0.0, 0.0, 0.0]]
MSE:[[6.57, 6.5, 5.78]][[0.4, 0.4, 0.4]]
MSE(-DR):[[0.0, -0.07, -0.79]][[-6.17, -6.17, -6.17]]
==============
```

```
[[ 0.7   1.02  1.04 70.88 70.88 10.1 ]
 [ 0.72  1.03  1.11 69.37 69.37  8.58]
 [ 3.13  3.39  3.36 68.94 68.94  8.15]
 [ 6.99  7.26  6.75 70.48 70.48  9.7 ]]


[[ 0.76  1.04  1.06 70.88 70.88 10.1 ]
 [ 0.77  1.04  1.12 69.37 69.37  8.58]
 [ 3.11  3.37  3.35 68.94 68.94  8.15]
 [ 6.96  7.24  6.76 70.48 70.48  9.7 ]]


[[ 0.82  1.06  1.08 70.88 70.88 10.1 ]
 [ 0.87  1.11  1.14 69.37 69.37  8.59]
 [ 3.09  3.32  3.36 68.94 68.94  8.15]
 [ 6.93  7.18  6.76 70.48 70.48  9.7 ]]


[[ 1.16  1.35  1.19 70.89 70.89 10.1 ]
 [ 1.28  1.45  1.26 69.38 69.38  8.6 ]
 [ 3.14  3.38  3.38 68.94 68.94  8.17]
 [ 7.03  7.27  6.79 70.49 70.49  9.7 ]]


time spent until now: 221.9 mins


--------------------------------------
[pattern_seed, day, sd_R] = [2, 14, 1]

max(u_O) =  197.9
O_threshold = 80
number of reward locations:  15
O_threshold = 90
number of reward locations:  12
O_threshold = 100
number of reward locations:  8
O_threshold = 110
number of reward locations:  6
1 -th region DONE!
```