```
Last login: Wed Apr 15 22:15:15 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-235-106-98.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 72.0

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

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

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

     https://multipass.run/

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

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


*** System restart required ***
Last login: Thu Apr 16 02:15:18 2020 from 107.13.161.147
ubuntu@ip-172-31-13-166:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
22:31, 04/15; num of cores:72
sd_u_O_35_uo_ud_0_10

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D_range, t_func] = [16, None, None, 35, 0.5, 1.5, [10, 20
], None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[100, 105, 110, 120], [0, 20, 40], [3, 7], [[0.0001, 5e-05], [0.0
001, 5e-05]]]

----------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 3, 0, 10]

max(u_O) =  180.2
O_threshold = 100
means of Order:

85.4 98.0 25.2 157.4 37.2

70.5 117.6 56.4 63.0 68.2

119.3 180.2 101.5 60.9 118.9

79.1 99.3 141.1 73.8 100.3

69.3 94.5 109.0 65.4 88.1

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 1

0 0 1 0 0

number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 120
number of reward locations:  3
target 1 in 1 DONE!
```

```
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

--------------------------------------
Value of Behaviour policy:49.096
O_threshold = 100
MC for this TARGET:[60.27, 0.118]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.44, -3.59, -3.92]][[-4.7, -60.27, -11.17]]
std:[[0.84, 0.86, 0.46]][[0.58, 0.0, 0.25]]
MSE:[[3.54, 3.69, 3.95]][[4.74, 60.27, 11.17]]
MSE(-DR):[[0.0, 0.15, 0.41]][[1.2, 56.73, 7.63]]
***
==============
O_threshold = 105
MC for this TARGET:[58.995, 0.115]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.88, -4.97, -5.29]][[-6.79, -59.0, -9.9]]
std:[[1.1, 1.12, 0.43]][[0.58, 0.0, 0.25]]
MSE:[[5.0, 5.09, 5.31]][[6.81, 59.0, 9.9]]
MSE(-DR):[[0.0, 0.09, 0.31]][[1.81, 54.0, 4.9]]
***
==============
O_threshold = 110
MC for this TARGET:[57.438, 0.113]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.24, -4.3, -4.86]][[-7.44, -57.44, -8.34]]
std:[[1.07, 1.08, 0.48]][[0.53, 0.0, 0.25]]
MSE:[[4.37, 4.43, 4.88]][[7.46, 57.44, 8.34]]
MSE(-DR):[[0.0, 0.06, 0.51]][[3.09, 53.07, 3.97]]
***
==============
O_threshold = 120
MC for this TARGET:[57.899, 0.091]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.69, -7.7, -8.33]][[-14.36, -57.9, -8.8]]
std:[[1.07, 1.09, 0.85]][[0.48, 0.0, 0.25]]
MSE:[[7.76, 7.78, 8.37]][[14.37, 57.9, 8.8]]
MSE(-DR):[[0.0, 0.02, 0.61]][[6.61, 50.14, 1.04]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 3.54  3.69  3.95  4.74 60.27 11.17]
 [ 5.    5.09  5.31  6.81 59.    9.9 ]
 [ 4.37  4.43  4.88  7.46 57.44  8.34]
 [ 7.76  7.78  8.37 14.37 57.9   8.8 ]]


time spent until now: 11.9 mins

22:43, 04/15

--------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 3, 0, 20]

max(u_O) =  180.2
O_threshold = 100
means of Order:

85.4 98.0 25.2 157.4 37.2

70.5 117.6 56.4 63.0 68.2

119.3 180.2 101.5 60.9 118.9

79.1 99.3 141.1 73.8 100.3

69.3 94.5 109.0 65.4 88.1

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1
```

```
0 0 1 0 1

0 0 1 0 0

number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 120
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

--------------------------------------
Value of Behaviour policy:45.21
O_threshold = 100
MC for this TARGET:[56.07, 0.084]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.21, -5.34, -5.26]][[-6.59, -56.07, -10.86]]
std:[[0.69, 0.71, 0.4]][[0.53, 0.0, 0.24]]
MSE:[[5.26, 5.39, 5.28]][[6.61, 56.07, 10.86]]
MSE(-DR):[[0.0, 0.13, 0.02]][[1.35, 50.81, 5.6]]
***
==============
O_threshold = 105
MC for this TARGET:[55.254, 0.105]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.33, -7.38, -7.14]][[-9.47, -55.25, -10.04]]
std:[[1.08, 1.1, 0.48]][[0.53, 0.0, 0.24]]
MSE:[[7.41, 7.46, 7.16]][[9.48, 55.25, 10.04]]
MSE(-DR):[[0.0, 0.05, -0.25]][[2.07, 47.84, 2.63]]
**
==============
O_threshold = 110
MC for this TARGET:[53.472, 0.102]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.26, -6.31, -6.54]][[-9.68, -53.47, -8.26]]
std:[[1.06, 1.07, 0.51]][[0.5, 0.0, 0.24]]
MSE:[[6.35, 6.4, 6.56]][[9.69, 53.47, 8.26]]
MSE(-DR):[[0.0, 0.05, 0.21]][[3.34, 47.12, 1.91]]
***
==============
O_threshold = 120
MC for this TARGET:[52.728, 0.077]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.14, -8.15, -8.78]][[-15.08, -52.73, -7.52]]
std:[[1.03, 1.03, 0.78]][[0.43, 0.0, 0.24]]
MSE:[[8.2, 8.21, 8.81]][[15.09, 52.73, 7.52]]
MSE(-DR):[[0.0, 0.01, 0.61]][[6.89, 44.53, -0.68]]
***
==============
[[ 3.54  3.69  3.95  4.74 60.27 11.17]
 [ 5.    5.09  5.31  6.81 59.    9.9 ]
 [ 4.37  4.43  4.88  7.46 57.44  8.34]
 [ 7.76  7.78  8.37 14.37 57.9   8.8 ]]


[[ 5.26  5.39  5.28  6.61 56.07 10.86]
 [ 7.41  7.46  7.16  9.48 55.25 10.04]
 [ 6.35  6.4   6.56  9.69 53.47  8.26]
 [ 8.2   8.21  8.81 15.09 52.73  7.52]]


time spent until now: 23.7 mins

22:55, 04/15

--------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 7, 0, 10]

max(u_O) =  180.2
O_threshold = 100
means of Order:

85.4 98.0 25.2 157.4 37.2
```

```
70.5 117.6 56.4 63.0 68.2

119.3 180.2 101.5 60.9 118.9

79.1 99.3 141.1 73.8 100.3

69.3 94.5 109.0 65.4 88.1

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 1

0 0 1 0 0

number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 120
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

--------------------------------------
Value of Behaviour policy:49.056
O_threshold = 100
MC for this TARGET:[60.247, 0.083]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.71, -2.87, -3.68]][[-4.36, -60.25, -11.19]]
std:[[0.59, 0.61, 0.42]][[0.26, 0.0, 0.24]]
MSE:[[2.77, 2.93, 3.7]][[4.37, 60.25, 11.19]]
MSE(-DR):[[0.0, 0.16, 0.93]][[1.6, 57.48, 8.42]]
***
==============
O_threshold = 105
MC for this TARGET:[58.982, 0.072]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.05, -4.19, -4.96]][[-6.48, -58.98, -9.93]]
std:[[0.62, 0.65, 0.38]][[0.3, 0.0, 0.24]]
MSE:[[4.1, 4.24, 4.97]][[6.49, 58.98, 9.93]]
MSE(-DR):[[0.0, 0.14, 0.87]][[2.39, 54.88, 5.83]]
***
==============
O_threshold = 110
MC for this TARGET:[57.42, 0.066]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.7, -3.83, -4.49]][[-7.13, -57.42, -8.36]]
std:[[0.66, 0.67, 0.4]][[0.28, 0.0, 0.24]]
MSE:[[3.76, 3.89, 4.51]][[7.14, 57.42, 8.36]]
MSE(-DR):[[0.0, 0.13, 0.75]][[3.38, 53.66, 4.6]]
***
==============
O_threshold = 120
MC for this TARGET:[57.889, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.08, -8.08, -8.29]][[-14.11, -57.89, -8.83]]
std:[[0.88, 0.92, 0.4]][[0.31, 0.0, 0.24]]
MSE:[[8.13, 8.13, 8.3]][[14.11, 57.89, 8.83]]
MSE(-DR):[[0.0, 0.0, 0.17]][[5.98, 49.76, 0.7]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 3.54  3.69  3.95  4.74 60.27 11.17]
 [ 5.    5.09  5.31  6.81 59.    9.9 ]
 [ 4.37  4.43  4.88  7.46 57.44  8.34]
 [ 7.76  7.78  8.37 14.37 57.9   8.8 ]]
```

```
[[ 5.26  5.39  5.28  6.61 56.07 10.86]
 [ 7.41  7.46  7.16  9.48 55.25 10.04]
 [ 6.35  6.4   6.56  9.69 53.47  8.26]
 [ 8.2   8.21  8.81 15.09 52.73  7.52]]


[[ 2.77  2.93  3.7   4.37 60.25 11.19]
 [ 4.1   4.24  4.97  6.49 58.98  9.93]
 [ 3.76  3.89  4.51  7.14 57.42  8.36]
 [ 8.13  8.13  8.3  14.11 57.89  8.83]]


time spent until now: 37.1 mins

23:08, 04/15

--------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 7, 0, 20]

max(u_O) =  180.2
O_threshold = 100
means of Order:

85.4 98.0 25.2 157.4 37.2

70.5 117.6 56.4 63.0 68.2

119.3 180.2 101.5 60.9 118.9

79.1 99.3 141.1 73.8 100.3

69.3 94.5 109.0 65.4 88.1

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 1

0 0 1 0 0

number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 120
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

--------------------------------------
Value of Behaviour policy:45.169
O_threshold = 100
MC for this TARGET:[56.061, 0.067]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.52, -4.69, -5.07]][[-6.33, -56.06, -10.89]]
std:[[0.44, 0.45, 0.36]][[0.23, 0.0, 0.2]]
MSE:[[4.54, 4.71, 5.08]][[6.33, 56.06, 10.89]]
MSE(-DR):[[0.0, 0.17, 0.54]][[1.79, 51.52, 6.35]]
***
==============
O_threshold = 105
MC for this TARGET:[55.247, 0.07]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.53, -6.66, -6.91]][[-9.25, -55.25, -10.08]]
std:[[0.52, 0.54, 0.35]][[0.26, 0.0, 0.2]]
MSE:[[6.55, 6.68, 6.92]][[9.25, 55.25, 10.08]]
MSE(-DR):[[0.0, 0.13, 0.37]][[2.7, 48.7, 3.53]]
***
==============
O_threshold = 110
```

```
MC for this TARGET:[53.459, 0.065]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.87, -5.96, -6.39]][[-9.47, -53.46, -8.29]]
std:[[0.6, 0.61, 0.37]][[0.24, 0.0, 0.2]]
MSE:[[5.9, 5.99, 6.4]][[9.47, 53.46, 8.29]]
MSE(-DR):[[0.0, 0.09, 0.5]][[3.57, 47.56, 2.39]]
***
==============
O_threshold = 120
MC for this TARGET:[52.716, 0.053]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.64, -8.62, -8.91]][[-14.9, -52.72, -7.55]]
std:[[0.75, 0.78, 0.4]][[0.25, 0.0, 0.2]]
MSE:[[8.67, 8.66, 8.92]][[14.9, 52.72, 7.55]]
MSE(-DR):[[0.0, -0.01, 0.25]][[6.23, 44.05, -1.12]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 3.54  3.69  3.95  4.74 60.27 11.17]
 [ 5.    5.09  5.31  6.81 59.    9.9 ]
 [ 4.37  4.43  4.88  7.46 57.44  8.34]
 [ 7.76  7.78  8.37 14.37 57.9   8.8 ]]


[[ 5.26  5.39  5.28  6.61 56.07 10.86]
 [ 7.41  7.46  7.16  9.48 55.25 10.04]
 [ 6.35  6.4   6.56  9.69 53.47  8.26]
 [ 8.2   8.21  8.81 15.09 52.73  7.52]]


[[ 2.77  2.93  3.7   4.37 60.25 11.19]
 [ 4.1   4.24  4.97  6.49 58.98  9.93]
 [ 3.76  3.89  4.51  7.14 57.42  8.36]
 [ 8.13  8.13  8.3  14.11 57.89  8.83]]


[[ 4.54  4.71  5.08  6.33 56.06 10.89]
 [ 6.55  6.68  6.92  9.25 55.25 10.08]
 [ 5.9   5.99  6.4   9.47 53.46  8.29]
 [ 8.67  8.66  8.92 14.9  52.72  7.55]]


time spent until now: 50.4 mins

23:21, 04/15

--------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 3, 20, 10]

max(u_O) =  180.2
O_threshold = 100
means of Order:

85.4 98.0 25.2 157.4 37.2

70.5 117.6 56.4 63.0 68.2

119.3 180.2 101.5 60.9 118.9

79.1 99.3 141.1 73.8 100.3

69.3 94.5 109.0 65.4 88.1

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 1

0 0 1 0 0

number of reward locations:  9
O_threshold = 105
number of reward locations:  7
```