```
Last login: Wed Apr 15 23:22:48 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 03:22:51 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
23:23, 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, 25, 0.5, 1.5, [0, 10]
, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[100, 105, 110, 115], [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, 0]

max(u_O) =  157.3
O_threshold = 100
means of Order:

89.6 98.6 46.6 141.0 55.2

79.0 112.6 68.9 73.6 77.3

113.8 157.3 101.0 72.1 113.5

85.1 99.5 129.4 81.3 100.2

78.0 96.1 106.4 75.3 91.5

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 = 115
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:54.244
O_threshold = 100
MC for this TARGET:[61.275, 0.114]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.53, -0.62, -1.43]][[-1.01, -61.28, -7.03]]
std:[[1.11, 1.13, 0.56]][[0.64, 0.0, 0.35]]
MSE:[[1.23, 1.29, 1.54]][[1.2, 61.28, 7.04]]
MSE(-DR):[[0.0, 0.06, 0.31]][[-0.03, 60.05, 5.81]]
==============
O_threshold = 105
MC for this TARGET:[61.836, 0.103]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.76, -2.84, -3.89]][[-4.3, -61.84, -7.59]]
std:[[1.25, 1.25, 0.61]][[0.62, 0.0, 0.35]]
MSE:[[3.03, 3.1, 3.94]][[4.34, 61.84, 7.6]]
MSE(-DR):[[0.0, 0.07, 0.91]][[1.31, 58.81, 4.57]]
***
==============
O_threshold = 110
MC for this TARGET:[60.853, 0.107]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.43, -2.51, -3.5]][[-5.41, -60.85, -6.61]]
std:[[1.14, 1.15, 0.65]][[0.59, 0.0, 0.35]]
MSE:[[2.68, 2.76, 3.56]][[5.44, 60.85, 6.62]]
MSE(-DR):[[0.0, 0.08, 0.88]][[2.76, 58.17, 3.94]]
***
==============
O_threshold = 115
MC for this TARGET:[62.677, 0.114]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.49, -6.48, -7.52]][[-13.03, -62.68, -8.43]]
std:[[1.04, 1.04, 0.69]][[0.5, 0.0, 0.35]]
MSE:[[6.57, 6.56, 7.55]][[13.04, 62.68, 8.44]]
MSE(-DR):[[0.0, -0.01, 0.98]][[6.47, 56.11, 1.87]]
***
==============
[[ 1.23  1.29  1.54  1.2  61.28  7.04]
 [ 3.03  3.1   3.94  4.34 61.84  7.6 ]
 [ 2.68  2.76  3.56  5.44 60.85  6.62]
 [ 6.57  6.56  7.55 13.04 62.68  8.44]]


time spent until now: 11.8 mins

23:35, 04/15

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

max(u_O) =  157.3
O_threshold = 100
means of Order:

89.6 98.6 46.6 141.0 55.2

79.0 112.6 68.9 73.6 77.3

113.8 157.3 101.0 72.1 113.5

85.1 99.5 129.4 81.3 100.2

78.0 96.1 106.4 75.3 91.5

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 = 115
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:51.765
O_threshold = 100
MC for this TARGET:[59.094, 0.112]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.55, -1.65, -2.15]][[-2.47, -59.09, -7.33]]
std:[[0.89, 0.9, 0.53]][[0.66, 0.0, 0.34]]
MSE:[[1.79, 1.88, 2.21]][[2.56, 59.09, 7.34]]
MSE(-DR):[[0.0, 0.09, 0.42]][[0.77, 57.3, 5.55]]
***
==============
O_threshold = 105
MC for this TARGET:[58.423, 0.1]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.21, -3.28, -3.98]][[-5.09, -58.42, -6.66]]
std:[[1.1, 1.1, 0.56]][[0.62, 0.0, 0.34]]
MSE:[[3.39, 3.46, 4.02]][[5.13, 58.42, 6.67]]
MSE(-DR):[[0.0, 0.07, 0.63]][[1.74, 55.03, 3.28]]
***
==============
O_threshold = 110
MC for this TARGET:[57.272, 0.094]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.82, -2.9, -3.74]][[-6.11, -57.27, -5.51]]
std:[[1.06, 1.07, 0.61]][[0.58, 0.0, 0.34]]
MSE:[[3.01, 3.09, 3.79]][[6.14, 57.27, 5.52]]
MSE(-DR):[[0.0, 0.08, 0.78]][[3.13, 54.26, 2.51]]
***
==============
O_threshold = 115
MC for this TARGET:[58.709, 0.101]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.56, -6.55, -7.92]][[-13.74, -58.71, -6.94]]
std:[[1.01, 1.01, 0.59]][[0.46, 0.0, 0.34]]
MSE:[[6.64, 6.63, 7.94]][[13.75, 58.71, 6.95]]
MSE(-DR):[[0.0, -0.01, 1.3]][[7.11, 52.07, 0.31]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.23  1.29  1.54  1.2  61.28  7.04]
 [ 3.03  3.1   3.94  4.34 61.84  7.6 ]
 [ 2.68  2.76  3.56  5.44 60.85  6.62]
 [ 6.57  6.56  7.55 13.04 62.68  8.44]]


[[ 1.79  1.88  2.21  2.56 59.09  7.34]
 [ 3.39  3.46  4.02  5.13 58.42  6.67]
 [ 3.01  3.09  3.79  6.14 57.27  5.52]
 [ 6.64  6.63  7.94 13.75 58.71  6.95]]


time spent until now: 23.8 mins

23:47, 04/15

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

max(u_O) =  157.3
O_threshold = 100
means of Order:

89.6 98.6 46.6 141.0 55.2
```

```
79.0 112.6 68.9 73.6 77.3

113.8 157.3 101.0 72.1 113.5

85.1 99.5 129.4 81.3 100.2

78.0 96.1 106.4 75.3 91.5

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 = 115
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:54.271
O_threshold = 100
MC for this TARGET:[61.251, 0.065]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.09, -0.22, -1.01]][[-0.77, -61.25, -6.98]]
std:[[0.55, 0.56, 0.44]][[0.3, 0.0, 0.25]]
MSE:[[0.56, 0.6, 1.1]][[0.83, 61.25, 6.98]]
MSE(-DR):[[0.0, 0.04, 0.54]][[0.27, 60.69, 6.42]]
***
==============
O_threshold = 105
MC for this TARGET:[61.816, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.62, -2.73, -3.5]][[-4.07, -61.82, -7.54]]
std:[[0.71, 0.72, 0.44]][[0.33, 0.0, 0.25]]
MSE:[[2.71, 2.82, 3.53]][[4.08, 61.82, 7.54]]
MSE(-DR):[[0.0, 0.11, 0.82]][[1.37, 59.11, 4.83]]
***
==============
O_threshold = 110
MC for this TARGET:[60.842, 0.063]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.49, -2.58, -3.18]][[-5.19, -60.84, -6.57]]
std:[[0.64, 0.65, 0.49]][[0.35, 0.0, 0.25]]
MSE:[[2.57, 2.66, 3.22]][[5.2, 60.84, 6.57]]
MSE(-DR):[[0.0, 0.09, 0.65]][[2.63, 58.27, 4.0]]
***
==============
O_threshold = 115
MC for this TARGET:[62.674, 0.058]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.31, -7.33, -7.43]][[-12.76, -62.67, -8.4]]
std:[[0.74, 0.79, 0.56]][[0.39, 0.0, 0.25]]
MSE:[[7.35, 7.37, 7.45]][[12.77, 62.67, 8.4]]
MSE(-DR):[[0.0, 0.02, 0.1]][[5.42, 55.32, 1.05]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.23  1.29  1.54  1.2  61.28  7.04]
 [ 3.03  3.1   3.94  4.34 61.84  7.6 ]
 [ 2.68  2.76  3.56  5.44 60.85  6.62]
 [ 6.57  6.56  7.55 13.04 62.68  8.44]]


[[ 1.79  1.88  2.21  2.56 59.09  7.34]
```

```
 [ 3.39  3.46  4.02  5.13 58.42  6.67]
 [ 3.01  3.09  3.79  6.14 57.27  5.52]
 [ 6.64  6.63  7.94 13.75 58.71  6.95]]


[[ 0.56  0.6   1.1   0.83 61.25  6.98]
 [ 2.71  2.82  3.53  4.08 61.82  7.54]
 [ 2.57  2.66  3.22  5.2  60.84  6.57]
 [ 7.35  7.37  7.45 12.77 62.67  8.4 ]]


time spent until now: 37.0 mins

00:00, 04/16

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

max(u_O) =  157.3
O_threshold = 100
means of Order:

89.6 98.6 46.6 141.0 55.2

79.0 112.6 68.9 73.6 77.3

113.8 157.3 101.0 72.1 113.5

85.1 99.5 129.4 81.3 100.2

78.0 96.1 106.4 75.3 91.5

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 = 115
number of reward locations:  3
```