```
Last login: Thu Apr 16 15:11:59 on ttys001
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-231-3-46.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 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

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


*** System restart required ***
Last login: Thu Apr 16 15:44:09 2020 from 107.13.161.147
ubuntu@ip-172-31-0-7:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
15:13, 04/16; num of cores:96
sd_u_O_25_sd

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, [15], None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[100, 100.5, 105, 110, 110.5, 111], [0, 15, 30], [7], [[0.0003, 0.0001, 5e-05], [
0.0003, 0.0001, 5e-05]]]

----------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 7, 0, 15]

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 = 100.5
number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 110.5
number of reward locations:  6
O_threshold = 111
number of reward locations:  6
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

----------------------------------------
Value of Behaviour policy:50.127
```

```
O_threshold = 100
MC for this TARGET:[57.544, 0.063]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.83, -2.03, -2.38]][[-3.17, -57.54, -7.42]]
std:[[0.47, 0.47, 0.4]][[0.28, 0.0, 0.24]]
MSE:[[1.89, 2.08, 2.41]][[3.18, 57.54, 7.42]]
MSE(-DR):[[0.0, 0.19, 0.52]][[1.29, 55.65, 5.53]]
***
==============
O_threshold = 100.5
MC for this TARGET:[54.712, 0.061]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.12, -0.3, -0.91]][[-2.27, -54.71, -4.59]]
std:[[0.46, 0.45, 0.41]][[0.28, 0.0, 0.24]]
MSE:[[0.48, 0.54, 1.0]][[2.29, 54.71, 4.6]]
MSE(-DR):[[0.0, 0.06, 0.52]][[1.81, 54.23, 4.12]]
***
==============
O_threshold = 105
MC for this TARGET:[56.425, 0.061]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.64, -3.82, -4.04]][[-5.6, -56.42, -6.3]]
std:[[0.64, 0.64, 0.5]][[0.32, 0.0, 0.24]]
MSE:[[3.7, 3.87, 4.07]][[5.61, 56.42, 6.3]]
MSE(-DR):[[0.0, 0.17, 0.37]][[1.91, 52.72, 2.6]]
***
==============
O_threshold = 110
MC for this TARGET:[55.129, 0.06]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.33, -3.45, -3.81]][[-6.43, -55.13, -5.0]]
std:[[0.65, 0.66, 0.53]][[0.33, 0.0, 0.24]]
MSE:[[3.39, 3.51, 3.85]][[6.44, 55.13, 5.01]]
MSE(-DR):[[0.0, 0.12, 0.46]][[3.05, 51.74, 1.62]]
***
==============
O_threshold = 110.5
MC for this TARGET:[55.129, 0.06]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.31, -3.45, -3.79]][[-6.43, -55.13, -5.0]]
std:[[0.66, 0.66, 0.48]][[0.33, 0.0, 0.24]]
MSE:[[3.38, 3.51, 3.82]][[6.44, 55.13, 5.01]]
MSE(-DR):[[0.0, 0.13, 0.44]][[3.06, 51.75, 1.63]]
***
==============
O_threshold = 111
MC for this TARGET:[55.129, 0.06]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.33, -3.45, -3.82]][[-6.43, -55.13, -5.0]]
std:[[0.66, 0.66, 0.52]][[0.33, 0.0, 0.24]]
MSE:[[3.39, 3.51, 3.86]][[6.44, 55.13, 5.01]]
MSE(-DR):[[0.0, 0.12, 0.47]][[3.05, 51.74, 1.62]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.89  2.08  2.41  3.18 57.54  7.42]
 [ 0.48  0.54  1.    2.29 54.71  4.6 ]
 [ 3.7   3.87  4.07  5.61 56.42  6.3 ]
 [ 3.39  3.51  3.85  6.44 55.13  5.01]
 [ 3.38  3.51  3.82  6.44 55.13  5.01]
 [ 3.39  3.51  3.86  6.44 55.13  5.01]]


time spent until now: 20.9 mins

15:34, 04/16

---------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 7, 15, 15]

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 = 100.5
number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 110.5
number of reward locations:  6
O_threshold = 111
number of reward locations:  6
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

--------------------------------------
Value of Behaviour policy:50.105
O_threshold = 100
MC for this TARGET:[57.568, 0.173]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.85, -2.07, -2.43]][[-3.23, -57.57, -7.46]]
std:[[0.69, 0.69, 0.49]][[0.51, 0.0, 0.22]]
MSE:[[1.97, 2.18, 2.48]][[3.27, 57.57, 7.46]]
MSE(-DR):[[0.0, 0.21, 0.51]][[1.3, 55.6, 5.49]]
***
==============
O_threshold = 100.5
MC for this TARGET:[54.736, 0.173]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.05, -0.26, -0.8]][[-2.36, -54.74, -4.63]]
std:[[0.73, 0.7, 0.36]][[0.43, 0.0, 0.22]]
MSE:[[0.73, 0.75, 0.88]][[2.4, 54.74, 4.64]]
MSE(-DR):[[0.0, 0.02, 0.15]][[1.67, 54.01, 3.91]]
***
==============
O_threshold = 105
MC for this TARGET:[56.449, 0.171]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.57, -3.75, -4.02]][[-5.69, -56.45, -6.34]]
std:[[0.98, 0.95, 0.46]][[0.44, 0.0, 0.22]]
MSE:[[3.7, 3.87, 4.05]][[5.71, 56.45, 6.34]]
MSE(-DR):[[0.0, 0.17, 0.35]][[2.01, 52.75, 2.64]]
***
==============
O_threshold = 110
MC for this TARGET:[55.153, 0.17]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.26, -3.38, -3.83]][[-6.52, -55.15, -5.05]]
std:[[1.21, 1.15, 0.62]][[0.42, 0.0, 0.22]]
MSE:[[3.48, 3.57, 3.88]][[6.53, 55.15, 5.05]]
MSE(-DR):[[0.0, 0.09, 0.4]][[3.05, 51.67, 1.57]]
***
==============
O_threshold = 110.5
MC for this TARGET:[55.153, 0.17]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.25, -3.38, -3.84]][[-6.53, -55.15, -5.05]]
std:[[1.22, 1.15, 0.6]][[0.42, 0.0, 0.22]]
MSE:[[3.47, 3.57, 3.89]][[6.54, 55.15, 5.05]]
MSE(-DR):[[0.0, 0.1, 0.42]][[3.07, 51.68, 1.58]]
***
==============
O_threshold = 111
MC for this TARGET:[55.153, 0.17]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.25, -3.38, -3.83]][[-6.52, -55.15, -5.05]]
std:[[1.22, 1.15, 0.62]][[0.43, 0.0, 0.22]]
MSE:[[3.47, 3.57, 3.88]][[6.53, 55.15, 5.05]]
MSE(-DR):[[0.0, 0.1, 0.41]][[3.06, 51.68, 1.58]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.89  2.08  2.41  3.18 57.54  7.42]
 [ 0.48  0.54  1.    2.29 54.71  4.6 ]
 [ 3.7   3.87  4.07  5.61 56.42  6.3 ]
 [ 3.39  3.51  3.85  6.44 55.13  5.01]
 [ 3.38  3.51  3.82  6.44 55.13  5.01]
 [ 3.39  3.51  3.86  6.44 55.13  5.01]]
```

```
[[ 1.97  2.18  2.48  3.27 57.57  7.46]
 [ 0.73  0.75  0.88  2.4  54.74  4.64]
 [ 3.7   3.87  4.05  5.71 56.45  6.34]
 [ 3.48  3.57  3.88  6.53 55.15  5.05]
 [ 3.47  3.57  3.89  6.54 55.15  5.05]
 [ 3.47  3.57  3.88  6.53 55.15  5.05]]


time spent until now: 41.6 mins

15:54, 04/16

--------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 7, 30, 15]

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 = 100.5
number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 110.5
number of reward locations:  6
O_threshold = 111
number of reward locations:  6
```