```
Last login: Thu Apr 16 10:01:09 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-34-201-49-219.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 14:00:12 2020 from 107.13.161.147
ubuntu@ip-172-31-0-227:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
10:27, 04/16; num of cores:96
sd_u_O_20_full_sigma

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, 20, 0.5, 1.5, [10, 0], None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[100, 100.5, 105, 110, 110.5, 111], [0, 15, 30, 45], [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, 10]

max(u_O) =  145.8
O_threshold = 100
means of Order:

91.7 98.9 57.3 132.8 64.1

83.2 110.1 75.1 78.8 81.8

111.0 145.8 100.8 77.6 110.8

88.1 99.6 123.5 85.0 100.2

82.4 96.9 105.1 80.2 93.2

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:  5
O_threshold = 111
number of reward locations:  4
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:52.865
```

```
O_threshold = 100
MC for this TARGET:[58.154, 0.081]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.36, -0.54, -1.08]][[-1.31, -58.15, -5.29]]
std:[[0.61, 0.6, 0.43]][[0.32, 0.0, 0.23]]
MSE:[[0.71, 0.81, 1.16]][[1.35, 58.15, 5.29]]
MSE(-DR):[[0.0, 0.1, 0.45]][[0.64, 57.44, 4.58]]
***
==============
O_threshold = 100.5
MC for this TARGET:[55.642, 0.072]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.1, 0.95, 0.22]][[-0.69, -55.64, -2.78]]
std:[[0.52, 0.48, 0.44]][[0.31, 0.0, 0.23]]
MSE:[[1.22, 1.06, 0.49]][[0.76, 55.64, 2.79]]
MSE(-DR):[[0.0, -0.16, -0.73]][[-0.46, 54.42, 1.57]]
==============
O_threshold = 105
MC for this TARGET:[57.708, 0.073]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.4, -2.55, -3.03]][[-4.12, -57.71, -4.84]]
std:[[0.68, 0.69, 0.4]][[0.33, 0.0, 0.23]]
MSE:[[2.49, 2.64, 3.06]][[4.13, 57.71, 4.85]]
MSE(-DR):[[0.0, 0.15, 0.57]][[1.64, 55.22, 2.36]]
***
==============
O_threshold = 110
MC for this TARGET:[56.697, 0.063]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.27, -2.36, -2.91]][[-5.24, -56.7, -3.83]]
std:[[0.59, 0.57, 0.44]][[0.33, 0.0, 0.23]]
MSE:[[2.35, 2.43, 2.94]][[5.25, 56.7, 3.84]]
MSE(-DR):[[0.0, 0.08, 0.59]][[2.9, 54.35, 1.49]]
***
==============
O_threshold = 110.5
MC for this TARGET:[59.437, 0.061]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.25, -5.35, -5.59]][[-9.49, -59.44, -6.57]]
std:[[0.68, 0.66, 0.52]][[0.32, 0.0, 0.23]]
MSE:[[5.29, 5.39, 5.61]][[9.5, 59.44, 6.57]]
MSE(-DR):[[0.0, 0.1, 0.32]][[4.21, 54.15, 1.28]]
***
==============
O_threshold = 111
MC for this TARGET:[57.578, 0.06]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.04, -5.11, -5.38]][[-9.85, -57.58, -4.71]]
std:[[0.69, 0.67, 0.53]][[0.31, 0.0, 0.23]]
MSE:[[5.09, 5.15, 5.41]][[9.85, 57.58, 4.72]]
MSE(-DR):[[0.0, 0.06, 0.32]][[4.76, 52.49, -0.37]]
***
==============
[[ 0.71  0.81  1.16  1.35 58.15  5.29]
 [ 1.22  1.06  0.49  0.76 55.64  2.79]
 [ 2.49  2.64  3.06  4.13 57.71  4.85]
 [ 2.35  2.43  2.94  5.25 56.7   3.84]
 [ 5.29  5.39  5.61  9.5  59.44  6.57]
 [ 5.09  5.15  5.41  9.85 57.58  4.72]]


time spent until now: 20.8 mins

10:48, 04/16

----------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 7, 15, 10]

max(u_O) =  145.8
O_threshold = 100
means of Order:

91.7 98.9 57.3 132.8 64.1

83.2 110.1 75.1 78.8 81.8

111.0 145.8 100.8 77.6 110.8

88.1 99.6 123.5 85.0 100.2

82.4 96.9 105.1 80.2 93.2

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:  5
O_threshold = 111
number of reward locations:  4
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:52.843
O_threshold = 100
MC for this TARGET:[58.178, 0.188]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.32, -0.49, -1.18]][[-1.35, -58.18, -5.33]]
std:[[0.7, 0.72, 0.56]][[0.31, 0.0, 0.21]]
MSE:[[0.77, 0.87, 1.31]][[1.39, 58.18, 5.33]]
MSE(-DR):[[0.0, 0.1, 0.54]][[0.62, 57.41, 4.56]]
***
==============
O_threshold = 100.5
MC for this TARGET:[55.666, 0.181]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.28, 1.12, 0.21]][[-0.74, -55.67, -2.82]]
std:[[0.68, 0.66, 0.49]][[0.32, 0.0, 0.21]]
MSE:[[1.45, 1.3, 0.53]][[0.81, 55.67, 2.83]]
MSE(-DR):[[0.0, -0.15, -0.92]][[-0.64, 54.22, 1.38]]
==============
O_threshold = 105
MC for this TARGET:[57.732, 0.182]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.25, -2.39, -3.2]][[-4.21, -57.73, -4.89]]
std:[[0.71, 0.69, 0.54]][[0.34, 0.0, 0.21]]
MSE:[[2.36, 2.49, 3.25]][[4.22, 57.73, 4.89]]
MSE(-DR):[[0.0, 0.13, 0.89]][[1.86, 55.37, 2.53]]
***
==============
O_threshold = 110
MC for this TARGET:[56.721, 0.18]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.29, -2.38, -3.02]][[-5.29, -56.72, -3.88]]
std:[[0.64, 0.6, 0.63]][[0.37, 0.0, 0.21]]
MSE:[[2.38, 2.45, 3.09]][[5.3, 56.72, 3.89]]
MSE(-DR):[[0.0, 0.07, 0.71]][[2.92, 54.34, 1.51]]
***
==============
O_threshold = 110.5
MC for this TARGET:[59.46, 0.18]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.46, -5.56, -5.65]][[-9.54, -59.46, -6.62]]
std:[[0.89, 0.8, 0.78]][[0.39, 0.0, 0.21]]
MSE:[[5.53, 5.62, 5.7]][[9.55, 59.46, 6.62]]
MSE(-DR):[[0.0, 0.09, 0.17]][[4.02, 53.93, 1.09]]
***
==============
O_threshold = 111
MC for this TARGET:[57.601, 0.181]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.08, -5.16, -5.34]][[-9.87, -57.6, -4.76]]
std:[[0.92, 0.91, 0.75]][[0.38, 0.0, 0.21]]
MSE:[[5.16, 5.24, 5.39]][[9.88, 57.6, 4.76]]
MSE(-DR):[[0.0, 0.08, 0.23]][[4.72, 52.44, -0.4]]
***
==============
[[ 0.71  0.81  1.16  1.35 58.15  5.29]
 [ 1.22  1.06  0.49  0.76 55.64  2.79]
 [ 2.49  2.64  3.06  4.13 57.71  4.85]
 [ 2.35  2.43  2.94  5.25 56.7   3.84]
 [ 5.29  5.39  5.61  9.5  59.44  6.57]
 [ 5.09  5.15  5.41  9.85 57.58  4.72]]


[[ 0.77  0.87  1.31  1.39 58.18  5.33]
 [ 1.45  1.3   0.53  0.81 55.67  2.83]
 [ 2.36  2.49  3.25  4.22 57.73  4.89]
```

```
 [ 2.38  2.45  3.09  5.3  56.72  3.89]
 [ 5.53  5.62  5.7   9.55 59.46  6.62]
 [ 5.16  5.24  5.39  9.88 57.6   4.76]]


time spent until now: 41.3 mins

11:08, 04/16

---------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 7, 30, 10]

max(u_O) =  145.8
O_threshold = 100
means of Order:

91.7 98.9 57.3 132.8 64.1

83.2 110.1 75.1 78.8 81.8

111.0 145.8 100.8 77.6 110.8

88.1 99.6 123.5 85.0 100.2

82.4 96.9 105.1 80.2 93.2

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:  5
O_threshold = 111
number of reward locations:  4
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:52.822
O_threshold = 100
MC for this TARGET:[58.202, 0.352]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.33, -0.51, -1.29]][[-1.4, -58.2, -5.38]]
std:[[1.22, 1.22, 0.89]][[0.55, 0.0, 0.3]]
MSE:[[1.26, 1.32, 1.57]][[1.5, 58.2, 5.39]]
MSE(-DR):[[0.0, 0.06, 0.31]][[0.24, 56.94, 4.13]]
***
==============
O_threshold = 100.5
MC for this TARGET:[55.689, 0.347]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.29, 1.17, 0.1]][[-0.8, -55.69, -2.87]]
std:[[1.22, 1.25, 0.71]][[0.54, 0.0, 0.3]]
MSE:[[1.78, 1.71, 0.72]][[0.97, 55.69, 2.89]]
MSE(-DR):[[0.0, -0.07, -1.06]][[-0.81, 53.91, 1.11]]
==============
O_threshold = 105
MC for this TARGET:[57.755, 0.348]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.19, -2.34, -3.31]][[-4.3, -57.76, -4.93]]
std:[[1.32, 1.3, 0.86]][[0.59, 0.0, 0.3]]
MSE:[[2.56, 2.68, 3.42]][[4.34, 57.76, 4.94]]
MSE(-DR):[[0.0, 0.12, 0.86]][[1.78, 55.2, 2.38]]
***
==============
O_threshold = 110
MC for this TARGET:[56.744, 0.347]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.34, -2.47, -3.07]][[-5.33, -56.74, -3.92]]
std:[[1.13, 1.12, 0.95]][[0.63, 0.0, 0.3]]
```

```
MSE:[[2.6, 2.71, 3.21]][[5.37, 56.74, 3.93]]
MSE(-DR):[[0.0, 0.11, 0.61]][[2.77, 54.14, 1.33]]
***
==============
O_threshold = 110.5
MC for this TARGET:[59.484, 0.347]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.67, -5.82, -5.7]][[-9.56, -59.48, -6.66]]
std:[[1.42, 1.29, 1.19]][[0.65, 0.0, 0.3]]
MSE:[[5.85, 5.96, 5.82]][[9.58, 59.48, 6.67]]
MSE(-DR):[[0.0, 0.11, -0.03]][[3.73, 53.63, 0.82]]
**
==============
O_threshold = 111
MC for this TARGET:[57.625, 0.349]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.14, -5.21, -5.36]][[-9.88, -57.62, -4.8]]
std:[[1.54, 1.5, 1.24]][[0.66, 0.0, 0.3]]
MSE:[[5.37, 5.42, 5.5]][[9.9, 57.62, 4.81]]
MSE(-DR):[[0.0, 0.05, 0.13]][[4.53, 52.25, -0.56]]
***
==============
[[ 0.71  0.81  1.16  1.35 58.15  5.29]
 [ 1.22  1.06  0.49  0.76 55.64  2.79]
 [ 2.49  2.64  3.06  4.13 57.71  4.85]
 [ 2.35  2.43  2.94  5.25 56.7   3.84]
 [ 5.29  5.39  5.61  9.5  59.44  6.57]
 [ 5.09  5.15  5.41  9.85 57.58  4.72]]


[[ 0.77  0.87  1.31  1.39 58.18  5.33]
 [ 1.45  1.3   0.53  0.81 55.67  2.83]
 [ 2.36  2.49  3.25  4.22 57.73  4.89]
 [ 2.38  2.45  3.09  5.3  56.72  3.89]
 [ 5.53  5.62  5.7   9.55 59.46  6.62]
 [ 5.16  5.24  5.39  9.88 57.6   4.76]]


[[ 1.26  1.32  1.57  1.5  58.2   5.39]
 [ 1.78  1.71  0.72  0.97 55.69  2.89]
 [ 2.56  2.68  3.42  4.34 57.76  4.94]
 [ 2.6   2.71  3.21  5.37 56.74  3.93]
 [ 5.85  5.96  5.82  9.58 59.48  6.67]
 [ 5.37  5.42  5.5   9.9  57.62  4.81]]


time spent until now: 61.8 mins

11:29, 04/16

--------------------------------------
[pattern_seed, day, sd_R, u_O_u_D] = [2, 7, 45, 10]

max(u_O) =  145.8
O_threshold = 100
means of Order:

91.7 98.9 57.3 132.8 64.1

83.2 110.1 75.1 78.8 81.8

111.0 145.8 100.8 77.6 110.8

88.1 99.6 123.5 85.0 100.2

82.4 96.9 105.1 80.2 93.2

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:  5
O_threshold = 111
```

```
number of reward locations:  4
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:52.8
O_threshold = 100
MC for this TARGET:[58.225, 0.522]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.42, -0.53, -1.41]][[-1.44, -58.22, -5.42]]
std:[[1.71, 1.72, 1.32]][[0.84, 0.0, 0.45]]
MSE:[[1.76, 1.8, 1.93]][[1.67, 58.22, 5.44]]
MSE(-DR):[[0.0, 0.04, 0.17]][[-0.09, 56.46, 3.68]]
==============
O_threshold = 100.5
MC for this TARGET:[55.713, 0.518]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.4, 1.25, 0.12]][[-0.87, -55.71, -2.91]]
std:[[1.86, 1.82, 1.1]][[0.83, 0.0, 0.45]]
MSE:[[2.33, 2.21, 1.11]][[1.2, 55.71, 2.94]]
MSE(-DR):[[0.0, -0.12, -1.22]][[-1.13, 53.38, 0.61]]
==============
O_threshold = 105
MC for this TARGET:[57.779, 0.518]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.09, -2.32, -3.37]][[-4.4, -57.78, -4.98]]
std:[[1.98, 1.92, 1.28]][[0.89, 0.0, 0.45]]
MSE:[[2.88, 3.01, 3.6]][[4.49, 57.78, 5.0]]
MSE(-DR):[[0.0, 0.13, 0.72]][[1.61, 54.9, 2.12]]
***
==============
O_threshold = 110
MC for this TARGET:[56.768, 0.519]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.51, -2.61, -3.18]][[-5.39, -56.77, -3.97]]
std:[[1.67, 1.66, 1.3]][[0.95, 0.0, 0.45]]
MSE:[[3.01, 3.09, 3.44]][[5.47, 56.77, 4.0]]
MSE(-DR):[[0.0, 0.08, 0.43]][[2.46, 53.76, 0.99]]
***
==============
O_threshold = 110.5
MC for this TARGET:[59.508, 0.519]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.9, -6.09, -5.75]][[-9.61, -59.51, -6.71]]
std:[[1.86, 1.8, 1.63]][[0.96, 0.0, 0.45]]
MSE:[[6.19, 6.35, 5.98]][[9.66, 59.51, 6.73]]
MSE(-DR):[[0.0, 0.16, -0.21]][[3.47, 53.32, 0.54]]
**
==============
O_threshold = 111
MC for this TARGET:[57.649, 0.521]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.12, -5.25, -5.36]][[-9.88, -57.65, -4.85]]
std:[[2.06, 2.01, 1.79]][[1.01, 0.0, 0.45]]
MSE:[[5.52, 5.62, 5.65]][[9.93, 57.65, 4.87]]
MSE(-DR):[[0.0, 0.1, 0.13]][[4.41, 52.13, -0.65]]
***
==============
[[ 0.71  0.81  1.16  1.35 58.15  5.29]
 [ 1.22  1.06  0.49  0.76 55.64  2.79]
 [ 2.49  2.64  3.06  4.13 57.71  4.85]
 [ 2.35  2.43  2.94  5.25 56.7   3.84]
 [ 5.29  5.39  5.61  9.5  59.44  6.57]
 [ 5.09  5.15  5.41  9.85 57.58  4.72]]


[[ 0.77  0.87  1.31  1.39 58.18  5.33]
 [ 1.45  1.3   0.53  0.81 55.67  2.83]
 [ 2.36  2.49  3.25  4.22 57.73  4.89]
 [ 2.38  2.45  3.09  5.3  56.72  3.89]
 [ 5.53  5.62  5.7   9.55 59.46  6.62]
 [ 5.16  5.24  5.39  9.88 57.6   4.76]]


[[ 1.26  1.32  1.57  1.5  58.2   5.39]
 [ 1.78  1.71  0.72  0.97 55.69  2.89]
 [ 2.56  2.68  3.42  4.34 57.76  4.94]
 [ 2.6   2.71  3.21  5.37 56.74  3.93]
 [ 5.85  5.96  5.82  9.58 59.48  6.67]
 [ 5.37  5.42  5.5   9.9  57.62  4.81]]


[[ 1.76  1.8   1.93  1.67 58.22  5.44]
 [ 2.33  2.21  1.11  1.2  55.71  2.94]
 [ 2.88  3.01  3.6   4.49 57.78  5.  ]
```

```
 [ 3.01  3.09  3.44  5.47 56.77  4.  ]
 [ 6.19  6.35  5.98  9.66 59.51  6.73]
 [ 5.52  5.62  5.65  9.93 57.65  4.87]]


time spent until now: 82.6 mins

11:50, 04/16

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

max(u_O) =  145.8
O_threshold = 100
means of Order:

91.7 98.9 57.3 132.8 64.1

83.2 110.1 75.1 78.8 81.8

111.0 145.8 100.8 77.6 110.8

88.1 99.6 123.5 85.0 100.2

82.4 96.9 105.1 80.2 93.2

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:  5
O_threshold = 111
number of reward locations:  4
```