```
Last login: Wed Apr 15 23:23:40 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 03:23:42 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
00:04, 04/16; num of cores:72
sd_u_0_25_uo_ud_10_20_sep_4

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, [10, 20], None]

[thre_range, sd_R_range, day_range, penalty_range):  [[101, 105, 109, 113], [0, 20, 40], [3, 7], [[0.0001, 5e-05], [0.0001, 5e-05]]]

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

max(u_O) =  157.3
O_threshold = 101
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 0

0 0 1 0 0

number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 109
number of reward locations:  6
O_threshold = 113
number of reward locations:  5
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 = 101
MC for this TARGET:[56.392, 0.104]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.39, 0.3, -0.63]][[-1.71, -56.39, -4.63]]
std:[[0.87, 0.88, 0.53]][[0.61, 0.0, 0.34]]
MSE:[[0.95, 0.93, 0.82]][[1.82, 56.39, 4.64]]
MSE(-DR):[[0.0, -0.02, -0.13]][[0.87, 55.44, 3.69]]
**
=============
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, -3.96]][[-5.09, -58.42, -6.66]]
std:[[1.09, 1.1, 0.54]][[0.61, 0.0, 0.34]]
MSE:[[3.39, 3.46, 4.0]][[5.13, 58.42, 6.67]]
MSE(-DR):[[0.0, 0.07, 0.61]][[1.74, 55.03, 3.28]]
***
=============
O_threshold = 109
MC for this TARGET:[57.272, 0.094]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.83, -2.9, -3.77]][[-6.11, -57.27, -5.51]]
std:[[1.04, 1.07, 0.61]][[0.58, 0.0, 0.34]]
MSE:[[3.02, 3.09, 3.82]][[6.14, 57.27, 5.52]]
MSE(-DR):[[0.0, 0.07, 0.8]][[3.12, 54.25, 2.5]]
***
=============
O_threshold = 113
MC for this TARGET:[60.076, 0.109]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.9, -5.94, -6.37]][[-10.59, -60.08, -8.31]]
std:[[1.26, 1.3, 0.64]][[0.57, 0.0, 0.34]]
MSE:[[6.03, 6.08, 6.4]][[10.61, 60.08, 8.32]]
MSE(-DR):[[0.0, 0.05, 0.37]][[4.58, 54.05, 2.29]]
***
=============
[[ 0.95  0.93  0.82  1.82 56.39  4.64]
 [ 3.39  3.46  4.    5.13 58.42  6.67]
 [ 3.02  3.09  3.82  6.14 57.27  5.52]
 [ 6.03  6.08  6.4  10.61 60.08  8.32]]


time spent until now: 12.0 mins

00:16, 04/16

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

max(u_O) =  157.3
O_threshold = 101
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 0

0 0 1 0 0

number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 109
number of reward locations:  6
O_threshold = 113
number of reward locations:  5
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

----------------------------------------
Value of Behaviour policy:48.185
O_threshold = 101
MC for this TARGET:[52.876, 0.1]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.4, -1.49, -1.99]][[-3.56, -52.88, -4.69]]
std:[[0.83, 0.84, 0.47]][[0.55, 0.0, 0.3]]
MSE:[[1.63, 1.71, 2.04]][[3.6, 52.88, 4.7]]
MSE(-DR):[[0.0, 0.08, 0.41]][[1.97, 51.25, 3.07]]
***
=============
O_threshold = 105
```

```
MC for this TARGET:[54.336, 0.099]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.66, -4.7, -4.86]][[-6.8, -54.34, -6.15]]
std:[[0.98, 0.99, 0.53]][[0.54, 0.0, 0.3]]
MSE:[[4.76, 4.8, 4.89]][[6.82, 54.34, 6.16]]
MSE(-DR):[[0.0, 0.04, 0.13]][[2.06, 49.58, 1.4]]
***
=============
O_threshold = 109
MC for this TARGET:[52.824, 0.09]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.83, -3.88, -4.46]][[-7.31, -52.82, -4.64]]
std:[[0.9, 0.91, 0.49]][[0.49, 0.0, 0.3]]
MSE:[[3.93, 3.99, 4.49]][[7.33, 52.82, 4.65]]
MSE(-DR):[[0.0, 0.06, 0.56]][[3.4, 48.89, 0.72]]
***
=============
O_threshold = 113
MC for this TARGET:[55.108, 0.093]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.38, -6.38, -6.5]][[-11.68, -55.11, -6.92]]
std:[[1.28, 1.3, 0.53]][[0.49, 0.0, 0.3]]
MSE:[[6.51, 6.51, 6.52]][[11.69, 55.11, 6.93]]
MSE(-DR):[[0.0, 0.0, 0.01]][[5.18, 48.6, 0.42]]
***
=============
******************** THIS SETTING IS GOOD ********************
[[ 0.95  0.93  0.82  1.82 56.39  4.64]
 [ 3.39  3.46  4.    5.13 58.42  6.67]
 [ 3.02  3.09  3.82  6.14 57.27  5.52]
 [ 6.03  6.08  6.4  10.61 60.08  8.32]]


[[ 1.63  1.71  2.04  3.6  52.88  4.7 ]
 [ 4.76  4.8   4.89  6.82 54.34  6.16]
 [ 3.93  3.99  4.49  7.33 52.82  4.65]
 [ 6.51  6.51  6.52 11.69 55.11  6.93]]


time spent until now: 24.0 mins

00:28, 04/16

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

max(u_O) =  157.3
O_threshold = 101
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 0

0 0 1 0 0

number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 109
number of reward locations:  6
O_threshold = 113
number of reward locations:  5
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!

----------------------------------------
Value of Behaviour policy:51.771
O_threshold = 101
MC for this TARGET:[56.374, 0.06]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.68, 0.57, -0.28]][[-1.46, -56.37, -4.6]]
std:[[0.54, 0.53, 0.45]][[0.3, 0.0, 0.25]]
MSE:[[0.87, 0.78, 0.53]][[1.49, 56.37, 4.61]]
MSE(-DR):[[0.0, -0.09, -0.34]][[0.62, 55.5, 3.74]]
**
=============
O_threshold = 105
MC for this TARGET:[58.402, 0.061]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.91, -3.01, -3.66]][[-4.88, -58.4, -6.63]]
std:[[0.71, 0.72, 0.5]][[0.34, 0.0, 0.25]]
MSE:[[3.0, 3.09, 3.69]][[4.89, 58.4, 6.63]]
MSE(-DR):[[0.0, 0.09, 0.69]][[1.89, 55.4, 3.63]]
***
=============
O_threshold = 109
MC for this TARGET:[57.263, 0.06]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.78, -2.85, -3.48]][[-5.91, -57.26, -5.49]]
std:[[0.71, 0.71, 0.54]][[0.35, 0.0, 0.25]]
MSE:[[2.87, 2.94, 3.52]][[5.92, 57.26, 5.5]]
MSE(-DR):[[0.0, 0.07, 0.65]][[3.05, 54.39, 2.63]]
***
=============
O_threshold = 113
MC for this TARGET:[60.069, 0.063]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.92, -5.97, -6.24]][[-10.39, -60.07, -8.3]]
std:[[0.81, 0.83, 0.56]][[0.33, 0.0, 0.25]]
MSE:[[5.98, 6.03, 6.27]][[10.4, 60.07, 8.3]]
MSE(-DR):[[0.0, 0.05, 0.29]][[4.42, 54.09, 2.32]]
***
=============
[[ 0.95  0.93  0.82  1.82 56.39  4.64]
 [ 3.39  3.46  4.    5.13 58.42  6.67]
 [ 3.02  3.09  3.82  6.14 57.27  5.52]
 [ 6.03  6.08  6.4  10.61 60.08  8.32]]


[[ 1.63  1.71  2.04  3.6  52.88  4.7 ]
 [ 4.76  4.8   4.89  6.82 54.34  6.16]
 [ 3.93  3.99  4.49  7.33 52.82  4.65]
 [ 6.51  6.51  6.52 11.69 55.11  6.93]]


[[ 0.87  0.78  0.53  1.49 56.37  4.61]
 [ 3.    3.09  3.69  4.89 58.4   6.63]
 [ 2.87  2.94  3.52  5.92 57.26  5.5 ]
 [ 5.98  6.03  6.27 10.4  60.07  8.3 ]]


time spent until now: 37.4 mins

00:41, 04/16

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

max(u_O) =  157.3
O_threshold = 101
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 0

0 0 1 0 0

number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 109
number of reward locations:  6
O_threshold = 113
number of reward locations:  5
target 1 in 1 DONE!
```

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

----------------------------------------
Value of Behaviour policy:48.158
O_threshold = 101
MC for this TARGET:[52.866, 0.061]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.1, -1.22, -1.83]][[-3.35, -52.87, -4.71]]
std:[[0.51, 0.5, 0.39]]][[0.26, 0.0, 0.21]]
MSE:[[1.21, 1.32, 1.87]]][[3.36, 52.87, 4.71]]
MSE(-DR):[[0.0, 0.11, 0.66]]][[2.15, 51.66, 3.5]]
***
=============
O_threshold = 105
MC for this TARGET:[54.317, 0.063]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.34, -4.42, -4.71]][[-6.64, -54.32, -6.16]]
std:[[0.63, 0.63, 0.44]]][[0.29, 0.0, 0.21]]
MSE:[[4.39, 4.46, 4.73]]][[6.65, 54.32, 6.16]]
MSE(-DR):[[0.0, 0.07, 0.34]]][[2.26, 49.93, 1.77]]
***
=============
O_threshold = 109
MC for this TARGET:[52.815, 0.061]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.86, -3.93, -4.34]][[-7.17, -52.82, -4.66]]
std:[[0.68, 0.68, 0.47]]][[0.29, 0.0, 0.21]]
MSE:[[3.92, 3.99, 4.37]]][[7.18, 52.82, 4.66]]
MSE(-DR):[[0.0, 0.07, 0.45]]][[3.26, 48.9, 0.74]]
***
=============
O_threshold = 113
MC for this TARGET:[55.104, 0.057]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.57, -6.62, -6.52]]][[-11.54, -55.1, -6.95]]
std:[[0.67, 0.68, 0.52]]][[0.26, 0.0, 0.21]]
MSE:[[6.6, 6.65, 6.54]]][[11.54, 55.1, 6.95]]
MSE(-DR):[[0.0, 0.05, -0.06]]][[4.94, 48.5, 0.35]]
**
=============
[[ 0.95  0.93  0.82  1.82 56.39  4.64]
 [ 3.39  3.46  4.    5.13 58.42  6.67]
 [ 3.02  3.09  3.82  6.14 57.27  5.52]
 [ 6.03  6.08  6.4  10.61 60.08  8.32]]


[[ 1.63  1.71  2.04  3.6  52.88  4.7 ]
 [ 4.76  4.8   4.89  6.82 54.34  6.16]
 [ 3.93  3.99  4.49  7.33 52.82  4.65]
 [ 6.51  6.51  6.52 11.69 55.11  6.93]]


[[ 0.87  0.78  0.53  1.49 56.37  4.61]
 [ 3.    3.09  3.69  4.89 58.4   6.63]
 [ 2.87  2.94  3.52  5.92 57.26  5.5 ]
 [ 5.98  6.03  6.27 10.4  60.07  8.3 ]]


[[ 1.21  1.32  1.87  3.36 52.87  4.71]
 [ 4.39  4.46  4.73  6.65 54.32  6.16]
 [ 3.92  3.99  4.37  7.18 52.82  4.66]
 [ 6.6   6.65  6.54 11.54 55.1   6.95]]


time spent until now: 50.9 mins

00:55, 04/16

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

max(u_O) =  157.3
O_threshold = 101
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 0

0 0 1 0 0

number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 109
number of reward locations:  6
O_threshold = 113
number of reward locations:  5
```