```
Last login: Mon Apr 13 14:17:25 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-235-59-105.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: Mon Apr 13 18:17:32 2020 from 107.13.161.147
ubuntu@ip-172-31-6-10:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
15:34, 04/13; num of cores:96
sd_u_O_30

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 40, 0.5, 1.5, 10, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[85, 95, 100, 110, 120, 125], [0, 20, 40], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]
]

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

max(u_O) =  191.7
O_threshold = 85
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  13
O_threshold = 95
number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
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!
target 1 in 1 DONE!
target 1 in 1 DONE!

--------------------------------------
Value of Behaviour policy:47.461
```

O_threshold = 85
MC for this TARGET:[61.672, 0.065]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.09, -0.41, -1.99]][[-0.53, -61.67, -14.21]]
std:[[0.42, 0.42, 0.29]][[0.32, 0.0, 0.26]]
MSE:[[0.43, 0.59, 2.01]][[0.62, 61.67, 14.21]]
MSE(-DR):[[0.0, 0.16, 1.58]][[0.19, 61.24, 13.78]]
***
==============
O_threshold = 95
MC for this TARGET:[59.924, 0.062]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.19, -1.44, -2.35]][[-2.05, -59.92, -12.46]]
std:[[0.42, 0.42, 0.3]][[0.33, 0.0, 0.26]]
MSE:[[1.26, 1.5, 2.37]][[2.08, 59.92, 12.46]]
MSE(-DR):[[0.0, 0.24, 1.11]][[0.82, 58.66, 11.2]]
***
==============
O_threshold = 100
MC for this TARGET:[60.269, 0.065]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.67, -3.86, -4.55]][[-5.41, -60.27, -12.81]]
std:[[0.58, 0.58, 0.38]][[0.35, 0.0, 0.26]]
MSE:[[3.72, 3.9, 4.57]][[5.42, 60.27, 12.81]]
MSE(-DR):[[0.0, 0.18, 0.85]][[1.7, 56.55, 9.09]]
***
==============
O_threshold = 110
MC for this TARGET:[58.929, 0.073]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.14, -5.28, -6.01]][[-7.61, -58.93, -11.47]]
std:[[0.62, 0.64, 0.33]][[0.3, 0.0, 0.26]]
MSE:[[5.18, 5.32, 6.02]][[7.62, 58.93, 11.47]]
MSE(-DR):[[0.0, 0.14, 0.84]][[2.44, 53.75, 6.29]]
***
==============
O_threshold = 120
MC for this TARGET:[57.176, 0.067]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.78, -4.92, -5.49]][[-8.03, -57.18, -9.72]]
std:[[0.69, 0.69, 0.41]][[0.31, 0.0, 0.26]]
MSE:[[4.83, 4.97, 5.51]][[8.04, 57.18, 9.72]]
MSE(-DR):[[0.0, 0.14, 0.68]][[3.21, 52.35, 4.89]]
***
==============
O_threshold = 125
MC for this TARGET:[56.866, 0.067]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.49, -8.51, -8.56]][[-14.33, -56.87, -9.41]]
std:[[1.06, 1.09, 0.37]][[0.32, 0.0, 0.26]]
MSE:[[8.56, 8.58, 8.57]][[14.33, 56.87, 9.41]]
MSE(-DR):[[0.0, 0.02, 0.01]][[5.77, 48.31, 0.85]]
***
==============
****************** THIS SETTING IS GOOD ******************
[[0.43 0.59 2.01 0.62 61.67 14.21]
 [1.26 1.5   2.37 2.08 59.92 12.46]
 [3.72 3.9   4.57 5.42 60.27 12.81]
 [5.18 5.32 6.02 7.62 58.93 11.47]
 [4.83 4.97 5.51 8.04 57.18 9.72]
 [8.56 8.58 8.57 14.33 56.87 9.41]]

time spent until now: 17.0 mins

15:51, 04/13

[pattern_seed, day, sd_R] = [2, 7, 20]

max(u_O) = 191.7
O_threshold = 85
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

```
1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  13
O_threshold = 95
number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
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!
target 1 in 1 DONE!
target 1 in 1 DONE!


----------------------------------------
Value of Behaviour policy:47.432
O_threshold = 85
MC for this TARGET:[61.704, 0.235]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.03, -0.4, -1.82]][[-0.72, -61.7, -14.27]]
std:[[1.27, 1.27, 0.8]][[0.68, 0.0, 0.27]]
MSE:[[1.27, 1.33, 1.99]][[0.99, 61.7, 14.27]]
MSE(-DR):[[0.0, 0.06, 0.72]][[-0.28, 60.43, 13.0]]
==============
O_threshold = 95
MC for this TARGET:[59.956, 0.237]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.12, -1.38, -2.31]][[-2.21, -59.96, -12.52]]
std:[[0.98, 0.98, 0.6]][[0.68, 0.0, 0.27]]
MSE:[[1.49, 1.69, 2.39]][[2.31, 59.96, 12.52]]
MSE(-DR):[[0.0, 0.2, 0.9]][[0.82, 58.47, 11.03]]
***
==============
O_threshold = 100
MC for this TARGET:[60.301, 0.237]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.43, -3.64, -4.46]][[-5.51, -60.3, -12.87]]
std:[[1.08, 1.07, 0.51]][[0.65, 0.0, 0.27]]
MSE:[[3.6, 3.79, 4.49]][[5.55, 60.3, 12.87]]
MSE(-DR):[[0.0, 0.19, 0.89]][[1.95, 56.7, 9.27]]
***
==============
O_threshold = 110
MC for this TARGET:[58.961, 0.239]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.21, -5.4, -5.92]][[-7.63, -58.96, -11.53]]
std:[[1.18, 1.19, 0.61]][[0.54, 0.0, 0.27]]
MSE:[[5.34, 5.53, 5.95]][[7.65, 58.96, 11.53]]
MSE(-DR):[[0.0, 0.19, 0.61]][[2.31, 53.62, 6.19]]
***
==============
O_threshold = 120
MC for this TARGET:[57.208, 0.238]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.92, -5.05, -5.55]][[-8.14, -57.21, -9.78]]
std:[[1.24, 1.25, 0.72]][[0.49, 0.0, 0.27]]
MSE:[[5.07, 5.2, 5.6]][[8.15, 57.21, 9.78]]
MSE(-DR):[[0.0, 0.13, 0.53]][[3.08, 52.14, 4.71]]
***
==============
O_threshold = 125
MC for this TARGET:[56.898, 0.238]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.08, -9.05, -8.84]][[-14.44, -56.9, -9.47]]
std:[[1.58, 1.69, 0.65]][[0.56, 0.0, 0.27]]
MSE:[[9.22, 9.21, 8.86]][[14.45, 56.9, 9.47]]
MSE(-DR):[[0.0, -0.01, -0.36]][[5.23, 47.68, 0.25]]
**
==============
[[ 0.43  0.59  2.01  0.62 61.67 14.21]
 [ 1.26  1.5   2.37  2.08 59.92 12.46]
 [ 3.72  3.9   4.57  5.42 60.27 12.81]
 [ 5.18  5.32  6.02  7.62 58.93 11.47]
 [ 4.83  4.97  5.51  8.04 57.18  9.72]
 [ 8.56  8.58  8.57 14.33 56.87  9.41]]


[[ 1.27  1.33  1.99  0.99 61.7  14.27]
```

```
 [ 1.49  1.69  2.39  2.31 59.96 12.52]
 [ 3.6   3.79  4.49  5.55 60.3  12.87]
 [ 5.34  5.53  5.95  7.65 58.96 11.53]
 [ 5.07  5.2   5.6   8.15 57.21  9.78]
 [ 9.22  9.21  8.86 14.45 56.9   9.47]]
```

time spent until now: 34.1 mins

max(u_0) = 191.7
O_threshold = 85
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations: 13
O_threshold = 95
number of reward locations: 11
O_threshold = 100
number of reward locations: 9
O_threshold = 110
number of reward locations: 7
O_threshold = 120
number of reward locations: 6
O_threshold = 125
number of reward locations: 3