```
Last login: Wed Apr  1 14:58:34 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-3-227-211-214.compute-1.amazonaws.com
Warning: Permanently added the ED25519 host key for IP address '3.227.211.214' to the list of known hosts.
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 as of Wed Apr  1 20:30:39 UTC 2020

  System load:  1.17               Processes:           399
  Usage of /:   55.4% of 15.45GB   Users logged in:     0
  Memory usage: 0%                 IP address for ens5: 172.31.79.27
  Swap usage:   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

93 packages can be updated.
43 updates are security updates.


Last login: Thu Mar  5 21:23:34 2020 from 107.13.161.147
ubuntu@ip-172-31-79-27:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
16:32, 04/01; num of cores:36

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, pois0, simple, u_O_u_D]] = [672, 10, 10, 10, 0.3, 0.5, 1,
 [True, False, True, False, 10]]


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

max(u_O) =  197.9
O_threshold = 70
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

1 1 0 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

number of reward locations:  22
O_threshold = 80
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
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 0

number of reward locations:  12
O_threshold = 100
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
Rep 1 DONE Rep 2 DONE Rep 3 DONE Rep 4 DONE Rep 5 DONE Rep 6 DONE Rep 7 DONE Rep 8 DONE
----------------------------------------
Value of Behaviour policy:60.797
O_threshold = 70
MC for this TARGET:[69.323, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.28, 2.14, 1.38]][[5.09, -69.32, -8.53]]
std:[[1.11, 1.11, 0.19]][[0.22, 0.0, 0.15]]
MSE:[[2.54, 2.41, 1.39]][[5.09, 69.32, 8.53]]
MSE(-DR):[[0.0, -0.13, -1.15]][[2.55, 66.78, 5.99]]
**
==============


O_threshold = 80
MC for this TARGET:[70.884, 0.141]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.71, 0.54, -0.96]][[1.26, -70.88, -10.09]]
std:[[0.59, 0.58, 0.42]][[0.31, 0.0, 0.15]]
MSE:[[0.92, 0.79, 1.05]][[1.3, 70.88, 10.09]]
MSE(-DR):[[0.0, -0.13, 0.13]][[0.38, 69.96, 9.17]]
***
==============


O_threshold = 90
MC for this TARGET:[69.371, 0.133]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.15, -0.32, -1.11]][[-0.51, -69.37, -8.57]]
std:[[0.7, 0.67, 0.63]][[0.31, 0.0, 0.15]]
MSE:[[0.72, 0.74, 1.28]][[0.6, 69.37, 8.57]]
MSE(-DR):[[0.0, 0.02, 0.56]][[-0.12, 68.65, 7.85]]
==============


O_threshold = 100
MC for this TARGET:[68.94, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.7, -2.83, -3.42]][[-4.84, -68.94, -8.14]]
std:[[0.48, 0.46, 0.33]][[0.3, 0.0, 0.15]]
MSE:[[2.74, 2.87, 3.44]][[4.85, 68.94, 8.14]]
MSE(-DR):[[0.0, 0.13, 0.7]][[2.11, 66.2, 5.4]]
***
==============


[[ 2.54  2.41  1.39  5.09 69.32  8.53]
 [ 0.92  0.79  1.05  1.3  70.88 10.09]
 [ 0.72  0.74  1.28  0.6  69.37  8.57]
 [ 2.74  2.87  3.44  4.85 68.94  8.14]]


time spent until now: 16.1 mins


----------------------------------------
[pattern_seed, day] = [2, 9]

max(u_O) =  197.9
O_threshold = 70
means of Order:

87.8 97.8 52.4 162.7 58.1
```

```
77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

1 1 0 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

number of reward locations:  22
O_threshold = 80
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
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 0

number of reward locations:  12
O_threshold = 100
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
Rep 1 DONE Rep 2 DONE Rep 3 DONE Rep 4 DONE Rep 5 DONE Rep 6 DONE Rep 7 DONE Rep 8 DONE
----------------------------------------
Value of Behaviour policy:60.741
O_threshold = 70
MC for this TARGET:[69.306, 0.104]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.8, 2.66, 1.61]][[5.2, -69.31, -8.56]]
std:[[0.97, 0.98, 0.49]][[0.22, 0.0, 0.1]]
MSE:[[2.96, 2.83, 1.68]][[5.2, 69.31, 8.56]]
MSE(-DR):[[0.0, -0.13, -1.28]][[2.24, 66.35, 5.6]]
**
==============


O_threshold = 80
MC for this TARGET:[70.872, 0.103]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.18, 0.01, -0.99]][[1.25, -70.87, -10.13]]
std:[[0.85, 0.83, 0.44]][[0.21, 0.0, 0.1]]
MSE:[[0.87, 0.83, 1.08]][[1.27, 70.87, 10.13]]
MSE(-DR):[[0.0, -0.04, 0.21]][[0.4, 70.0, 9.26]]
***
==============
```

```
O_threshold = 90
MC for this TARGET:[69.356, 0.108]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.25, -0.41, -1.11]][[-0.5, -69.36, -8.61]]
std:[[0.26, 0.26, 0.28]][[0.26, 0.0, 0.1]]
MSE:[[0.36, 0.49, 1.14]][[0.56, 69.36, 8.61]]
MSE(-DR):[[0.0, 0.13, 0.78]][[0.2, 69.0, 8.25]]
***
==============


O_threshold = 100
MC for this TARGET:[68.927, 0.115]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.9, -3.01, -3.26]][[-4.82, -68.93, -8.19]]
std:[[0.4, 0.4, 0.2]][[0.27, 0.0, 0.1]]
MSE:[[2.93, 3.04, 3.27]][[4.83, 68.93, 8.19]]
MSE(-DR):[[0.0, 0.11, 0.34]][[1.9, 66.0, 5.26]]
***
==============


[[ 2.54  2.41  1.39  5.09 69.32  8.53]
 [ 0.92  0.79  1.05  1.3  70.88 10.09]
 [ 0.72  0.74  1.28  0.6  69.37  8.57]
 [ 2.74  2.87  3.44  4.85 68.94  8.14]]


[[ 2.96  2.83  1.68  5.2  69.31  8.56]
 [ 0.87  0.83  1.08  1.27 70.87 10.13]
 [ 0.36  0.49  1.14  0.56 69.36  8.61]
 [ 2.93  3.04  3.27  4.83 68.93  8.19]]


time spent until now: 32.2 mins


----------------------------------------
[pattern_seed, day] = [2, 11]

max(u_O) =  197.9
O_threshold = 70
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

1 1 0 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

number of reward locations:  22
O_threshold = 80
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
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 0

number of reward locations:  12
O_threshold = 100
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
Rep 1 DONE Rep 2 DONE Rep 3 DONE Rep 4 DONE Rep 5 DONE Rep 6 DONE Rep 7 DONE Rep 8 DONE
--------------------------------------
Value of Behaviour policy:60.692
O_threshold = 70
MC for this TARGET:[69.325, 0.091]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[3.29, 3.19, 1.5]][[5.23, -69.32, -8.63]]
std:[[0.5, 0.51, 0.4]][[0.21, 0.0, 0.1]]
MSE:[[3.33, 3.23, 1.55]][[5.23, 69.32, 8.63]]
MSE(-DR):[[0.0, -0.1, -1.78]][[1.9, 65.99, 5.3]]
**
==============


O_threshold = 80
MC for this TARGET:[70.892, 0.092]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.68, 0.51, -0.91]][[1.45, -70.89, -10.2]]
std:[[0.58, 0.58, 0.26]][[0.22, 0.0, 0.1]]
MSE:[[0.89, 0.77, 0.95]][[1.47, 70.89, 10.2]]
MSE(-DR):[[0.0, -0.12, 0.06]][[0.58, 70.0, 9.31]]
***
==============


O_threshold = 90
MC for this TARGET:[69.378, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.21, 0.07, -0.96]][[-0.34, -69.38, -8.69]]
std:[[0.36, 0.38, 0.24]][[0.27, 0.0, 0.1]]
MSE:[[0.42, 0.39, 0.99]][[0.43, 69.38, 8.69]]
MSE(-DR):[[0.0, -0.03, 0.57]][[0.01, 68.96, 8.27]]
***
==============


O_threshold = 100
MC for this TARGET:[68.934, 0.096]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.58, -2.7, -3.29]][[-4.69, -68.93, -8.24]]
std:[[0.58, 0.58, 0.34]][[0.27, 0.0, 0.1]]
MSE:[[2.64, 2.76, 3.31]][[4.7, 68.93, 8.24]]
MSE(-DR):[[0.0, 0.12, 0.67]][[2.06, 66.29, 5.6]]
***
==============


[[ 2.54  2.41  1.39  5.09 69.32  8.53]
 [ 0.92  0.79  1.05  1.3  70.88 10.09]
 [ 0.72  0.74  1.28  0.6  69.37  8.57]
 [ 2.74  2.87  3.44  4.85 68.94  8.14]]


[[ 2.96  2.83  1.68  5.2  69.31  8.56]
 [ 0.87  0.83  1.08  1.27 70.87 10.13]
 [ 0.36  0.49  1.14  0.56 69.36  8.61]
 [ 2.93  3.04  3.27  4.83 68.93  8.19]]


[[ 3.33  3.23  1.55  5.23 69.32  8.63]
 [ 0.89  0.77  0.95  1.47 70.89 10.2 ]
 [ 0.42  0.39  0.99  0.43 69.38  8.69]
 [ 2.64  2.76  3.31  4.7  68.93  8.24]]


time spent until now: 48.4 mins


--------------------------------------
```

```
[pattern_seed, day] = [2, 14]

max(u_O) =  197.9
O_threshold = 70
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

1 1 0 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

number of reward locations:  22
O_threshold = 80
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 90
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 0

number of reward locations:  12
O_threshold = 100
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
Rep 1 DONE
```