```
Last login: Tue Apr 21 01:30:47 on ttys001
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-34-205-133-58.compute-1.amazonaws.com
Warning: Permanently added the ED25519 host key for IP address '34.205.133.58' 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 Tue Apr 21 14:50:46 UTC 2020

  System load:  1.15               Processes:           877
  Usage of /:   28.0% of 30.96GB   Users logged in:     0
  Memory usage: 0%                 IP address for ens5: 172.31.6.185
  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

89 packages can be updated.
39 updates are security updates.


Last login: Fri Apr  3 19:45:17 2020 from 107.13.161.147
ubuntu@ip-172-31-6-185:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
10:53, 04/21; num of cores:96
final_T_small

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_D_range, t_func] = [96, None, None, 25, 0.5, 1.5, [80], None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[100, 101, 105, 110], [15], [2, 3, 4, 5], [[0.0003, 0.0001, 5e-05], [0.0003, 0.000
1, 5e-05]]]

---------------------------------------
[pattern_seed, day, sd_R, u_D] = [2, 2, 15, 80]

max(u_O) =  157.3 mean(u_O) =  93.7
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 = 101
number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:50.559
O_threshold = 100
```

```
MC for this TARGET:[58.028, 0.299]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.4, -2.55, -2.55]][[-3.19, -58.03, -7.47]]
std:[[1.64, 1.63, 1.0]][[0.93, 0.0, 0.53]]
sd_MSE:[array([0.91, 0.94, 0.53])][array([0.61, 0.  , 0.8 ])]
MSE:[[8.45, 9.16, 7.5]][[11.04, 3367.48, 56.08]]
RMSE(-DR):[[0.0, 0.71, -0.95]][[2.59, 3359.03, 47.63]]
**
==============
O_threshold = 101
MC for this TARGET:[55.198, 0.302]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.32, -0.44, -1.02]][[-2.21, -55.2, -4.64]]
std:[[1.75, 1.72, 1.0]][[0.91, 0.0, 0.53]]
sd_MSE:[array([0.5 , 0.49, 0.25])][array([0.42, 0.  , 0.5 ])]
MSE:[[3.16, 3.15, 2.04]][[5.71, 3047.04, 21.81]]
RMSE(-DR):[[0.0, -0.01, -1.12]][[2.55, 3043.88, 18.65]]
**
==============
O_threshold = 105
MC for this TARGET:[56.985, 0.298]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.16, -4.27, -4.23]][[-5.56, -56.98, -6.43]]
std:[[1.67, 1.68, 1.2]][[0.91, 0.0, 0.53]]
sd_MSE:[array([1.56, 1.58, 1.11])][array([1.03, 0.  , 0.69])]
MSE:[[20.09, 21.06, 19.33]][[31.74, 3246.72, 41.63]]
RMSE(-DR):[[0.0, 0.97, -0.76]][[11.65, 3226.63, 21.54]]
**
==============
O_threshold = 110
MC for this TARGET:[55.721, 0.295]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.47, -3.57, -3.97]][[-6.42, -55.72, -5.16]]
std:[[1.85, 1.84, 1.3]][[0.95, 0.0, 0.53]]
sd_MSE:[array([1.61, 1.62, 1.17])][array([1.24, 0.  , 0.55])]
MSE:[[15.46, 16.13, 17.45]][[42.12, 3104.72, 26.91]]
RMSE(-DR):[[0.0, 0.67, 1.99]][[26.66, 3089.26, 11.45]]
***
==============
[[8.4500e+00 9.1600e+00 7.5000e+00 1.1040e+01 3.3675e+03 5.6080e+01]
 [3.1600e+00 3.1500e+00 2.0400e+00 5.7100e+00 3.0470e+03 2.1810e+01]
 [2.0090e+01 2.1060e+01 1.9330e+01 3.1740e+01 3.2467e+03 4.1630e+01]
 [1.5460e+01 1.6130e+01 1.7450e+01 4.2120e+01 3.1047e+03 2.6910e+01]]


time spent until now: 60.5 mins

11:54, 04/21

--------------------------------------
[pattern_seed, day, sd_R, u_D] = [2, 3, 15, 80]

max(u_O) =  157.3 mean(u_O) =  93.7
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 = 101
number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!
```

```
--------------------------------------
Value of Behaviour policy:50.684
O_threshold = 100
MC for this TARGET:[58.021, 0.287]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.92, -2.1, -2.41]][[-3.04, -58.02, -7.34]]
std:[[1.43, 1.41, 0.92]][[0.75, 0.0, 0.41]]
sd_MSE:[array([0.65, 0.67, 0.45])][array([0.47, 0.  , 0.6 ])]
MSE:[[5.73, 6.4, 6.65]][[9.8, 3366.32, 54.04]]
RMSE(-DR):[[0.0, 0.67, 0.92]][[4.07, 3360.59, 48.31]]
***
==============
O_threshold = 101
MC for this TARGET:[55.201, 0.285]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.15, -0.29, -0.89]][[-2.13, -55.2, -4.52]]
std:[[1.45, 1.45, 0.95]][[0.75, 0.0, 0.41]]
sd_MSE:[array([0.32, 0.33, 0.22])][array([0.34, 0.  , 0.37])]
MSE:[[2.12, 2.19, 1.69]][[5.1, 3047.04, 20.6]]
RMSE(-DR):[[0.0, 0.07, -0.43]][[2.98, 3044.92, 18.48]]
**
==============
O_threshold = 105
MC for this TARGET:[56.998, 0.278]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.71, -3.83, -4.08]][[-5.46, -57.0, -6.31]]
std:[[1.6, 1.59, 1.03]][[0.75, 0.0, 0.41]]
sd_MSE:[array([1.33, 1.36, 0.87])][array([0.85, 0.  , 0.52])]
MSE:[[16.32, 17.2, 17.71]][[30.37, 3249.0, 39.98]]
RMSE(-DR):[[0.0, 0.88, 1.39]][[14.05, 3232.68, 23.66]]
***
==============
O_threshold = 110
MC for this TARGET:[55.738, 0.27]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.15, -3.24, -3.93]][[-6.31, -55.74, -5.05]]
std:[[1.63, 1.61, 0.99]][[0.75, 0.0, 0.41]]
sd_MSE:[array([1.13, 1.15, 0.77])][array([0.99, 0.  , 0.41])]
MSE:[[12.58, 13.09, 16.42]][[40.38, 3106.95, 25.67]]
RMSE(-DR):[[0.0, 0.51, 3.84]][[27.8, 3094.37, 13.09]]
***
==============
[[8.4500e+00 9.1600e+00 7.5000e+00 1.1040e+01 3.3675e+03 5.6080e+01]
 [3.1600e+00 3.1500e+00 2.0400e+00 5.7100e+00 3.0470e+03 2.1810e+01]
 [2.0090e+01 2.1060e+01 1.9330e+01 3.1740e+01 3.2467e+03 4.1630e+01]
 [1.5460e+01 1.6130e+01 1.7450e+01 4.2120e+01 3.1047e+03 2.6910e+01]]


[[5.7300e+00 6.4000e+00 6.6500e+00 9.8000e+00 3.3663e+03 5.4040e+01]
 [2.1200e+00 2.1900e+00 1.6900e+00 5.1000e+00 3.0470e+03 2.0600e+01]
 [1.6320e+01 1.7200e+01 1.7710e+01 3.0370e+01 3.2490e+03 3.9980e+01]
 [1.2580e+01 1.3090e+01 1.6420e+01 4.0380e+01 3.1069e+03 2.5670e+01]]


time spent until now: 123.0 mins

12:56, 04/21

--------------------------------------
[pattern_seed, day, sd_R, u_D] = [2, 4, 15, 80]

max(u_O) =  157.3 mean(u_O) =  93.7
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 = 101
number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:50.65
O_threshold = 100
MC for this TARGET:[58.01, 0.251]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.76, -1.95, -2.18]][[-2.98, -58.01, -7.36]]
std:[[0.97, 0.96, 0.74]][[0.65, 0.0, 0.35]]
sd_MSE:[array([0.38, 0.41, 0.36])][array([0.41, 0.  , 0.53])]
MSE:[[4.04, 4.72, 5.3]][[9.3, 3365.16, 54.29]]
RMSE(-DR):[[0.0, 0.68, 1.26]][[5.26, 3361.12, 50.25]]
***
==============
O_threshold = 101
MC for this TARGET:[55.185, 0.254]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.11, -0.04, -0.67]][[-2.1, -55.18, -4.54]]
std:[[0.95, 0.95, 0.76]][[0.67, 0.0, 0.35]]
sd_MSE:[array([0.13, 0.12, 0.15])][array([0.3 , 0.  , 0.33])]
MSE:[[0.91, 0.9, 1.03]][[4.86, 3044.83, 20.73]]
RMSE(-DR):[[0.0, -0.01, 0.12]][[3.95, 3043.92, 19.82]]
***
==============
O_threshold = 105
MC for this TARGET:[56.976, 0.255]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.48, -3.62, -3.92]][[-5.44, -56.98, -6.33]]
std:[[1.2, 1.17, 0.83]][[0.68, 0.0, 0.35]]
sd_MSE:[array([0.92, 0.93, 0.69])][array([0.76, 0.  , 0.46])]
MSE:[[13.55, 14.47, 16.06]][[30.06, 3246.72, 40.19]]
RMSE(-DR):[[0.0, 0.92, 2.51]][[16.51, 3233.17, 26.64]]
***
==============
O_threshold = 110
MC for this TARGET:[55.723, 0.245]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.06, -3.19, -3.83]][[-6.32, -55.72, -5.07]]
std:[[1.29, 1.27, 0.85]][[0.7, 0.0, 0.35]]
sd_MSE:[array([0.91, 0.93, 0.69])][array([0.9 , 0.  , 0.37])]
MSE:[[11.03, 11.79, 15.39]][[40.43, 3104.72, 25.83]]
RMSE(-DR):[[0.0, 0.76, 4.36]][[29.4, 3093.69, 14.8]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[8.4500e+00 9.1600e+00 7.5000e+00 1.1040e+01 3.3675e+03 5.6080e+01]
 [3.1600e+00 3.1500e+00 2.0400e+00 5.7100e+00 3.0470e+03 2.1810e+01]
 [2.0090e+01 2.1060e+01 1.9330e+01 3.1740e+01 3.2467e+03 4.1630e+01]
 [1.5460e+01 1.6130e+01 1.7450e+01 4.2120e+01 3.1047e+03 2.6910e+01]]


[[5.7300e+00 6.4000e+00 6.6500e+00 9.8000e+00 3.3663e+03 5.4040e+01]
 [2.1200e+00 2.1900e+00 1.6900e+00 5.1000e+00 3.0470e+03 2.0600e+01]
 [1.6320e+01 1.7200e+01 1.7710e+01 3.0370e+01 3.2490e+03 3.9980e+01]
 [1.2580e+01 1.3090e+01 1.6420e+01 4.0380e+01 3.1069e+03 2.5670e+01]]


[[4.0400e+00 4.7200e+00 5.3000e+00 9.3000e+00 3.3652e+03 5.4290e+01]
 [9.1000e-01 9.0000e-01 1.0300e+00 4.8600e+00 3.0448e+03 2.0730e+01]
 [1.3550e+01 1.4470e+01 1.6060e+01 3.0060e+01 3.2467e+03 4.0190e+01]
 [1.1030e+01 1.1790e+01 1.5390e+01 4.0430e+01 3.1047e+03 2.5830e+01]]


time spent until now: 188.4 mins

14:01, 04/21

---------------------------------------
[pattern_seed, day, sd_R, u_D] = [2, 5, 15, 80]

max(u_O) =  157.3 mean(u_O) =  93.7
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 = 101
number of reward locations:  8
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:50.686
O_threshold = 100
MC for this TARGET:[57.993, 0.217]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.62, -1.82, -2.22]][[-2.95, -57.99, -7.31]]
std:[[0.95, 0.94, 0.68]][[0.59, 0.0, 0.32]]
sd_MSE:[array([0.34, 0.36, 0.31])][array([0.36, 0.  , 0.48])]
MSE:[[3.53, 4.2, 5.39]][[9.05, 3362.84, 53.54]]
RMSE(-DR):[[0.0, 0.67, 1.86]][[5.52, 3359.31, 50.01]]
***
==============
O_threshold = 101
MC for this TARGET:[55.178, 0.219]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.07, -0.09, -0.72]][[-2.08, -55.18, -4.49]]
std:[[0.99, 0.99, 0.7]][[0.61, 0.0, 0.32]]
sd_MSE:[array([0.15, 0.15, 0.11])][array([0.27, 0.  , 0.29])]
MSE:[[0.98, 0.99, 1.01]][[4.7, 3044.83, 20.26]]
RMSE(-DR):[[0.0, 0.01, 0.03]][[3.72, 3043.85, 19.28]]
***
==============
O_threshold = 105
MC for this TARGET:[56.972, 0.217]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.39, -3.54, -3.88]][[-5.43, -56.97, -6.29]]
std:[[1.0, 1.0, 0.75]][[0.61, 0.0, 0.32]]
sd_MSE:[array([0.7 , 0.71, 0.6 ])][array([0.68, 0.  , 0.41])]
MSE:[[12.49, 13.53, 15.62]][[29.86, 3245.58, 39.67]]
RMSE(-DR):[[0.0, 1.04, 3.13]][[17.37, 3233.09, 27.18]]
***
==============
O_threshold = 110
MC for this TARGET:[55.715, 0.217]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.03, -3.15, -3.69]][[-6.27, -55.72, -5.03]]
std:[[1.01, 1.01, 0.81]][[0.62, 0.0, 0.32]]
sd_MSE:[array([0.65, 0.66, 0.61])][array([0.79, 0.  , 0.33])]
MSE:[[10.2, 10.94, 14.27]][[39.7, 3104.72, 25.4]]
RMSE(-DR):[[0.0, 0.74, 4.07]][[29.5, 3094.52, 15.2]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[8.4500e+00 9.1600e+00 7.5000e+00 1.1040e+01 3.3675e+03 5.6080e+01]
 [3.1600e+00 3.1500e+00 2.0400e+00 5.7100e+00 3.0470e+03 2.1810e+01]
 [2.0090e+01 2.1060e+01 1.9330e+01 3.1740e+01 3.2467e+03 4.1630e+01]
 [1.5460e+01 1.6130e+01 1.7450e+01 4.2120e+01 3.1047e+03 2.6910e+01]]


[[5.7300e+00 6.4000e+00 6.6500e+00 9.8000e+00 3.3663e+03 5.4040e+01]
 [2.1200e+00 2.1900e+00 1.6900e+00 5.1000e+00 3.0470e+03 2.0600e+01]
 [1.6320e+01 1.7200e+01 1.7710e+01 3.0370e+01 3.2490e+03 3.9980e+01]
 [1.2580e+01 1.3090e+01 1.6420e+01 4.0380e+01 3.1069e+03 2.5670e+01]]


[[4.0400e+00 4.7200e+00 5.3000e+00 9.3000e+00 3.3652e+03 5.4290e+01]
 [9.1000e-01 9.0000e-01 1.0300e+00 4.8600e+00 3.0448e+03 2.0730e+01]
 [1.3550e+01 1.4470e+01 1.6060e+01 3.0060e+01 3.2467e+03 4.0190e+01]
 [1.1030e+01 1.1790e+01 1.5390e+01 4.0430e+01 3.1047e+03 2.5830e+01]]
```

```
[[3.5300e+00 4.2000e+00 5.3900e+00 9.0500e+00 3.3628e+03 5.3540e+01]
 [9.8000e-01 9.9000e-01 1.0100e+00 4.7000e+00 3.0448e+03 2.0260e+01]
 [1.2490e+01 1.3530e+01 1.5620e+01 2.9860e+01 3.2456e+03 3.9670e+01]
 [1.0200e+01 1.0940e+01 1.4270e+01 3.9700e+01 3.1047e+03 2.5400e+01]]


time spent until now: 255.7 mins

15:09, 04/21
ubuntu@ip-172-31-6-185:~$
```