```
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-34-204-191-144.compute-1.amazonaws.com
The authenticity of host 'ec2-34-204-191-144.compute-1.amazonaws.com (34.204.191.144)' can't be established.
ECDSA key fingerprint is SHA256:ojFAJ9Hl5yAwQk9B5sVMNOBhjWy2sXsTWtQoJsfqU3g.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-34-204-191-144.compute-1.amazonaws.com,34.204.191.144' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1063-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Tue Apr  7 23:45:09 UTC 2020

  System load:  0.83               Processes:           819
  Usage of /:   57.0% of 15.45GB   Users logged in:     0
  Memory usage: 0%                 IP address for ens5: 172.31.75.110
  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

53 packages can be updated.
0 updates are security updates.


Last login: Wed Apr  1 20:30:39 2020 from 107.13.161.147
ubuntu@ip-172-31-75-110:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
19:46, 04/07; num of cores:96

final sd_R trend for[25, 50, 100] the same

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, u_O_u_D], sd_R_range, t_func] = [None,
96, None, None, None, 0.3, 0.5, 1, [True, False, True, 10], [25, 50, 100], None]


---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 25]

max(u_O) =  197.9
O_threshold = 95
number of reward locations:  11
O_threshold = 100
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:60.817
O_threshold = 95
MC for this TARGET:[69.18, 0.295]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.07, -1.22, -1.59]][[-1.56, -69.18, -8.36]]
std:[[1.1, 1.08, 0.71]][[0.62, 0.0, 0.34]]
MSE:[[1.53, 1.63, 1.74]][[1.68, 69.18, 8.37]]
MSE(-DR):[[0.0, 0.1, 0.21]][[0.15, 67.65, 6.84]]
***
==============
O_threshold = 100
MC for this TARGET:[68.98, 0.299]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.85, -2.98, -3.31]][[-4.72, -68.98, -8.16]]
std:[[1.22, 1.2, 0.69]][[0.63, 0.0, 0.34]]
MSE:[[3.1, 3.21, 3.38]][[4.76, 68.98, 8.17]]
MSE(-DR):[[0.0, 0.11, 0.28]][[1.66, 65.88, 5.07]]
***
==============
O_threshold = 105
MC for this TARGET:[71.436, 0.3]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.32, -6.44, -6.62]][[-7.94, -71.44, -10.62]]
std:[[1.47, 1.46, 0.77]][[0.67, 0.0, 0.34]]
```

```
MSE:[[6.49, 6.6, 6.66]][[7.97, 71.44, 10.63]]
MSE(-DR):[[0.0, 0.11, 0.17]][[1.48, 64.95, 4.14]]
***
==============
O_threshold = 110
MC for this TARGET:[70.519, 0.301]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.65, -6.73, -6.78]][[-8.72, -70.52, -9.7]]
std:[[1.56, 1.53, 0.9]][[0.66, 0.0, 0.34]]
MSE:[[6.83, 6.9, 6.84]][[8.74, 70.52, 9.71]]
MSE(-DR):[[0.0, 0.07, 0.01]][[1.91, 63.69, 2.88]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.53  1.63  1.74  1.68 69.18  8.37]
 [ 3.1   3.21  3.38  4.76 68.98  8.17]
 [ 6.49  6.6   6.66  7.97 71.44 10.63]
 [ 6.83  6.9   6.84  8.74 70.52  9.71]]


time spent until now: 64.7 mins


--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 50]

max(u_O) =  197.9
O_threshold = 95
number of reward locations:  11
O_threshold = 100
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:60.849
O_threshold = 95
MC for this TARGET:[69.22, 0.58]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.18, -1.33, -1.64]][[-1.66, -69.22, -8.37]]
std:[[2.13, 2.08, 1.37]][[1.2, 0.0, 0.58]]
MSE:[[2.44, 2.47, 2.14]][[2.05, 69.22, 8.39]]
MSE(-DR):[[0.0, 0.03, -0.3]][[-0.39, 66.78, 5.95]]
==============
O_threshold = 100
MC for this TARGET:[69.02, 0.583]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.04, -3.18, -3.4]][[-4.82, -69.02, -8.17]]
std:[[2.38, 2.35, 1.35]][[1.22, 0.0, 0.58]]
MSE:[[3.86, 3.95, 3.66]][[4.97, 69.02, 8.19]]
MSE(-DR):[[0.0, 0.09, -0.2]][[1.11, 65.16, 4.33]]
**
==============
O_threshold = 105
MC for this TARGET:[71.476, 0.586]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.5, -6.59, -6.69]][[-8.03, -71.48, -10.63]]
std:[[2.9, 2.83, 1.55]][[1.27, 0.0, 0.58]]
MSE:[[7.12, 7.17, 6.87]][[8.13, 71.48, 10.65]]
MSE(-DR):[[0.0, 0.05, -0.25]][[1.01, 64.36, 3.53]]
**
==============
O_threshold = 110
MC for this TARGET:[70.559, 0.586]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.93, -6.99, -6.91]][[-8.8, -70.56, -9.71]]
std:[[3.0, 2.94, 1.79]][[1.25, 0.0, 0.58]]
MSE:[[7.55, 7.58, 7.14]][[8.89, 70.56, 9.73]]
MSE(-DR):[[0.0, 0.03, -0.41]][[1.34, 63.01, 2.18]]
**
==============
[[ 1.53  1.63  1.74  1.68 69.18  8.37]
 [ 3.1   3.21  3.38  4.76 68.98  8.17]
 [ 6.49  6.6   6.66  7.97 71.44 10.63]
 [ 6.83  6.9   6.84  8.74 70.52  9.71]]


[[ 2.44  2.47  2.14  2.05 69.22  8.39]
 [ 3.86  3.95  3.66  4.97 69.02  8.19]
 [ 7.12  7.17  6.87  8.13 71.48 10.65]
 [ 7.55  7.58  7.14  8.89 70.56  9.73]]
```

```
time spent until now: 128.8 mins


--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 100]

max(u_O) =  197.9
O_threshold = 95
number of reward locations:  11
O_threshold = 100
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:60.913
O_threshold = 95
MC for this TARGET:[69.299, 1.155]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.36, -1.54, -1.67]][[-1.85, -69.3, -8.39]]
std:[[4.27, 4.13, 2.7]][[2.37, 0.0, 1.12]]
MSE:[[4.48, 4.41, 3.17]][[3.01, 69.3, 8.46]]
MSE(-DR):[[0.0, -0.07, -1.31]][[-1.47, 64.82, 3.98]]
==============
O_threshold = 100
MC for this TARGET:[69.099, 1.159]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.41, -3.53, -3.62]][[-5.0, -69.1, -8.19]]
std:[[4.77, 4.71, 2.75]][[2.41, 0.0, 1.12]]
MSE:[[5.86, 5.89, 4.55]][[5.55, 69.1, 8.27]]
MSE(-DR):[[0.0, 0.03, -1.31]][[-0.31, 63.24, 2.41]]
==============
O_threshold = 105
MC for this TARGET:[71.555, 1.161]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.71, -6.81, -6.69]][[-8.19, -71.56, -10.64]]
std:[[5.76, 5.66, 3.08]][[2.53, 0.0, 1.12]]
MSE:[[8.84, 8.86, 7.36]][[8.57, 71.56, 10.7]]
MSE(-DR):[[0.0, 0.02, -1.48]][[-0.27, 62.72, 1.86]]
==============
O_threshold = 110
MC for this TARGET:[70.638, 1.161]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.36, -7.45, -7.0]][[-8.98, -70.64, -9.72]]
std:[[5.98, 5.83, 3.57]][[2.47, 0.0, 1.12]]
MSE:[[9.48, 9.46, 7.86]][[9.31, 70.64, 9.78]]
MSE(-DR):[[0.0, -0.02, -1.62]][[-0.17, 61.16, 0.3]]
==============
[[ 1.53  1.63  1.74  1.68 69.18  8.37]
 [ 3.1   3.21  3.38  4.76 68.98  8.17]
 [ 6.49  6.6   6.66  7.97 71.44 10.63]
 [ 6.83  6.9   6.84  8.74 70.52  9.71]]


[[ 2.44  2.47  2.14  2.05 69.22  8.39]
 [ 3.86  3.95  3.66  4.97 69.02  8.19]
 [ 7.12  7.17  6.87  8.13 71.48 10.65]
 [ 7.55  7.58  7.14  8.89 70.56  9.73]]


[[ 4.48  4.41  3.17  3.01 69.3   8.46]
 [ 5.86  5.89  4.55  5.55 69.1   8.27]
 [ 8.84  8.86  7.36  8.57 71.56 10.7 ]
 [ 9.48  9.46  7.86  9.31 70.64  9.78]]


time spent until now: 194.6 mins

ubuntu@ip-172-31-75-110:~$
```