```
Last login: Wed Apr  8 00:18:56 on ttys001
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-232-95-73.compute-1.amazonaws.com
ssh: connect to host ec2-3-232-95-73.compute-1.amazonaws.com port 22: Connection refused
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-232-95-73.compute-1.amazonaws.com
ssh: connect to host ec2-3-232-95-73.compute-1.amazonaws.com port 22: Connection refused
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-232-95-73.compute-1.amazonaws.com
Warning: Permanently added the ED25519 host key for IP address '3.232.95.73' 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  8 14:55:23 UTC 2020

  System load:  0.84               Processes:           817
  Usage of /:   28.0% of 30.96GB   Users logged in:     0
  Memory usage: 0%                 IP address for ens5: 172.31.7.194
  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
export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
ubuntu@ip-172-31-7-194:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py

10:57, 04/08; num of cores:96

final sd_R trend for[10] the same

Basic setting:[T, rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, u_O_u_D], sd_R_range, t_func] = [None, 96, No
ne, None, 30, 0.5, 1, [True, False, True, 10], [10], None]


---------------------------------------
[pattern_seed, day, sd_R] = [2, 6, 10]

max(u_O) =  168.8
O_threshold = 80
number of reward locations:  15
O_threshold = 90
number of reward locations:  12
O_threshold = 100
number of reward locations:  9
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:57.75
O_threshold = 80
MC for this TARGET:[68.351, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.35, -0.55, -1.16]][[1.11, -68.35, -10.6]]
std:[[0.76, 0.76, 0.44]][[0.37, 0.0, 0.25]]
MSE:[[0.84, 0.94, 1.24]][[1.17, 68.35, 10.6]]
MSE(-DR):[[0.0, 0.1, 0.4]][[0.33, 67.51, 9.76]]
***
==============
O_threshold = 90
MC for this TARGET:[66.713, 0.14]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.15, -0.33, -0.8]][[-0.5, -66.71, -8.96]]
std:[[0.73, 0.73, 0.46]][[0.36, 0.0, 0.25]]
MSE:[[0.75, 0.8, 0.92]][[0.62, 66.71, 8.96]]
MSE(-DR):[[0.0, 0.05, 0.17]][[-0.13, 65.96, 8.21]]
==============
O_threshold = 100
MC for this TARGET:[66.955, 0.145]
```

```
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.76, -2.89, -3.11]][[-4.2, -66.96, -9.21]]
std:[[0.69, 0.7, 0.4]][[0.38, 0.0, 0.25]]
MSE:[[2.84, 2.97, 3.14]][[4.22, 66.96, 9.21]]
MSE(-DR):[[0.0, 0.13, 0.3]][[1.38, 64.12, 6.37]]
***
==============
O_threshold = 110
MC for this TARGET:[65.975, 0.144]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.07, -5.13, -5.35]][[-7.37, -65.97, -8.23]]
std:[[0.93, 0.93, 0.48]][[0.39, 0.0, 0.25]]
MSE:[[5.15, 5.21, 5.37]][[7.38, 65.97, 8.23]]
MSE(-DR):[[0.0, 0.06, 0.22]][[2.23, 60.82, 3.08]]
***
==============
[[ 0.84  0.94  1.24  1.17 68.35 10.6 ]
 [ 0.75  0.8   0.92  0.62 66.71  8.96]
 [ 2.84  2.97  3.14  4.22 66.96  9.21]
 [ 5.15  5.21  5.37  7.38 65.97  8.23]]


time spent until now: 61.6 mins


---------------------------------------
[pattern_seed, day, sd_R] = [2, 10, 10]

max(u_O) =  168.8
O_threshold = 80
number of reward locations:  15
O_threshold = 90
number of reward locations:  12
O_threshold = 100
number of reward locations:  9
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:57.766
O_threshold = 80
MC for this TARGET:[68.369, 0.103]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.3, 0.09, -1.04]][[1.18, -68.37, -10.6]]
std:[[0.58, 0.57, 0.36]][[0.26, 0.0, 0.18]]
MSE:[[0.65, 0.58, 1.1]][[1.21, 68.37, 10.6]]
MSE(-DR):[[0.0, -0.07, 0.45]][[0.56, 67.72, 9.95]]
***
==============
O_threshold = 90
MC for this TARGET:[66.736, 0.109]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.12, -0.06, -0.73]][[-0.44, -66.74, -8.97]]
std:[[0.46, 0.45, 0.34]][[0.26, 0.0, 0.18]]
MSE:[[0.48, 0.45, 0.81]][[0.51, 66.74, 8.97]]
MSE(-DR):[[0.0, -0.03, 0.33]][[0.03, 66.26, 8.49]]
***
==============
O_threshold = 100
MC for this TARGET:[66.966, 0.116]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.77, -2.89, -3.12]][[-4.17, -66.97, -9.2]]
std:[[0.52, 0.52, 0.36]][[0.26, 0.0, 0.18]]
MSE:[[2.82, 2.94, 3.14]][[4.18, 66.97, 9.2]]
MSE(-DR):[[0.0, 0.12, 0.32]][[1.36, 64.15, 6.38]]
***
==============
O_threshold = 110
MC for this TARGET:[65.981, 0.114]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.22, -5.29, -5.42]][[-7.36, -65.98, -8.22]]
std:[[0.62, 0.62, 0.39]][[0.27, 0.0, 0.18]]
MSE:[[5.26, 5.33, 5.43]][[7.36, 65.98, 8.22]]
MSE(-DR):[[0.0, 0.07, 0.17]][[2.1, 60.72, 2.96]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 0.84  0.94  1.24  1.17 68.35 10.6 ]
 [ 0.75  0.8   0.92  0.62 66.71  8.96]
 [ 2.84  2.97  3.14  4.22 66.96  9.21]
 [ 5.15  5.21  5.37  7.38 65.97  8.23]]


[[ 0.65  0.58  1.1   1.21 68.37 10.6 ]
 [ 0.48  0.45  0.81  0.51 66.74  8.97]
```

```
 [ 2.82  2.94  3.14  4.18 66.97  9.2 ]
 [ 5.26  5.33  5.43  7.36 65.98  8.22]]


time spent until now: 141.9 mins


--------------------------------------
[pattern_seed, day, sd_R] = [2, 14, 10]

max(u_0) =  168.8
O_threshold = 80
number of reward locations:  15
O_threshold = 90
number of reward locations:  12
O_threshold = 100
number of reward locations:  9
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:57.728
O_threshold = 80
MC for this TARGET:[68.37, 0.099]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.46, 0.22, -0.99]][[1.17, -68.37, -10.64]]
std:[[0.47, 0.47, 0.31]][[0.22, 0.0, 0.14]]
MSE:[[0.66, 0.52, 1.04]][[1.19, 68.37, 10.64]]
MSE(-DR):[[0.0, -0.14, 0.38]][[0.53, 67.71, 9.98]]
***
==============
O_threshold = 90
MC for this TARGET:[66.735, 0.098]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.11, -0.09, -0.68]][[-0.47, -66.74, -9.01]]
std:[[0.41, 0.4, 0.26]][[0.2, 0.0, 0.14]]
MSE:[[0.42, 0.41, 0.73]][[0.51, 66.74, 9.01]]
MSE(-DR):[[0.0, -0.01, 0.31]][[0.09, 66.32, 8.59]]
***
==============
O_threshold = 100
MC for this TARGET:[66.957, 0.102]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.86, -3.0, -3.09]][[-4.19, -66.96, -9.23]]
std:[[0.39, 0.38, 0.27]][[0.21, 0.0, 0.14]]
MSE:[[2.89, 3.02, 3.1]][[4.2, 66.96, 9.23]]
MSE(-DR):[[0.0, 0.13, 0.21]][[1.31, 64.07, 6.34]]
***
==============
O_threshold = 110
MC for this TARGET:[65.978, 0.098]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.31, -5.4, -5.36]][[-7.38, -65.98, -8.25]]
std:[[0.48, 0.47, 0.35]][[0.22, 0.0, 0.14]]
MSE:[[5.33, 5.42, 5.37]][[7.38, 65.98, 8.25]]
MSE(-DR):[[0.0, 0.09, 0.04]][[2.05, 60.65, 2.92]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 0.84  0.94  1.24  1.17 68.35 10.6 ]
 [ 0.75  0.8   0.92  0.62 66.71  8.96]
 [ 2.84  2.97  3.14  4.22 66.96  9.21]
 [ 5.15  5.21  5.37  7.38 65.97  8.23]]


[[ 0.65  0.58  1.1   1.21 68.37 10.6 ]
 [ 0.48  0.45  0.81  0.51 66.74  8.97]
 [ 2.82  2.94  3.14  4.18 66.97  9.2 ]
 [ 5.26  5.33  5.43  7.36 65.98  8.22]]


[[ 0.66  0.52  1.04  1.19 68.37 10.64]
 [ 0.42  0.41  0.73  0.51 66.74  9.01]
 [ 2.89  3.02  3.1   4.2  66.96  9.23]
 [ 5.33  5.42  5.37  7.38 65.98  8.25]]


time spent until now: 267.7 mins

ubuntu@ip-172-31-7-194:~$
ubuntu@ip-172-31-7-194:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
15:26, 04/08; num of cores:96

final sd_R trend for[10] the same
```

Basic setting:[T, rep_times, sd_O, sd_D, 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, 30, 0.5, 1, [True, False, True, 10], [10], None]

[pattern_seed, day, sd_R] = [2, 4, 10]

max(u_O) = 168.8
O_threshold = 80
number of reward locations: 15
O_threshold = 90
number of reward locations: 12
O_threshold = 100
number of reward locations: 9
O_threshold = 110
number of reward locations: 6