```
Last login: Fri Apr 10 13:09:15 on ttys001
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-35-169-116-116.compute-1.amazonaws.com
Warning: Permanently added the ED25519 host key for IP address '35.169.116.116' 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 Fri Apr 10 17:37:53 UTC 2020

  System load:  3.52               Processes:           384
  Usage of /:   28.0% of 30.96GB   Users logged in:     0
  Memory usage: 0%                 IP address for ens5: 172.31.8.69
  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-172-31-8-69:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
13:39, 04/10; num of cores:36vary_T

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, sd_R_range, t_func] = [36, None, None, 20, 0.5, 1, 0, [25], None]


---------------------------------------
[pattern_seed, day, sd_R] = [2, 3, 25]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:64.834
O_threshold = 100
MC for this TARGET:[70.804, 0.478]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.38, -0.46, -1.27]][[-1.4, -70.8, -5.97]]
std:[[2.24, 2.2, 1.56]][[1.13, 0.0, 0.63]]
MSE:[[2.27, 2.25, 2.01]][[1.8, 70.8, 6.0]]
MSE(-DR):[[0.0, -0.02, -0.26]][[-0.47, 68.53, 3.73]]
==============
O_threshold = 105
MC for this TARGET:[71.81, 0.475]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.02, -3.11, -3.46]][[-4.06, -71.81, -6.98]]
std:[[2.08, 2.07, 1.43]][[1.08, 0.0, 0.63]]
MSE:[[3.67, 3.74, 3.74]][[4.2, 71.81, 7.01]]
MSE(-DR):[[0.0, 0.07, 0.07]][[0.53, 68.14, 3.34]]
***
==============
O_threshold = 110
MC for this TARGET:[70.913, 0.474]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.07, -3.12, -3.39]][[-4.82, -70.91, -6.08]]
std:[[2.41, 2.37, 1.61]][[1.1, 0.0, 0.63]]
MSE:[[3.9, 3.92, 3.75]][[4.94, 70.91, 6.11]]
MSE(-DR):[[0.0, 0.02, -0.15]][[1.04, 67.01, 2.21]]
**
==============
O_threshold = 115
MC for this TARGET:[71.828, 0.457]
```

```
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.89, -5.88, -5.64]][[-9.61, -71.83, -6.99]]
std:[[3.09, 3.04, 1.93]][[1.11, 0.0, 0.63]]
MSE:[[6.65, 6.62, 5.96]][[9.67, 71.83, 7.02]]
MSE(-DR):[[0.0, -0.03, -0.69]][[3.02, 65.18, 0.37]]
**
==============
[[ 2.27  2.25  2.01  1.8  70.8   6.  ]
 [ 3.67  3.74  3.74  4.2  71.81  7.01]
 [ 3.9   3.92  3.75  4.94 70.91  6.11]
 [ 6.65  6.62  5.96  9.67 71.83  7.02]]


time spent until now: 42.9 mins

14:22, 04/10

--------------------------------------
[pattern_seed, day, sd_R] = [2, 5, 25]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

--------------------------------------
Value of Behaviour policy:64.868
O_threshold = 100
MC for this TARGET:[70.782, 0.334]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.21, -0.29, -1.12]][[-1.32, -70.78, -5.91]]
std:[[1.49, 1.49, 0.75]][[0.83, 0.0, 0.39]]
MSE:[[1.5, 1.52, 1.35]][[1.56, 70.78, 5.92]]
MSE(-DR):[[0.0, 0.02, -0.15]][[0.06, 69.28, 4.42]]
**
==============
O_threshold = 105
MC for this TARGET:[71.786, 0.328]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.45, -2.51, -3.53]][[-4.1, -71.79, -6.92]]
std:[[1.73, 1.72, 0.9]][[0.91, 0.0, 0.39]]
MSE:[[3.0, 3.04, 3.64]][[4.2, 71.79, 6.93]]
MSE(-DR):[[0.0, 0.04, 0.64]][[1.2, 68.79, 3.93]]
***
==============
O_threshold = 110
MC for this TARGET:[70.885, 0.33]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.42, -2.47, -3.4]][[-4.83, -70.89, -6.02]]
std:[[1.75, 1.75, 1.03]][[0.9, 0.0, 0.39]]
MSE:[[2.99, 3.03, 3.55]][[4.91, 70.89, 6.03]]
MSE(-DR):[[0.0, 0.04, 0.56]][[1.92, 67.9, 3.04]]
***
==============
O_threshold = 115
MC for this TARGET:[71.798, 0.327]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.36, -5.41, -5.73]][[-9.63, -71.8, -6.93]]
std:[[2.26, 2.22, 1.15]][[0.82, 0.0, 0.39]]
MSE:[[5.82, 5.85, 5.84]][[9.66, 71.8, 6.94]]
MSE(-DR):[[0.0, 0.03, 0.02]][[3.84, 65.98, 1.12]]
***
==============
[[ 2.27  2.25  2.01  1.8  70.8   6.  ]
 [ 3.67  3.74  3.74  4.2  71.81  7.01]
 [ 3.9   3.92  3.75  4.94 70.91  6.11]
 [ 6.65  6.62  5.96  9.67 71.83  7.02]]


[[ 1.5   1.52  1.35  1.56 70.78  5.92]
 [ 3.    3.04  3.64  4.2  71.79  6.93]
 [ 2.99  3.03  3.55  4.91 70.89  6.03]
 [ 5.82  5.85  5.84  9.66 71.8   6.94]]


time spent until now: 87.9 mins

15:07, 04/10

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

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

max(u_0) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:64.892
O_threshold = 100
MC for this TARGET:[70.815, 0.306]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.26, -0.35, -1.0]][[-1.31, -70.82, -5.92]]
std:[[1.16, 1.2, 0.78]][[0.6, 0.0, 0.31]]
MSE:[[1.19, 1.25, 1.27]][[1.44, 70.82, 5.93]]
MSE(-DR):[[0.0, 0.06, 0.08]][[0.25, 69.63, 4.74]]
***
==============
O_threshold = 105
MC for this TARGET:[71.819, 0.301]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.53, -2.6, -3.52]][[-4.04, -71.82, -6.93]]
std:[[1.5, 1.5, 0.79]][[0.62, 0.0, 0.31]]
MSE:[[2.94, 3.0, 3.61]][[4.09, 71.82, 6.94]]
MSE(-DR):[[0.0, 0.06, 0.67]][[1.15, 68.88, 4.0]]
***
==============
O_threshold = 110
MC for this TARGET:[70.913, 0.301]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.71, -2.79, -3.21]][[-4.76, -70.91, -6.02]]
std:[[1.3, 1.31, 0.77]][[0.61, 0.0, 0.31]]
MSE:[[3.01, 3.08, 3.3]][[4.8, 70.91, 6.03]]
MSE(-DR):[[0.0, 0.07, 0.29]][[1.79, 67.9, 3.02]]
***
==============
O_threshold = 115
MC for this TARGET:[71.837, 0.297]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.86, -5.88, -5.86]][[-9.62, -71.84, -6.95]]
std:[[1.68, 1.67, 1.11]][[0.61, 0.0, 0.31]]
MSE:[[6.1, 6.11, 5.96]][[9.64, 71.84, 6.96]]
MSE(-DR):[[0.0, 0.01, -0.14]][[3.54, 65.74, 0.86]]
**
==============
[[ 2.27  2.25  2.01  1.8  70.8   6.  ]
 [ 3.67  3.74  3.74  4.2  71.81  7.01]
 [ 3.9   3.92  3.75  4.94 70.91  6.11]
 [ 6.65  6.62  5.96  9.67 71.83  7.02]]


[[ 1.5   1.52  1.35  1.56 70.78  5.92]
 [ 3.    3.04  3.64  4.2  71.79  6.93]
 [ 2.99  3.03  3.55  4.91 70.89  6.03]
 [ 5.82  5.85  5.84  9.66 71.8   6.94]]


[[ 1.19  1.25  1.27  1.44 70.82  5.93]
 [ 2.94  3.    3.61  4.09 71.82  6.94]
 [ 3.01  3.08  3.3   4.8  70.91  6.03]
 [ 6.1   6.11  5.96  9.64 71.84  6.96]]


time spent until now: 136.7 mins

15:56, 04/10

---------------------------------------
[pattern_seed, day, sd_R] = [2, 9, 25]

max(u_0) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
```

```
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

----------------------------------------
Value of Behaviour policy:64.887
O_threshold = 100
MC for this TARGET:[70.795, 0.254]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.4, -0.51, -1.02]][[-1.46, -70.8, -5.91]]
std:[[1.16, 1.13, 0.8]][[0.6, 0.0, 0.26]]
MSE:[[1.23, 1.24, 1.3]][[1.58, 70.8, 5.92]]
MSE(-DR):[[0.0, 0.01, 0.07]][[0.35, 69.57, 4.69]]
***
==============
O_threshold = 105
MC for this TARGET:[71.793, 0.252]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.84, -2.93, -3.38]][[-4.15, -71.79, -6.91]]
std:[[1.46, 1.46, 0.83]][[0.58, 0.0, 0.26]]
MSE:[[3.19, 3.27, 3.48]][[4.19, 71.79, 6.91]]
MSE(-DR):[[0.0, 0.08, 0.29]][[1.0, 68.6, 3.72]]
***
==============
O_threshold = 110
MC for this TARGET:[70.888, 0.25]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.1, -3.2, -3.16]][[-4.88, -70.89, -6.0]]
std:[[1.55, 1.54, 0.84]][[0.6, 0.0, 0.26]]
MSE:[[3.47, 3.55, 3.27]][[4.92, 70.89, 6.01]]
MSE(-DR):[[0.0, 0.08, -0.2]][[1.45, 67.42, 2.54]]
**
==============
O_threshold = 115
MC for this TARGET:[71.817, 0.256]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.39, -6.46, -5.86]][[-9.79, -71.82, -6.93]]
std:[[1.73, 1.68, 1.0]][[0.57, 0.0, 0.26]]
MSE:[[6.62, 6.67, 5.94]][[9.81, 71.82, 6.93]]
MSE(-DR):[[0.0, 0.05, -0.68]][[3.19, 65.2, 0.31]]
**
==============
[[ 2.27  2.25  2.01  1.8  70.8   6.  ]
 [ 3.67  3.74  3.74  4.2  71.81  7.01]
 [ 3.9   3.92  3.75  4.94 70.91  6.11]
 [ 6.65  6.62  5.96  9.67 71.83  7.02]]


[[ 1.5   1.52  1.35  1.56 70.78  5.92]
 [ 3.    3.04  3.64  4.2  71.79  6.93]
 [ 2.99  3.03  3.55  4.91 70.89  6.03]
 [ 5.82  5.85  5.84  9.66 71.8   6.94]]


[[ 1.19  1.25  1.27  1.44 70.82  5.93]
 [ 2.94  3.    3.61  4.09 71.82  6.94]
 [ 3.01  3.08  3.3   4.8  70.91  6.03]
 [ 6.1   6.11  5.96  9.64 71.84  6.96]]


[[ 1.23  1.24  1.3   1.58 70.8   5.92]
 [ 3.19  3.27  3.48  4.19 71.79  6.91]
 [ 3.47  3.55  3.27  4.92 70.89  6.01]
 [ 6.62  6.67  5.94  9.81 71.82  6.93]]


time spent until now: 192.2 mins

16:51, 04/10
ubuntu@ip-172-31-8-69:~$
```