```
Last login: Thu Apr  9 16:36:21 on ttys001
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-235-86-16.compute-1.amazonaws.com
The authenticity of host 'ec2-3-235-86-16.compute-1.amazonaws.com (3.235.86.16)' can't be established.
ECDSA key fingerprint is SHA256:L5q8YeNxmqwssxAudfqCfC8+R/ISK0E2u1CF/co7vqI.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-235-86-16.compute-1.amazonaws.com,3.235.86.16' (ECDSA) 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 Thu Apr  9 20:52:19 UTC 2020

  System load:  1.22                Processes:            225
  Usage of /:   28.0% of 30.96GB    Users logged in:      0
  Memory usage: 1%                  IP address for ens5: 172.31.15.219
  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.pyubuntu@ip-172-31-15-219:~$ export openblas_num_threads=1; export O
MP_NUM_THREADS=1; python EC2.py
16:53, 04/09; num of cores:16

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, sd_R_range, t_func] = [16, None, None, 20, 0.5, 1, 0, [0, 10, 20, 30],
None]


---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 0]

max(u_O) =  145.8
O_threshold = 80
number of reward locations:  20
O_threshold = 90
number of reward locations:  14
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
O_threshold = 120
number of reward locations:  3
target 1 in 5 DONE!
target 2 in 5 DONE!
target 3 in 5 DONE!
target 4 in 5 DONE!
target 5 in 5 DONE!


---------------------------------------
Value of Behaviour policy:64.845
O_threshold = 80
MC for this TARGET:[76.333, 0.082]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.92, 0.71, 0.26]][[2.02, -76.33, -11.49]]
std:[[0.36, 0.38, 0.27]][[0.27, 0.0, 0.2]]
MSE:[[0.99, 0.81, 0.37]][[2.04, 76.33, 11.49]]
MSE(-DR):[[0.0, -0.18, -0.62]][[1.05, 75.34, 10.5]]
**
==============
O_threshold = 90
MC for this TARGET:[73.912, 0.082]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.46, 0.29, -1.06]][[0.83, -73.91, -9.07]]
std:[[0.3, 0.31, 0.29]][[0.31, 0.0, 0.2]]
MSE:[[0.55, 0.42, 1.1]][[0.89, 73.91, 9.07]]
MSE(-DR):[[0.0, -0.13, 0.55]][[0.34, 73.36, 8.52]]
***
==============
O_threshold = 100
MC for this TARGET:[70.776, 0.095]
```

```
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.3, -0.4, -0.99]][[-1.28, -70.78, -5.93]]
std:[[0.51, 0.51, 0.34]][[0.31, 0.0, 0.2]]
MSE:[[0.59, 0.65, 1.05]][[1.32, 70.78, 5.93]]
MSE(-DR):[[0.0, 0.06, 0.46]][[0.73, 70.19, 5.34]]
***
==============
O_threshold = 110
MC for this TARGET:[70.873, 0.086]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.61, -2.65, -3.14]][[-4.71, -70.87, -6.03]]
std:[[0.46, 0.45, 0.38]][[0.32, 0.0, 0.2]]
MSE:[[2.65, 2.69, 3.16]][[4.72, 70.87, 6.03]]
MSE(-DR):[[0.0, 0.04, 0.51]][[2.07, 68.22, 3.38]]
***
==============
O_threshold = 120
MC for this TARGET:[71.797, 0.073]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.7, -5.72, -5.63]][[-9.59, -71.8, -6.95]]
std:[[0.56, 0.57, 0.4]][[0.31, 0.0, 0.2]]
MSE:[[5.73, 5.75, 5.64]][[9.6, 71.8, 6.95]]
MSE(-DR):[[0.0, 0.02, -0.09]][[3.87, 66.07, 1.22]]
**
==============
[[ 0.99  0.81  0.37  2.04 76.33 11.49]
 [ 0.55  0.42  1.1   0.89 73.91  9.07]
 [ 0.59  0.65  1.05  1.32 70.78  5.93]
 [ 2.65  2.69  3.16  4.72 70.87  6.03]
 [ 5.73  5.75  5.64  9.6  71.8   6.95]]


time spent until now: 48.2 mins

17:42, 04/09

---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 10]

max(u_O) =  145.8
O_threshold = 80
number of reward locations:  20
O_threshold = 90
number of reward locations:  14
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
O_threshold = 120
number of reward locations:  3
target 1 in 5 DONE!
target 2 in 5 DONE!
target 3 in 5 DONE!
target 4 in 5 DONE!
target 5 in 5 DONE!

---------------------------------------
Value of Behaviour policy:64.831
O_threshold = 80
MC for this TARGET:[76.349, 0.142]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.04, 0.81, 0.36]][[2.09, -76.35, -11.52]]
std:[[0.73, 0.73, 0.44]][[0.36, 0.0, 0.15]]
MSE:[[1.27, 1.09, 0.57]][[2.12, 76.35, 11.52]]
MSE(-DR):[[0.0, -0.18, -0.7]][[0.85, 75.08, 10.25]]
**
==============
O_threshold = 90
MC for this TARGET:[73.927, 0.143]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.65, 0.49, -0.92]][[0.88, -73.93, -9.1]]
std:[[0.4, 0.41, 0.35]][[0.38, 0.0, 0.15]]
MSE:[[0.76, 0.64, 0.98]][[0.96, 73.93, 9.1]]
MSE(-DR):[[0.0, -0.12, 0.22]][[0.2, 73.17, 8.34]]
***
==============
O_threshold = 100
MC for this TARGET:[70.792, 0.152]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.31, -0.37, -1.08]][[-1.32, -70.79, -5.96]]
std:[[0.51, 0.54, 0.42]][[0.29, 0.0, 0.15]]
MSE:[[0.6, 0.65, 1.16]][[1.35, 70.79, 5.96]]
MSE(-DR):[[0.0, 0.05, 0.56]][[0.75, 70.19, 5.36]]
***
==============
O_threshold = 110
MC for this TARGET:[70.889, 0.144]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.5, -2.55, -3.17]][[-4.75, -70.89, -6.06]]
```

```
std:[[0.53, 0.57, 0.43]][[0.31, 0.0, 0.15]]
MSE:[[2.56, 2.61, 3.2]][[4.76, 70.89, 6.06]]
MSE(-DR):[[0.0, 0.05, 0.64]][[2.2, 68.33, 3.5]]
***
==============
O_threshold = 120
MC for this TARGET:[71.813, 0.136]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.67, -5.69, -5.66]][[-9.62, -71.81, -6.98]]
std:[[0.58, 0.54, 0.53]][[0.29, 0.0, 0.15]]
MSE:[[5.7, 5.72, 5.68]][[9.62, 71.81, 6.98]]
MSE(-DR):[[0.0, 0.02, -0.02]][[3.92, 66.11, 1.28]]
**
==============
[[ 0.99  0.81  0.37  2.04 76.33 11.49]
 [ 0.55  0.42  1.1   0.89 73.91  9.07]
 [ 0.59  0.65  1.05  1.32 70.78  5.93]
 [ 2.65  2.69  3.16  4.72 70.87  6.03]
 [ 5.73  5.75  5.64  9.6  71.8   6.95]]


[[ 1.27  1.09  0.57  2.12 76.35 11.52]
 [ 0.76  0.64  0.98  0.96 73.93  9.1 ]
 [ 0.6   0.65  1.16  1.35 70.79  5.96]
 [ 2.56  2.61  3.2   4.76 70.89  6.06]
 [ 5.7   5.72  5.68  9.62 71.81  6.98]]


time spent until now: 99.6 mins

18:33, 04/09

---------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 20]

max(u_O) =  145.8
O_threshold = 80
number of reward locations:  20
O_threshold = 90
number of reward locations:  14
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
O_threshold = 120
number of reward locations:  3
=target 1 in 5 DONE!
target 2 in 5 DONE!
target 3 in 5 DONE!
target 4 in 5 DONE!
target 5 in 5 DONE!


---------------------------------------
Value of Behaviour policy:64.817
O_threshold = 80
MC for this TARGET:[76.364, 0.245]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.11, 0.88, 0.41]][[2.15, -76.36, -11.55]]
std:[[1.22, 1.21, 0.74]][[0.51, 0.0, 0.18]]
MSE:[[1.65, 1.5, 0.85]][[2.21, 76.36, 11.55]]
MSE(-DR):[[0.0, -0.15, -0.8]][[0.56, 74.71, 9.9]]
**
==============
O_threshold = 90
MC for this TARGET:[73.943, 0.247]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.84, 0.67, -0.78]][[0.92, -73.94, -9.13]]
std:[[0.63, 0.65, 0.49]][[0.55, 0.0, 0.18]]
MSE:[[1.05, 0.93, 0.92]][[1.07, 73.94, 9.13]]
MSE(-DR):[[0.0, -0.12, -0.13]][[0.02, 72.89, 8.08]]
**
==============
O_threshold = 100
MC for this TARGET:[70.808, 0.252]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.31, -0.35, -1.14]][[-1.36, -70.81, -5.99]]
std:[[0.85, 0.87, 0.63]][[0.41, 0.0, 0.18]]
MSE:[[0.9, 0.94, 1.3]][[1.42, 70.81, 5.99]]
MSE(-DR):[[0.0, 0.04, 0.4]][[0.52, 69.91, 5.09]]
***
==============
O_threshold = 110
MC for this TARGET:[70.905, 0.246]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.38, -2.44, -3.21]][[-4.78, -70.9, -6.09]]
std:[[0.96, 1.01, 0.66]][[0.42, 0.0, 0.18]]
MSE:[[2.57, 2.64, 3.28]][[4.8, 70.9, 6.09]]
MSE(-DR):[[0.0, 0.07, 0.71]][[2.23, 68.33, 3.52]]
***
```

```
==============
O_threshold = 120
MC for this TARGET:[71.829, 0.241]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.65, -5.65, -5.72]][[-9.65, -71.83, -7.01]]
std:[[1.01, 0.96, 0.85]][[0.43, 0.0, 0.18]]
MSE:[[5.74, 5.73, 5.78]][[9.66, 71.83, 7.01]]
MSE(-DR):[[0.0, -0.01, 0.04]][[3.92, 66.09, 1.27]]
***
==============
[[ 0.99  0.81  0.37  2.04 76.33 11.49]
 [ 0.55  0.42  1.1   0.89 73.91  9.07]
 [ 0.59  0.65  1.05  1.32 70.78  5.93]
 [ 2.65  2.69  3.16  4.72 70.87  6.03]
 [ 5.73  5.75  5.64  9.6  71.8   6.95]]


[[ 1.27  1.09  0.57  2.12 76.35 11.52]
 [ 0.76  0.64  0.98  0.96 73.93  9.1 ]
 [ 0.6   0.65  1.16  1.35 70.79  5.96]
 [ 2.56  2.61  3.2   4.76 70.89  6.06]
 [ 5.7   5.72  5.68  9.62 71.81  6.98]]


[[ 1.65  1.5   0.85  2.21 76.36 11.55]
 [ 1.05  0.93  0.92  1.07 73.94  9.13]
 [ 0.9   0.94  1.3   1.42 70.81  5.99]
 [ 2.57  2.64  3.28  4.8  70.9   6.09]
 [ 5.74  5.73  5.78  9.66 71.83  7.01]]


time spent until now: 151.3 mins

19:25, 04/09

----------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 30]

max(u_O) =  145.8
O_threshold = 80
number of reward locations:  20
O_threshold = 90
number of reward locations:  14
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
O_threshold = 120
number of reward locations:  3
target 1 in 5 DONE!
target 2 in 5 DONE!
target 3 in 5 DONE!
target 4 in 5 DONE!
target 5 in 5 DONE!


----------------------------------------
Value of Behaviour policy:64.802
O_threshold = 80
MC for this TARGET:[76.38, 0.356]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.14, 0.91, 0.46]][[2.2, -76.38, -11.58]]
std:[[1.79, 1.7, 1.1]][[0.7, 0.0, 0.26]]
MSE:[[2.12, 1.93, 1.19]][[2.31, 76.38, 11.58]]
MSE(-DR):[[0.0, -0.19, -0.93]][[0.19, 74.26, 9.46]]
**
==============
O_threshold = 90
MC for this TARGET:[73.959, 0.358]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.03, 0.85, -0.66]][[0.97, -73.96, -9.16]]
std:[[0.89, 0.92, 0.67]][[0.75, 0.0, 0.26]]
MSE:[[1.36, 1.25, 0.94]][[1.23, 73.96, 9.16]]
MSE(-DR):[[0.0, -0.11, -0.42]][[-0.13, 72.6, 7.8]]
==============
O_threshold = 100
MC for this TARGET:[70.823, 0.362]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.29, -0.36, -1.2]][[-1.4, -70.82, -6.02]]
std:[[1.25, 1.29, 0.9]][[0.59, 0.0, 0.26]]
MSE:[[1.28, 1.34, 1.5]][[1.52, 70.82, 6.03]]
MSE(-DR):[[0.0, 0.06, 0.22]][[0.24, 69.54, 4.75]]
***
==============
O_threshold = 110
MC for this TARGET:[70.921, 0.356]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.35, -2.38, -3.28]][[-4.82, -70.92, -6.12]]
std:[[1.5, 1.53, 0.87]][[0.61, 0.0, 0.26]]
MSE:[[2.79, 2.83, 3.39]][[4.86, 70.92, 6.13]]
```

MSE(-DR):[[0.0, 0.04, 0.6]][[2.07, 68.13, 3.34]]
***
==============
O_threshold = 120
MC for this TARGET:[71.845, 0.353]
     [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.65, -5.65, -5.75]][[-9.68, -71.84, -7.04]]
std:[[1.59, 1.5, 1.32]][[0.64, 0.0, 0.26]]
MSE:[[5.87, 5.85, 5.9]][[9.7, 71.84, 7.04]]
MSE(-DR):[[0.0, -0.02, 0.03]][[3.83, 65.97, 1.17]]
***
==============
[[0.99 0.81 0.37 2.04 76.33 11.49]
 [0.55 0.42 1.1   0.89 73.91 9.07]
 [0.59 0.65 1.05 1.32 70.78 5.93]
 [2.65 2.69 3.16 4.72 70.87 6.03]
 [5.73 5.75 5.64 9.6 71.8   6.95]]

[[1.27 1.09 0.57 2.12 76.35 11.52]
 [0.76 0.64 0.98 0.96 73.93 9.1]
 [0.6   0.65 1.16 1.35 70.79 5.96]
 [2.56 2.61 3.2   4.76 70.89 6.06]
 [5.7   5.72 5.68 9.62 71.81 6.98]]

[[1.65 1.5   0.85 2.21 76.36 11.55]
 [1.05 0.93 0.92 1.07 73.94 9.13]
 [0.9   0.94 1.3   1.42 70.81 5.99]
 [2.57 2.64 3.28 4.8 70.9   6.09]
 [5.74 5.73 5.78 9.66 71.83 7.01]]

[[2.12 1.93 1.19 2.31 76.38 11.58]
 [1.36 1.25 0.94 1.23 73.96 9.16]
 [1.28 1.34 1.5   1.52 70.82 6.03]
 [2.79 2.83 3.39 4.86 70.92 6.13]
 [5.87 5.85 5.9   9.7 71.84 7.04]]

time spent until now: 203.0 mins

20:16, 04/09
ubuntu@ip-172-31-15-219:~$ =