```
Last login: Mon Mar 30 10:50:07 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-215-134-165.compute-1.amazonaws.com
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 Mon Mar 30 15:15:55 UTC 2020

  System load:  1.19               Processes:           212
  Usage of /:   56.9% of 15.45GB   Users logged in:     0
  Memory usage: 1%                 IP address for ens5: 172.31.9.80
  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

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


*** System restart required ***
Last login: Mon Mar 30 14:50:10 2020 from 107.13.161.147
ubuntu@ip-172-31-9-80:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1
ubuntu@ip-172-31-9-80:~$ python EC2.py
11:16, 03/30; num of cores:16

Basic setting:[sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, lam, simple, M_in_R] = [5, 5, 5, 0.2, 2, 2, 1e-05, True, True]


--------------------------------------
[pattern_seed, T, sd_R] = [0, 336, 5]

max(u_O) =  156.6
O_threshold = 100
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = -4
target policy:

1 1 1 1 1

0 1 0 0 0

1 0 1 1 0

1 0 0 0 1

0 1 1 1 0

number of reward locations:  14
O_threshold = -3
target policy:
```

```
1 1 0 1 1

1 0 0 0 0

0 0 1 0 1

1 0 1 0 0

0 1 0 0 1

number of reward locations:  11
O_threshold = -2
target policy:

0 0 1 0 0

0 0 1 0 0

1 1 0 1 0

1 1 0 1 0

1 0 0 0 0

number of reward locations:  9
O_threshold = -1
target policy:

0 1 0 0 0

0 0 0 0 1

0 1 0 1 0

1 0 1 0 0

1 1 0 1 1

number of reward locations:  10
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
1 2 3 4 5 6 1 2 3 4 5 6
--------------------------------------
Value of Behaviour policy:85.584
O_threshold = 100
MC for this TARGET:[96.117, 0.112]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.52, -0.65, -0.92]][[0.1, -0.1, -0.21]][[-96.12, -96.12, -96.12]][[-1.04, -10.53]]
std:[[0.33, 0.24, 0.18]][[0.19, 0.16, 0.2]][[0.0, 0.0, 0.0]][[0.09, 0.06]]
MSE:[[0.62, 0.69, 0.94]][[0.21, 0.19, 0.29]][[96.12, 96.12, 96.12]][[1.04, 10.53]]
MSE(-DR):[[0.0, 0.07, 0.32]][[-0.41, -0.43, -0.33]][[95.5, 95.5, 95.5]][[0.42, 9.91]]
==============


O_threshold = -4
MC for this TARGET:[88.81, 0.106]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.23, 0.23, -0.68]][[1.06, 0.97, 0.9]][[-88.81, -88.81, -88.81]][[-0.69, -3.23]]
std:[[0.05, 0.07, 0.04]][[0.04, 0.03, 0.02]][[0.0, 0.0, 0.0]][[0.06, 0.06]]
MSE:[[0.24, 0.24, 0.68]][[1.06, 0.97, 0.9]][[88.81, 88.81, 88.81]][[0.69, 3.23]]
MSE(-DR):[[0.0, 0.0, 0.44]][[0.82, 0.73, 0.66]][[88.57, 88.57, 88.57]][[0.45, 2.99]]
******
MC-based ATE = -7.31
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.76, 0.88, 0.23]][[0.96, 1.07, 1.11]][[7.31, 7.31, 7.31]][0.35]
std:[[0.38, 0.31, 0.22]][[0.15, 0.13, 0.19]][[0.0, 0.0, 0.0]][0.15]
MSE:[[0.85, 0.93, 0.32]][[0.97, 1.08, 1.13]][[7.31, 7.31, 7.31]][0.38]
MSE(-DR):[[0.0, 0.08, -0.53]][[0.12, 0.23, 0.28]][[6.46, 6.46, 6.46]][-0.47]
better than DR_NO_MARL
==============


O_threshold = -3
MC for this TARGET:[90.122, 0.107]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.54, -2.61, -2.92]][[-2.07, -2.15, -2.22]][[-90.12, -90.12, -90.12]][[-3.0, -4.54]]
```

```
std:[[0.35, 0.37, 0.13]][[0.11, 0.04, 0.01]][[0.0, 0.0, 0.0]][[0.15, 0.06]]
MSE:[[2.56, 2.64, 2.92]][[2.07, 2.15, 2.22]][[90.12, 90.12, 90.12]][[3.0, 4.54]]
MSE(-DR):[[0.0, 0.08, 0.36]][[-0.49, -0.41, -0.34]][[87.56, 87.56, 87.56]][[0.44, 1.98]]
MC-based ATE = -6.0
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.02, -1.97, -2.01]][[-2.17, -2.05, -2.01]][[6.0, 6.0, 6.0]][-1.96]
std:[[0.68, 0.61, 0.31]][[0.09, 0.12, 0.2]][[0.0, 0.0, 0.0]][0.24]
MSE:[[2.13, 2.06, 2.03]][[2.17, 2.05, 2.02]][[6.0, 6.0, 6.0]][1.97]
MSE(-DR):[[0.0, -0.07, -0.1]][[0.04, -0.08, -0.11]][[3.87, 3.87, 3.87]][-0.16]
==============


O_threshold = -2
MC for this TARGET:[85.228, 0.107]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.28, 0.24, 0.08]][[-1.08, -1.11, -1.11]][[-85.23, -85.23, -85.23]][[0.04, 0.36]]
std:[[0.56, 0.55, 0.19]][[0.09, 0.07, 0.17]][[0.0, 0.0, 0.0]][[0.18, 0.06]]
MSE:[[0.63, 0.6, 0.21]][[1.08, 1.11, 1.12]][[85.23, 85.23, 85.23]][[0.18, 0.36]]
MSE(-DR):[[0.0, -0.03, -0.42]][[0.45, 0.48, 0.49]][[84.6, 84.6, 84.6]][[-0.45, -0.27]]
better than DR_NO_MARL
MC-based ATE = -10.89
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.8, 0.88, 1.0]][[-1.18, -1.0, -0.91]][[10.89, 10.89, 10.89]][1.08]
std:[[0.23, 0.31, 0.0]][[0.11, 0.09, 0.03]][[0.0, 0.0, 0.0]][0.09]
MSE:[[0.83, 0.93, 1.0]][[1.19, 1.0, 0.91]][[10.89, 10.89, 10.89]][1.08]
MSE(-DR):[[0.0, 0.1, 0.17]][[0.36, 0.17, 0.08]][[10.06, 10.06, 10.06]][0.25]
******
==============


O_threshold = -1
MC for this TARGET:[86.753, 0.1]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.91, -2.93, -3.64]][[-3.11, -3.1, -3.1]][[-86.75, -86.75, -86.75]][[-3.67, -1.17]]
std:[[0.17, 0.16, 0.08]][[0.12, 0.07, 0.11]][[0.0, 0.0, 0.0]][[0.08, 0.06]]
MSE:[[2.91, 2.93, 3.64]][[3.11, 3.1, 3.1]][[86.75, 86.75, 86.75]][[3.67, 1.17]]
MSE(-DR):[[0.0, 0.02, 0.73]][[0.2, 0.19, 0.19]][[83.84, 83.84, 83.84]][[0.76, -1.74]]
******
MC-based ATE = -9.36
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.38, -2.29, -2.73]][[-3.22, -3.0, -2.9]][[9.36, 9.36, 9.36]][-2.63]
std:[[0.5, 0.4, 0.11]][[0.07, 0.09, 0.1]][[0.0, 0.0, 0.0]][0.01]
MSE:[[2.43, 2.32, 2.73]][[3.22, 3.0, 2.9]][[9.36, 9.36, 9.36]][2.63]
MSE(-DR):[[0.0, -0.11, 0.3]][[0.79, 0.57, 0.47]][[6.93, 6.93, 6.93]][0.2]
******
==============


O_threshold = 90
MC for this TARGET:[95.767, 0.112]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.16, 0.04, -0.09]][[0.22, 0.08, -0.05]][[-95.77, -95.77, -95.77]][[-0.21, -10.18]]
std:[[0.05, 0.03, 0.09]][[0.1, 0.08, 0.11]][[0.0, 0.0, 0.0]][[0.06, 0.06]]
MSE:[[0.17, 0.05, 0.13]][[0.24, 0.11, 0.12]][[95.77, 95.77, 95.77]][[0.22, 10.18]]
MSE(-DR):[[0.0, -0.12, -0.04]][[0.07, -0.06, -0.05]][[95.6, 95.6, 95.6]][[0.05, 10.01]]
MC-based ATE = -0.35
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.68, 0.69, 0.82]][[0.12, 0.18, 0.16]][[0.35, 0.35, 0.35]][0.83]
std:[[0.28, 0.21, 0.09]][[0.09, 0.09, 0.09]][[0.0, 0.0, 0.0]][0.03]
MSE:[[0.74, 0.72, 0.82]][[0.15, 0.2, 0.18]][[0.35, 0.35, 0.35]][0.83]
MSE(-DR):[[0.0, -0.02, 0.08]][[-0.59, -0.54, -0.56]][[-0.39, -0.39, -0.39]][0.09]
==============


time spent until now: 4.2 mins


-------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 5]

max(u_O) =  156.6
O_threshold = 100
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

0 1 0 0 1
```

```
1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = -4
target policy:

1 1 1 1 1

0 1 0 0 0

1 0 1 1 0

1 0 0 0 1

0 1 1 1 0

number of reward locations:  14
O_threshold = -3
target policy:

1 1 0 1 1

1 0 0 0 0

0 0 1 0 1

1 0 1 0 0

0 1 0 0 1

number of reward locations:  11
O_threshold = -2
target policy:

0 0 1 0 0

0 0 1 0 0

1 1 0 1 0

1 1 0 1 0

1 0 0 0 0

number of reward locations:  9
O_threshold = -1
target policy:

0 1 0 0 0

0 0 0 0 1

0 1 0 1 0

1 0 1 0 0

1 1 0 1 1

number of reward locations:  10
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
1 2 3 4 5 6 1 2 3 4 5 6
--------------------------------------
Value of Behaviour policy:85.52
O_threshold = 100
MC for this TARGET:[96.107, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.09, -0.25, -1.01]][[-0.14, -0.38, -0.38]][[-96.11, -96.11, -96.11]][[-1.17, -10.59]]
std:[[0.06, 0.07, 0.08]][[0.14, 0.09, 0.13]][[0.0, 0.0, 0.0]][[0.07, 0.03]]
MSE:[[0.11, 0.26, 1.01]][[0.2, 0.39, 0.4]][[96.11, 96.11, 96.11]][[1.17, 10.59]]
MSE(-DR):[[0.0, 0.15, 0.9]][[0.09, 0.28, 0.29]][[96.0, 96.0, 96.0]][[1.06, 10.48]]
******
```

```
==============


O_threshold = -4
MC for this TARGET:[88.801, 0.082]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.06, 0.0, -0.73]][[0.57, 0.45, 0.4]][[-88.8, -88.8, -88.8]][[-0.79, -3.28]]
std:[[0.27, 0.21, 0.29]][[0.23, 0.19, 0.19]][[0.0, 0.0, 0.0]][[0.24, 0.03]]
MSE:[[0.28, 0.21, 0.79]][[0.61, 0.49, 0.44]][[88.8, 88.8, 88.8]][[0.83, 3.28]]
MSE(-DR):[[0.0, -0.07, 0.51]][[0.33, 0.21, 0.16]][[88.52, 88.52, 88.52]][[0.55, 3.0]]
******
MC-based ATE = -7.31
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.15, 0.25, 0.28]][[0.71, 0.83, 0.78]][[7.31, 7.31, 7.31]][0.38]
std:[[0.2, 0.14, 0.38]][[0.09, 0.1, 0.07]][[0.0, 0.0, 0.0]][0.31]
MSE:[[0.25, 0.29, 0.47]][[0.72, 0.84, 0.78]][[7.31, 7.31, 7.31]][0.49]
MSE(-DR):[[0.0, 0.04, 0.22]][[0.47, 0.59, 0.53]][[7.06, 7.06, 7.06]][0.24]
******
==============


O_threshold = -3
MC for this TARGET:[90.118, 0.074]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.23, -2.32, -2.89]][[-2.13, -2.28, -2.38]][[-90.12, -90.12, -90.12]][[-2.98, -4.6]]
std:[[0.22, 0.17, 0.08]][[0.1, 0.05, 0.1]][[0.0, 0.0, 0.0]][[0.04, 0.03]]
MSE:[[2.24, 2.33, 2.89]][[2.13, 2.28, 2.38]][[90.12, 90.12, 90.12]][[2.98, 4.6]]
MSE(-DR):[[0.0, 0.09, 0.65]][[-0.11, 0.04, 0.14]][[87.88, 87.88, 87.88]][[0.74, 2.36]]
MC-based ATE = -5.99
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.14, -2.07, -1.88]][[-1.99, -1.9, -2.0]][[5.99, 5.99, 5.99]][-1.81]
std:[[0.15, 0.1, 0.17]][[0.05, 0.04, 0.02]][[0.0, 0.0, 0.0]][0.11]
MSE:[[2.15, 2.07, 1.89]][[1.99, 1.9, 2.0]][[5.99, 5.99, 5.99]][1.81]
MSE(-DR):[[0.0, -0.08, -0.26]][[-0.16, -0.25, -0.15]][[3.84, 3.84, 3.84]][-0.34]
==============


O_threshold = -2
MC for this TARGET:[85.221, 0.076]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.56, 0.59, 0.26]][[-0.99, -1.01, -1.06]][[-85.22, -85.22, -85.22]][[0.29, 0.3]]
std:[[0.01, 0.03, 0.18]][[0.09, 0.05, 0.15]][[0.0, 0.0, 0.0]][[0.14, 0.03]]
MSE:[[0.56, 0.59, 0.32]][[0.99, 1.01, 1.07]][[85.22, 85.22, 85.22]][[0.32, 0.3]]
MSE(-DR):[[0.0, 0.03, -0.24]][[0.43, 0.45, 0.51]][[84.66, 84.66, 84.66]][[-0.24, -0.26]]
better than DR_NO_MARL
MC-based ATE = -10.89
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.64, 0.83, 1.26]][[-0.85, -0.63, -0.69]][[10.89, 10.89, 10.89]][1.46]
std:[[0.06, 0.1, 0.26]][[0.05, 0.03, 0.02]][[0.0, 0.0, 0.0]][0.21]
MSE:[[0.64, 0.84, 1.29]][[0.85, 0.63, 0.69]][[10.89, 10.89, 10.89]][1.48]
MSE(-DR):[[0.0, 0.2, 0.65]][[0.21, -0.01, 0.05]][[10.25, 10.25, 10.25]][0.84]
******
==============


O_threshold = -1
MC for this TARGET:[86.748, 0.073]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.38, -3.34, -3.85]][[-3.02, -3.04, -3.05]][[-86.75, -86.75, -86.75]][[-3.81, -1.23]]
std:[[0.34, 0.27, 0.26]][[0.01, 0.05, 0.07]][[0.0, 0.0, 0.0]][[0.18, 0.03]]
MSE:[[3.4, 3.35, 3.86]][[3.02, 3.04, 3.05]][[86.75, 86.75, 86.75]][[3.81, 1.23]]
MSE(-DR):[[0.0, -0.05, 0.46]][[-0.38, -0.36, -0.35]][[83.35, 83.35, 83.35]][[0.41, -2.17]]
MC-based ATE = -9.36
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.29, -3.1, -2.84]][[-2.88, -2.66, -2.68]][[9.36, 9.36, 9.36]][-2.64]
std:[[0.28, 0.19, 0.34]][[0.15, 0.14, 0.06]][[0.0, 0.0, 0.0]][0.26]
MSE:[[3.3, 3.11, 2.86]][[2.88, 2.66, 2.68]][[9.36, 9.36, 9.36]][2.65]
MSE(-DR):[[0.0, -0.19, -0.44]][[-0.42, -0.64, -0.62]][[6.06, 6.06, 6.06]][-0.65]
==============


O_threshold = 90
MC for this TARGET:[95.768, 0.082]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.62, 0.49, 0.05]][[-0.11, -0.31, -0.25]][[-95.77, -95.77, -95.77]][[-0.08, -10.25]]
std:[[0.08, 0.07, 0.2]][[0.14, 0.09, 0.09]][[0.0, 0.0, 0.0]][[0.19, 0.03]]
MSE:[[0.63, 0.49, 0.21]][[0.18, 0.32, 0.27]][[95.77, 95.77, 95.77]][[0.21, 10.25]]
MSE(-DR):[[0.0, -0.14, -0.42]][[-0.45, -0.31, -0.36]][[95.14, 95.14, 95.14]][[-0.42, 9.62]]
MC-based ATE = -0.34
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.7, 0.73, 1.06]][[0.03, 0.07, 0.12]][[0.34, 0.34, 0.34]][1.09]
std:[[0.14, 0.14, 0.11]][[0.0, 0.0, 0.04]][[0.0, 0.0, 0.0]][0.11]
MSE:[[0.71, 0.74, 1.07]][[0.03, 0.07, 0.13]][[0.34, 0.34, 0.34]][1.1]
MSE(-DR):[[0.0, 0.03, 0.36]][[-0.68, -0.64, -0.58]][[-0.37, -0.37, -0.37]][0.39]
==============


time spent until now: 9.0 mins
```