

```

Last login: Wed Apr  1 13:42:54 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:~ ssh mac$
Run-Mac:~ ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-18-204-44-50.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 disabled due to load higher than 36.0

 * 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.

*** System restart required ***
Last login: Wed Apr  1 17:42:57 2020 from 107.13.161.147
ubuntu@ip-172-31-9-82:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
14:58, 04/01; num of cores:36

Basic setting:[T, sd_0, sd_D, sd_R, sd_u_0, w_0, w_A, [M_in_R, mean_reversion, pois0, simple, u_0_u_D]] = [672, 10, 10, 20, 0.3, 0.5, 1,
[True, False, True, False, 10]]

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

max(u_0) = 197.9
Q_threshold = 70
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

1 1 0 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

number of reward locations: 22
Q_threshold = 80
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations: 15
Q_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

```

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations: 12

0_threshold = 100

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations: 8

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

Value of Behaviour policy:60.68

0_threshold = 70

MC for this TARGET:[69.326, 0.229]

[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]

bias:[[2.01, 1.89, 1.18]][[5.01, -69.33, -8.65]]

std:[[1.74, 1.72, 0.55]][[0.47, 0.0, 0.25]]

MSE:[[2.66, 2.56, 1.3]][[5.03, 69.33, 8.65]]

MSE(-DR):[[0.0, -0.1, -1.36]][[2.37, 66.67, 5.99]]

=====

0_threshold = 80

MC for this TARGET:[70.887, 0.239]

[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]

bias:[[0.58, 0.4, -1.01]][[1.27, -70.89, -10.21]]

std:[[1.13, 1.08, 0.73]][[0.57, 0.0, 0.25]]

MSE:[[1.27, 1.15, 1.25]][[1.39, 70.89, 10.21]]

MSE(-DR):[[0.0, -0.12, -0.02]][[0.12, 69.62, 8.94]]

=====

0_threshold = 90

MC for this TARGET:[69.375, 0.232]

[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]

bias:[[-0.66, -0.81, -1.34]][[-0.56, -69.38, -8.7]]

std:[[1.41, 1.35, 1.01]][[0.48, 0.0, 0.25]]

MSE:[[1.56, 1.57, 1.68]][[0.74, 69.38, 8.7]]

MSE(-DR):[[0.0, 0.01, 0.12]][[-0.82, 67.82, 7.14]]

=====

0_threshold = 100

MC for this TARGET:[68.943, 0.229]

[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]

bias:[[-3.06, -3.19, -3.56]][[-4.85, -68.94, -8.26]]

std:[[0.87, 0.88, 0.49]][[0.45, 0.0, 0.25]]

MSE:[[3.18, 3.31, 3.59]][[4.87, 68.94, 8.26]]

MSE(-DR):[[0.0, 0.13, 0.41]][[1.69, 65.76, 5.08]]

=====

[[2.66 2.56 1.3 5.03 69.33 8.65]
[1.27 1.15 1.25 1.39 70.89 10.21]
[1.56 1.57 1.68 0.74 69.38 8.7]
[3.18 3.31 3.59 4.87 68.94 8.26]]

time spent until now: 19.8 mins

[pattern_seed, day] = [2, 9]

```

max(u_0) = 197.9
0_threshold = 70
means of Order:

87.8 97.8 52.4 162.7 58.1

77.3 115.7 68.5 72.4 75.7

117.4 197.9 100.7 71.1 116.9

83.2 98.9 141.5 79.5 99.8

76.4 94.9 107.4 73.9 89.9

target policy:

1 1 0 1 0

1 1 0 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

number of reward locations: 22
0_threshold = 80
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations: 15
0_threshold = 90
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations: 12
0_threshold = 100
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations: 8
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

-----
Value of Behaviour policy:60.743
0_threshold = 70
MC for this TARGET:[69.301, 0.187]
[DR_QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.96, 2.82, 1.8]][[5.27, -69.3, -8.56]]
std:[[1.42, 1.42, 0.83]][[0.3, 0.0, 0.19]]
MSE:[[3.28, 3.16, 1.98]][[5.28, 69.3, 8.56]]

```

```
MSE(-DR):[[0.0, -0.12, -1.3]][[2.0, 66.02, 5.28]]
```

```
***
```

```
=====
```

```
Q_threshold = 80
```

```
MC for this TARGET:[70.866, 0.187]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[0.11, -0.06, -0.96]][[1.24, -70.87, -10.12]]
```

```
std:[[1.11, 1.11, 0.65]][[0.35, 0.0, 0.19]]
```

```
MSE:[[1.12, 1.11, 1.16]][[1.29, 70.87, 10.12]]
```

```
MSE(-DR):[[0.0, -0.01, 0.04]][[0.17, 69.75, 9.0]]
```

```
***
```

```
=====
```

```
Q_threshold = 90
```

```
MC for this TARGET:[69.351, 0.193]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[-0.54, -0.68, -1.25]][[-0.51, -69.35, -8.61]]
```

```
std:[[0.66, 0.69, 0.53]][[0.44, 0.0, 0.19]]
```

```
MSE:[[0.85, 0.97, 1.36]][[0.67, 69.35, 8.61]]
```

```
MSE(-DR):[[0.0, 0.12, 0.51]][[-0.18, 68.5, 7.76]]
```

```
=====
```

```
Q_threshold = 100
```

```
MC for this TARGET:[68.922, 0.198]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[-2.77, -2.89, -3.35]][[-4.9, -68.92, -8.18]]
```

```
std:[[0.81, 0.82, 0.33]][[0.42, 0.0, 0.19]]
```

```
MSE:[[2.89, 3.0, 3.37]][[4.92, 68.92, 8.18]]
```

```
MSE(-DR):[[0.0, 0.11, 0.48]][[2.03, 66.03, 5.29]]
```

```
***
```

```
=====
```

```
[[ 2.66  2.56  1.3   5.03 69.33  8.65]
 [ 1.27  1.15  1.25  1.39 70.89 10.21]
 [ 1.56  1.57  1.68  0.74 69.38  8.7 ]
 [ 3.18  3.31  3.59  4.87 68.94  8.26]]
```

```
[[ 3.28  3.16  1.98  5.28 69.3   8.56]
 [ 1.12  1.11  1.16  1.29 70.87 10.12]
 [ 0.85  0.97  1.36  0.67 69.35  8.61]
 [ 2.89  3.    3.37  4.92 68.92  8.18]]
```

```
time spent until now: 39.8 mins
```

```
-----
[pattern_seed, day] = [2, 11]
```

```
max(u_0) = 197.9
```

```
Q_threshold = 70
```

```
means of Order:
```

```
87.8 97.8 52.4 162.7 58.1
```

```
77.3 115.7 68.5 72.4 75.7
```

```
117.4 197.9 100.7 71.1 116.9
```

```
83.2 98.9 141.5 79.5 99.8
```

```
76.4 94.9 107.4 73.9 89.9
```

```
target policy:
```

```
1 1 0 1 0
```

```
1 1 0 1 1
```

```
1 1 1 1 1
```

```
1 1 1 1 1
```

```
1 1 1 1 1
```

```
number of reward locations: 22
```

```
Q_threshold = 80
```

```
target policy:
```

```
1 1 0 1 0
```

```
0 1 0 0 0
```

1 1 1 0 1

1 1 1 0 1

0 1 1 0 1

number of reward locations: 15

0_threshold = 90

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations: 12

0_threshold = 100

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations: 8

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

1 -th target; 2 -th target; 3 -th target; 4 -th target; one rep DONE

Value of Behaviour policy:60.641

0_threshold = 70

MC for this TARGET:[69.321, 0.162]

[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]

bias:[[2.99, 2.89, 1.5]][[5.22, -69.32, -8.68]]

std:[[1.1, 1.13, 0.54]][[0.3, 0.0, 0.13]]

MSE:[[3.19, 3.1, 1.59]][[5.23, 69.32, 8.68]]

MSE(-DR):[[0.0, -0.09, -1.6]][[2.04, 66.13, 5.49]]

=====

0_threshold = 80

MC for this TARGET:[70.888, 0.164]

[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]

bias:[[0.77, 0.57, -0.73]][[1.41, -70.89, -10.25]]

std:[[0.89, 0.88, 0.43]][[0.39, 0.0, 0.13]]

MSE:[[1.18, 1.05, 0.85]][[1.46, 70.89, 10.25]]

MSE(-DR):[[0.0, -0.13, -0.33]][[0.28, 69.71, 9.07]]

=====

0_threshold = 90

MC for this TARGET:[69.374, 0.168]

[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]

bias:[[0.13, 0.01, -0.99]][[-0.4, -69.37, -8.73]]

std:[[0.7, 0.73, 0.39]][[0.42, 0.0, 0.13]]

MSE:[[0.71, 0.73, 1.06]][[0.58, 69.37, 8.73]]

MSE(-DR):[[0.0, 0.02, 0.35]][[-0.13, 68.66, 8.02]]

=====

0_threshold = 100

MC for this TARGET:[68.93, 0.166]

[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]

bias:[[-2.48, -2.6, -3.31]][[-4.84, -68.93, -8.29]]

std:[[0.93, 0.92, 0.49]][[0.5, 0.0, 0.13]]

MSE:[[2.65, 2.76, 3.35]][[4.87, 68.93, 8.29]]

MSE(-DR):[[0.0, 0.11, 0.7]][[2.22, 66.28, 5.64]]

=====

```
[[ 2.66  2.56  1.3   5.03 69.33  8.65]
 [ 1.27  1.15  1.25  1.39 70.89 10.21]
 [ 1.56  1.57  1.68  0.74 69.38  8.7 ]
 [ 3.18  3.31  3.59  4.87 68.94  8.26]]
```

```
[[ 3.28  3.16  1.98  5.28 69.3   8.56]
 [ 1.12  1.11  1.16  1.29 70.87 10.12]
 [ 0.85  0.97  1.36  0.67 69.35  8.61]
 [ 2.89  3.    3.37  4.92 68.92  8.18]]
```

```
[[ 3.19  3.1   1.59  5.23 69.32  8.68]
 [ 1.18  1.05  0.85  1.46 70.89 10.25]
 [ 0.71  0.73  1.06  0.58 69.37  8.73]
 [ 2.65  2.76  3.35  4.87 68.93  8.29]]
```

time spent until now: 60.0 mins

```
-----
[pattern_seed, day] = [2, 14]
```

```
max(u_0) = 197.9
0_threshold = 70
means of Order:
```

```
87.8 97.8 52.4 162.7 58.1
```

```
77.3 115.7 68.5 72.4 75.7
```

```
117.4 197.9 100.7 71.1 116.9
```

```
83.2 98.9 141.5 79.5 99.8
```

```
76.4 94.9 107.4 73.9 89.9
```

target policy:

```
1 1 0 1 0
```

```
1 1 0 1 1
```

```
1 1 1 1 1
```

```
1 1 1 1 1
```

```
1 1 1 1 1
```

number of reward locations: 22

```
0_threshold = 80
```

target policy:

```
1 1 0 1 0
```

```
0 1 0 0 0
```

```
1 1 1 0 1
```

```
1 1 1 0 1
```

```
0 1 1 0 1
```

number of reward locations: 15

```
0_threshold = 90
```

target policy:

```
0 1 0 1 0
```

```
0 1 0 0 0
```

```
1 1 1 0 1
```

```
0 1 1 0 1
```

```
0 1 1 0 0
```

number of reward locations: 12

```
0_threshold = 100
```

target policy:

```
0 0 0 1 0
```

```
0 1 0 0 0
```

```
1 1 1 0 1
```

0 0 1 0 0

0 0 1 0 0

number of reward locations: 8
1 -th target; 2 -th target; 3 -th target; Connection to ec2-18-204-44-50.compute-1.amazonaws.com closed by remote host.
Connection to ec2-18-204-44-50.compute-1.amazonaws.com closed.
Run-Mac:ssh mac\$