

```
Last login: Mon Apr 20 18:43:29 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:~.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-91-2-13.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 96.0

 * 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

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

*** System restart required ***
Last login: Mon Apr 20 22:44:50 2020 from 107.13.161.147
ubuntu@ip-172-31-0-252:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
19:34, 04/20; num of cores:96
final_sd

Basic setting:[rep_times, sd_0, sd_D, sd_u_0, w_0, w_A, u_0_range, t_func] = [96, None, None, 25, 0.5, 1.5, [00], None]

[thre_range, sd_R_range, day_range, penalty_range]: [[100, 101, 105, 110], [0, 5, 10, 15, 20, 25, 30], [7], [[0.0003, 0.0001, 5e-05], [0.0003, 0.0001, 5e-05]]]

-----
[pattern_seed, day, sd_R, u_D] = [2, 7, 0, 80]

max(u_0) = 157.3 mean(u_0) = 93.7
0_threshold = 100
means of Order:

89.6 98.6 46.6 141.0 55.2
79.0 112.6 68.9 73.6 77.3
113.8 157.3 101.0 72.1 113.5
85.1 99.5 129.4 81.3 100.2
78.0 96.1 106.4 75.3 91.5

target policy:

0 0 0 1 0
0 1 0 0 0
1 1 1 0 1
0 0 1 0 1
0 0 1 0 0

number of reward locations: 9
0_threshold = 101
number of reward locations: 8
0_threshold = 105
number of reward locations: 7
0_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:50.623
0_threshold = 100
MC for this TARGET:[57.979, 0.063]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.64, -1.86, -2.17]][[-2.91, -57.98, -7.36]]
std:[0.51, 0.52, 0.34][0.29, 0.0, 0.23]]
sd_MSE:[array([0.19, 0.21, 0.15])][array([0.18, 0. , 0.35])]
MSE:[12.95, 3.73, 4.82][0.55, 3361.68, 54.22]]
RMSE(-DR):[[0.0, 0.78, 1.87]][[5.6, 3358.73, 51.27]]
MC
=====
0_threshold = 101
MC for this TARGET:[55.156, 0.061]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[0.00, -0.09, -0.68][[-2.05, -55.16, -4.53]]
std:[0.5, 0.5, 0.32][0.3, 0.0, 0.23]]
sd_MSE:[array([0.03, 0.03, 0.05])][array([0.13, 0. , 0.21])]
MSE:[10.26, 0.26, 0.56][[4.21, 3042.63, 28.57]]
RMSE(-DR):[[0.0, 0.0, 0.3]][[3.95, 3042.37, 28.31]]
MC
=====
0_threshold = 105
MC for this TARGET:[56.95, 0.061]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.44, -3.62, -3.96]][[-5.38, -56.95, -6.33]]
std:[0.61, 0.62, 0.36][0.31, 0.0, 0.23]]
sd_MSE:[array([0.43, 0.46, 0.29])][array([0.34, 0. , 0.3 ])]
MSE:[12.21, 13.49, 15.81][[29.04, 3243.3, 40.12]]
RMSE(-DR):[[0.0, 1.28, 3.6]][[16.83, 3231.09, 27.91]]
MC
=====
0_threshold = 110
MC for this TARGET:[55.7, 0.059]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.22, -3.35, -3.82]][[-6.26, -55.7, -5.08]]
std:[0.61, 0.63, 0.39][0.31, 0.0, 0.23]]
sd_MSE:[array([0.41, 0.44, 0.3 ])]array([0.41, 0. , 0.24])]
MSE:[110.74, 11.62, 14.74][[39.28, 3102.49, 25.86]]
RMSE(-DR):[[0.0, 0.89, 4.0]][[28.54, 3091.75, 15.12]]
MC
=====
***** THIS SETTING IS GOOD *****
[[2.9500e+00 3.7300e+00 4.8200e+00 8.5500e+00 3.3617e+03 5.4220e+01]
[2.6000e-01 2.6000e-01 5.6000e-01 4.2100e+00 3.0426e+03 2.0570e+01]
[1.2210e+01 1.3490e+01 1.5810e+01 2.9040e+01 3.2433e+03 4.0120e+01]
[1.0740e+01 1.1620e+01 1.4740e+01 3.9280e+01 3.1025e+03 2.5860e+01]]

time spent until now: 80.0 mins

20:53, 04/20

-----
[pattern_seed, day, sd_R, u_D] = [2, 7, 5, 80]

max(u_0) = 157.3 mean(u_0) = 93.7
0_threshold = 100
means of Order:

89.6 98.6 46.6 141.0 55.2
79.0 112.6 68.9 73.6 77.3
113.8 157.3 101.0 72.1 113.5
85.1 99.5 129.4 81.3 100.2
78.0 96.1 106.4 75.3 91.5

target policy:

0 0 0 1 0
0 1 0 0 0
1 1 1 0 1
```

```
0 0 1 0 1
0 0 1 0 0

number of reward locations: 9
Q_threshold = 101
number of reward locations: 8
Q_threshold = 105
number of reward locations: 7
Q_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:50.63
Q_threshold = 100
MC for this TARGET:[57.987, 0.077]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.61, -1.82, -2.15]][[-2.92, -57.99, -7.36]]
std:[10.57, 0.58, 0.36]][[0.33, 0.0, 0.23]]
sd_MSE:[array([0.2, 0.23, 0.16])][array([0.2, 0. , 0.35])]
MSE:[12.92, 3.65, 4.75]][[0.64, 3362.84, 54.22]]
RMSE(-DR):[[0.0, 0.73, 1.83]][[5.72, 3359.92, 51.3]]
***

=====
Q_threshold = 101
MC for this TARGET:[55.164, 0.076]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[0.1, -0.08, -0.68]][[-2.05, -55.16, -4.53]]
std:[0.56, 0.55, 0.35]][[0.33, 0.0, 0.23]]
sd_MSE:[array([0.04, 0.05, 0.05])][array([0.14, 0. , 0.22])]
MSE:[10.32, 0.31, 0.58]][[4.31, 3042.63, 20.57]]
RMSE(-DR):[[0.0, -0.01, 0.26]][[3.99, 3042.31, 20.25]]
***

=====
Q_threshold = 105
MC for this TARGET:[56.957, 0.075]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.44, -3.61, -3.96]][[-5.39, -56.96, -6.33]]
std:[10.67, 0.67, 0.39]][[0.33, 0.0, 0.23]]
sd_MSE:[array([0.5, 0.52, 0.32])][array([0.37, 0. , 0.3 ])]
MSE:[12.28, 13.48, 15.83]][[29.16, 3244.44, 40.12]]
RMSE(-DR):[[0.0, 1.2, 3.55]][[16.88, 3232.16, 27.84]]
***

=====
Q_threshold = 110
MC for this TARGET:[55.708, 0.072]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.24, -3.36, -3.83]][[-6.27, -55.71, -5.08]]
std:[10.67, 0.68, 0.44]][[0.33, 0.0, 0.23]]
sd_MSE:[array([0.45, 0.47, 0.34])][array([0.42, 0. , 0.24])]
MSE:[110.95, 11.75, 14.86]][[39.42, 3103.0, 25.86]]
RMSE(-DR):[[0.0, 0.8, 3.91]][[28.47, 3092.65, 14.91]]
***

=====
***** THIS SETTING IS GOOD *****
[[2.9500e+00 3.7300e+00 4.8200e+00 8.5500e+00 3.3617e+03 5.4220e+01]
[2.6000e-01 2.6000e-01 5.6000e-01 4.2100e+00 3.0426e+03 2.0570e+01]
[1.2210e+01 1.3490e+01 1.5810e+01 2.9040e+01 3.2433e+03 4.0120e+01]
[1.0740e+01 1.1620e+01 1.4740e+01 3.9280e+01 3.1025e+03 2.5860e+01]]

[[2.9200e+00 3.6500e+00 4.7500e+00 8.6400e+00 3.3628e+03 5.4220e+01]
[3.2000e-01 3.1000e-01 5.8000e-01 4.3100e+00 3.0426e+03 2.0570e+01]
[1.2280e+01 1.3480e+01 1.5830e+01 2.9160e+01 3.2444e+03 4.0120e+01]
[1.0950e+01 1.1750e+01 1.4860e+01 3.9420e+01 3.1036e+03 2.5860e+01]]

time spent until now: 158.9 mins

22:12, 04/20

-----
[pattern_seed, day, sd_R, u_0] = [2, 7, 10, 80]

max(u_0) = 157.3 mean(u_0) = 93.7
Q_threshold = 100
means of Order:

89.6 98.6 46.6 141.0 55.2
79.0 112.6 68.9 73.6 77.3

113.8 157.3 101.0 72.1 113.5
85.1 99.5 129.4 81.3 100.2
78.0 96.1 106.4 75.3 91.5

target policy:

0 0 0 1 0
0 1 0 0 0
1 1 1 0 1
0 0 1 0 1
0 0 1 0 0

number of reward locations: 9
Q_threshold = 101
number of reward locations: 8
Q_threshold = 105
number of reward locations: 7
Q_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:50.636
Q_threshold = 100
MC for this TARGET:[57.995, 0.121]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.56, -1.77, -2.15]][[-2.92, -58.0, -7.36]]
std:[10.71, 0.7, 0.46]][[0.4, 0.0, 0.25]]
sd_MSE:[array([0.25, 0.27, 0.2 ])] [array([0.24, 0. , 0.37])]
MSE:[12.94, 3.62, 4.83]][[0.69, 3364.0, 54.23]]
RMSE(-DR):[[0.0, 0.68, 1.89]][[5.75, 3361.06, 51.29]]
***

=====
Q_threshold = 101
MC for this TARGET:[55.172, 0.12]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[0.14, -0.03, -0.67]][[-2.05, -55.17, -4.54]]
std:[10.71, 0.69, 0.44]][[0.4, 0.0, 0.25]]
sd_MSE:[array([0.07, 0.07, 0.07])][array([0.17, 0. , 0.23])]
MSE:[10.52, 0.48, 0.64]][[4.36, 3043.73, 20.67]]
RMSE(-DR):[[0.0, -0.04, 0.12]][[3.04, 3043.21, 20.15]]
***

=====
Q_threshold = 105
MC for this TARGET:[56.965, 0.119]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.41, -3.57, -3.96]][[-5.41, -56.96, -6.33]]
std:[10.82, 0.81, 0.51]][[0.39, 0.0, 0.25]]
sd_MSE:[array([0.61, 0.62, 0.41])][array([0.43, 0. , 0.32])]
MSE:[12.3, 13.4, 15.93]][[29.42, 3244.44, 40.13]]
RMSE(-DR):[[0.0, 1.1, 3.63]][[17.12, 3232.14, 27.83]]
***

=====
Q_threshold = 110
MC for this TARGET:[55.716, 0.117]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.24, -3.34, -3.83]][[-6.28, -55.72, -5.08]]
std:[10.83, 0.81, 0.55]][[0.38, 0.0, 0.25]]
sd_MSE:[array([0.55, 0.56, 0.43])][array([0.5, 0. , 0.26])]
MSE:[111.19, 11.81, 14.97]][[39.38, 3104.72, 25.87]]
RMSE(-DR):[[0.0, 0.62, 3.78]][[28.39, 3093.53, 14.68]]
***

=====
```

```
***** THIS SETTING IS GOOD *****
[[2.9500e+00 3.7300e+00 4.8200e+00 8.5500e+00 3.3617e+03 5.4220e+01]
 [2.6000e-01 2.6000e-01 5.6000e-01 4.2100e+00 3.0426e+03 2.0570e+01]
 [1.2210e+01 1.3490e+01 1.5810e+01 2.9040e+01 3.2433e+03 4.0120e+01]
 [1.0740e+01 1.1620e+01 1.4740e+01 3.9280e+01 3.1025e+03 2.5860e+01]]
```

```
[[2.9200e+00 3.6500e+00 4.7500e+00 8.6400e+00 3.3628e+03 5.4220e+01]
 [3.2000e-01 3.1000e-01 5.8000e-01 4.3100e+00 3.0426e+03 2.0570e+01]
 [1.2280e+01 1.3480e+01 1.5830e+01 2.9160e+01 3.2444e+03 4.0120e+01]
 [1.0950e+01 1.1750e+01 1.4860e+01 3.9420e+01 3.1036e+03 2.5860e+01]]
```

```
[[2.9400e+00 3.6200e+00 4.8300e+00 8.6900e+00 3.3640e+03 5.4230e+01]
 [5.2000e-01 4.8000e-01 6.4000e-01 4.3600e+00 3.0437e+03 2.0670e+01]
 [1.2300e+01 1.3400e+01 1.5930e+01 2.9420e+01 3.2444e+03 4.0130e+01]
 [1.1190e+01 1.1810e+01 1.4970e+01 3.9580e+01 3.1047e+03 2.5870e+01]]
```

time spent until now: 235.2 mins

23:29, 04/20

[pattern\_seed, day, sd\_R, u\_0] = [2, 7, 15, 80]

max(u\_0) = 157.3 mean(u\_0) = 93.7  
0\_threshold = 100  
means of Order:

89.6 98.6 46.6 141.0 55.2

79.0 112.6 68.9 73.6 77.3

113.8 157.3 101.0 72.1 113.5

85.1 99.5 129.4 81.3 100.2

78.0 96.1 106.4 75.3 91.5

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 1

0 0 1 0 0

number of reward locations: 9

0\_threshold = 101

number of reward locations: 8

0\_threshold = 105

number of reward locations: 7

0\_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:50.643

0\_threshold = 100

MC for this TARGET:[58.003, 0.174]

[DR/QV/IS]: [DR\_NO\_MARL, DR\_NO\_MF, V\_behav]

bias:[[-1.54, -1.74, -2.14]][[-2.93, -58.0, -7.36]]

std:[0.9, 0.88, 0.57]][0.5, 0.0, 0.28]]

sd\_MSE:[array[0.32, 0.34, 0.26]][array[0.31, 0. , 0.41]]]

MSE:[0.18, 3.8, 4.9]][0.83, 3364.0, 54.25]]

RMSE(-DR):[0.0, 0.62, 1.72]][[5.65, 3360.82, 51.07]]

MC

0\_threshold = 101

MC for this TARGET:[55.18, 0.172]

[DR/QV/IS]: [DR\_NO\_MARL, DR\_NO\_MF, V\_behav]

bias:[0.15, -0.03, -0.67]][[-2.06, -55.18, -4.54]]

std:[0.91, 0.89, 0.55]][0.5, 0.0, 0.28]]

sd\_MSE:[array[0.11, 0.11, 0.09]][array[0.22, 0. , 0.25]]]

MSE:[0.85, 0.79, 0.75]][[4.49, 3044.83, 20.69]]

RMSE(-DR):[0.0, -0.06, -0.1]][[3.64, 3043.98, 19.84]]

MC

0\_threshold = 105

MC for this TARGET:[56.973, 0.171]

[DR/QV/IS]: [DR\_NO\_MARL, DR\_NO\_MF, V\_behav]

bias:[[-3.41, -3.57, -3.98]][[-5.42, -56.97, -6.33]]

std:[1.0, 0.99, 0.62]][0.49, 0.0, 0.28]]

sd\_MSE:[array[0.73, 0.76, 0.51]][array[0.54, 0. , 0.36]]]

MSE:[12.63, 13.72, 16.22]][[29.62, 3245.58, 40.15]]

RMSE(-DR):[0.0, 1.09, 3.59]][[16.99, 3232.95, 27.52]]

MC

0\_threshold = 110

MC for this TARGET:[55.724, 0.169]

[DR/QV/IS]: [DR\_NO\_MARL, DR\_NO\_MF, V\_behav]

bias:[[-3.26, -3.37, -3.84]][[-6.29, -55.72, -5.08]]

std:[1.02, 1.0, 0.66]][[0.48, 0.0, 0.28]]

sd\_MSE:[array[0.67, 0.69, 0.51]][array[0.61, 0. , 0.28]]]

MSE:[11.67, 12.36, 15.18]][[39.79, 3104.72, 25.88]]

RMSE(-DR):[0.0, 0.69, 3.51]][[26.12, 3093.05, 14.21]]

MC

```
*****
[[2.9500e+00 3.7300e+00 4.8200e+00 8.5500e+00 3.3617e+03 5.4220e+01]
 [2.6000e-01 2.6000e-01 5.6000e-01 4.2100e+00 3.0426e+03 2.0570e+01]
 [1.2210e+01 1.3490e+01 1.5810e+01 2.9040e+01 3.2433e+03 4.0120e+01]
 [1.0740e+01 1.1620e+01 1.4740e+01 3.9280e+01 3.1025e+03 2.5860e+01]]
```

```
[[2.9200e+00 3.6500e+00 4.7500e+00 8.6400e+00 3.3628e+03 5.4220e+01]
 [3.2000e-01 3.1000e-01 5.8000e-01 4.3100e+00 3.0426e+03 2.0570e+01]
 [1.2280e+01 1.3480e+01 1.5830e+01 2.9160e+01 3.2444e+03 4.0120e+01]
 [1.0950e+01 1.1750e+01 1.4860e+01 3.9420e+01 3.1036e+03 2.5860e+01]]
```

```
[[2.9400e+00 3.6200e+00 4.8300e+00 8.6900e+00 3.3640e+03 5.4230e+01]
 [5.2000e-01 4.8000e-01 6.4000e-01 4.3600e+00 3.0437e+03 2.0670e+01]
 [1.2300e+01 1.3400e+01 1.5930e+01 2.9420e+01 3.2444e+03 4.0130e+01]
 [1.1190e+01 1.1810e+01 1.4970e+01 3.9580e+01 3.1047e+03 2.5870e+01]]
```

```
[[3.1800e+00 3.8000e+00 4.9000e+00 8.8300e+00 3.3640e+03 5.4250e+01]
 [8.5000e-01 7.9000e-01 7.5000e-01 4.4900e+00 3.0448e+03 2.0690e+01]
 [1.2630e+01 1.3720e+01 1.6220e+01 2.9620e+01 3.2456e+03 4.0150e+01]
 [1.1670e+01 1.2360e+01 1.5180e+01 3.9790e+01 3.1047e+03 2.5860e+01]]
```

time spent until now: 313.8 mins

00:47, 04/21

[pattern\_seed, day, sd\_R, u\_0] = [2, 7, 20, 80]

max(u\_0) = 157.3 mean(u\_0) = 93.7  
0\_threshold = 100  
means of Order:

89.6 98.6 46.6 141.0 55.2

79.0 112.6 68.9 73.6 77.3

113.8 157.3 101.0 72.1 113.5

85.1 99.5 129.4 81.3 100.2

78.0 96.1 106.4 75.3 91.5

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

```
0 0 1 0 1
0 0 1 0 0

number of reward locations: 9
0_threshold = 101
number of reward locations: 8
0_threshold = 105
number of reward locations: 7
0_threshold = 110
number of reward locations: 6
target 1 in 4 DONE!
^[[Atarget 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

-----
Value of Behaviour policy:50.649
0_threshold = 100
MC for this TARGET:[58.011, 0.228]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.5, -1.72, -2.12]][[-2.93, -58.01, -7.36]]
std:[[1.11, 1.07, 0.73]][[0.62, 0.0, 0.31]]
sd_MSE:[array([0.4, 0.41, 0.32])][array([0.38, 0. , 0.46])]
MSE:[[3.48, 4.1, 5.01]][[8.97, 3365.16, 54.27]]
RMSE(-DR):[[0.0, 0.62, 1.53]][[5.49, 3361.68, 50.79]]
+-----+
=====
0_threshold = 101
MC for this TARGET:[55.188, 0.227]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.16, -0.02, -0.67]][[-2.07, -55.19, -4.54]]
std:[[1.11, 1.07, 0.7]][[0.62, 0.0, 0.31]]
sd_MSE:[array([0.18, 0.16, 0.13])][array([0.26, 0. , 0.29])]
MSE:[[1.26, 1.15, 0.94]][[4.67, 3045.94, 20.71]]
RMSE(-DR):[[0.0, -0.11, -0.32]][[3.41, 3044.68, 19.45]]
+-----+
=====
0_threshold = 105
MC for this TARGET:[56.981, 0.226]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.41, -3.58, -3.98]][[-5.43, -56.98, -6.33]]
std:[[1.22, 1.18, 0.77]][[0.59, 0.0, 0.31]]
sd_MSE:[array([0.9, 0.9, 0.65])][array([0.66, 0. , 0.4 ])]
MSE:[[13.12, 14.21, 16.43]][[29.83, 3246.72, 40.16]]
RMSE(-DR):[[0.0, 1.09, 3.31]][[16.71, 3233.6, 27.04]]
+-----+
=====
0_threshold = 110
MC for this TARGET:[55.731, 0.224]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.27, -3.38, -3.85]][[-6.29, -55.73, -5.08]]
std:[[1.24, 1.22, 0.83]][[0.57, 0.0, 0.31]]
sd_MSE:[array([0.81, 0.82, 0.64])][array([0.73, 0. , 0.32])]
MSE:[[12.23, 12.91, 15.51]][[39.89, 3305.83, 25.9]]
RMSE(-DR):[[0.0, 0.68, 3.28]][[27.66, 3093.6, 13.67]]
+-----+
=====
[[2.9500e+00 3.7300e+00 4.8200e+00 8.5500e+00 3.3617e+03 5.4220e+01]
[2.6000e-01 2.6000e-01 5.6000e-01 4.2100e+00 3.0426e+03 2.0570e+01]
[1.2210e+01 1.3490e+01 1.5810e+01 2.9040e+01 3.2433e+03 4.0120e+01]
[1.0740e+01 1.1620e+01 1.4740e+01 3.9280e+01 3.1025e+03 2.5860e+01]]

[[2.9200e+00 3.6500e+00 4.7500e+00 8.6400e+00 3.3628e+03 5.4220e+01]
[3.2000e-01 3.1000e-01 5.8000e-01 4.3100e+00 3.0426e+03 2.0570e+01]
[1.2280e+01 1.3480e+01 1.5830e+01 2.9160e+01 3.2444e+03 4.0120e+01]
[1.0950e+01 1.1750e+01 1.4860e+01 3.9420e+01 3.1036e+03 2.5860e+01]]

[[2.9400e+00 3.6200e+00 4.8300e+00 8.6900e+00 3.3640e+03 5.4230e+01]
[5.2000e-01 4.8000e-01 6.4000e-01 4.3600e+00 3.0437e+03 2.0670e+01]
[1.2300e+01 1.3400e+01 1.5930e+01 2.9420e+01 3.2444e+03 4.0130e+01]
[1.1190e+01 1.1810e+01 1.4970e+01 3.9580e+01 3.1047e+03 2.5870e+01]]

[[3.1800e+00 3.8000e+00 4.9000e+00 8.8300e+00 3.3640e+03 5.4250e+01]
[8.5000e-01 7.9000e-01 7.5000e-01 4.4900e+00 3.0448e+03 2.0690e+01]
[1.2630e+01 1.3720e+01 1.6220e+01 2.9620e+01 3.2456e+03 4.0150e+01]
[1.1670e+01 1.2360e+01 1.5180e+01 3.9790e+01 3.1047e+03 2.5880e+01]]

[[3.4800e+00 4.1000e+00 5.0100e+00 8.9700e+00 3.3652e+03 5.4270e+01]
[1.2600e+00 1.1500e+00 9.4000e-01 4.6700e+00 3.0459e+03 2.0710e+01]
[1.3120e+01 1.4210e+01 1.6430e+01 2.9830e+01 3.2467e+03 4.0160e+01]
[1.2230e+01 1.2910e+01 1.5510e+01 3.9890e+01 3.1058e+03 2.5900e+01]]

time spent until now: 393.2 mins

02:07, 04/21

-----
[pattern_seed, day, sd_R, u_D] = [2, 7, 25, 80]

max(u_0) = 157.3 mean(u_0) = 93.7
0_threshold = 100
means of Order:

89.6 98.6 46.6 141.0 55.2

79.0 112.6 68.9 73.6 77.3

113.8 157.3 101.0 72.1 113.5

85.1 99.5 129.4 81.3 100.2

78.0 96.1 106.4 75.3 91.5

target policy:

0 0 0 1 0
0 1 0 0 0
1 1 1 0 1
0 0 1 0 1
0 0 1 0 0

number of reward locations: 9
0_threshold = 101
number of reward locations: 8
0_threshold = 105
number of reward locations: 7
0_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:50.655
0_threshold = 100
MC for this TARGET:[58.019, 0.284]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.48, -1.7, -2.12]][[-2.94, -58.02, -7.36]]
std:[[1.31, 1.25, 0.86]][[0.74, 0.0, 0.35]]
sd_MSE:[array([0.48, 0.49, 0.39])][array([0.45, 0. , 0.53])]
MSE:[[3.31, 4.45, 5.23]][[9.19, 3366.32, 54.29]]
RMSE(-DR):[[0.0, 0.54, 1.32]][[5.28, 3362.41, 50.38]]
+-----+
=====
0_threshold = 101
MC for this TARGET:[55.195, 0.284]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.18, 0.0, -0.67]][[-2.08, -55.2, -4.54]]
std:[[1.34, 1.26, 0.83]][[0.74, 0.0, 0.35]]
sd_MSE:[array([0.25, 0.22, 0.16])][array([0.32, 0. , 0.32])]
MSE:[[1.83, 1.64, 1.14]][[4.87, 3047.04, 20.73]]
RMSE(-DR):[[0.0, -0.19, -0.69]][[3.04, 3045.21, 18.9]]
+-----+
=====
0_threshold = 105
MC for this TARGET:[56.989, 0.282]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.4, -3.57, -3.98]]][[-5.44, -56.99, -6.33]]
std:[11.43, 1.39, 0.9]]][[0.71, 0.0, 0.35]]
sd_MSE:[array([1.06, 3.06, 0.76])][array([0.78, 0. , 0.45])]
MSE:[13.6, 14.68, 16.65]]][[30.1, 3247.86, 40.19]]
RMSE(-DR):[0.0, 1.08, 3.05]]][[16.5, 3234.26, 26.59]]
#####
0_threshold = 110
MC for this TARGET:[55.739, 0.281]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.3, -3.39, -3.68]]][[-6.3, -55.74, -5.08]]
std:[11.5, 1.44, 1.0]]][[0.68, 0.0, 0.35]]
sd_MSE:[array([0.98, 0.98, 0.77])][array([0.87, 0. , 0.36])]
MSE:[13.14, 13.57, 16.07]]][[40.15, 3106.95, 25.93]]
RMSE(-DR):[0.0, 0.43, 2.93]]][[27.01, 3093.81, 12.79]]
#####
[[2.9500e+00 3.7300e+00 4.8200e+00 8.5500e+00 3.3617e+03 5.4220e+01]
[2.6000e-01 2.6000e-01 5.0000e-01 4.2100e+00 3.0426e+03 2.0570e+01]
[1.2210e+01 1.3490e+01 1.5810e+01 2.9040e+01 3.2433e+03 4.0120e+01]
[1.0740e+01 1.1620e+01 1.4740e+01 3.9280e+01 3.1025e+03 2.5860e+01]]

[[2.9200e+00 3.6500e+00 4.7500e+00 8.6400e+00 3.3628e+03 5.4220e+01]
[3.2000e-01 3.1000e-01 5.8000e-01 4.3100e+00 3.0426e+03 2.0570e+01]
[1.2280e+01 1.3480e+01 1.5830e+01 2.9160e+01 3.2444e+03 4.0120e+01]
[1.0950e+01 1.1750e+01 1.4860e+01 3.9420e+01 3.1036e+03 2.5860e+01]]

[[2.9400e+00 3.6200e+00 4.8300e+00 8.6900e+00 3.3640e+03 5.4230e+01]
[5.2000e-01 4.8000e-01 6.4000e-01 4.3600e+00 3.0427e+03 2.0670e+01]
[1.2300e+01 1.3400e+01 1.5930e+01 2.9420e+01 3.2444e+03 4.0130e+01]
[1.1190e+01 1.1810e+01 1.4970e+01 3.9580e+01 3.1047e+03 2.5870e+01]]

[[3.1800e+00 3.8000e+00 4.9000e+00 8.8300e+00 3.3640e+03 5.4250e+01]
[8.5000e-01 7.9000e-01 7.5000e-01 4.4900e+00 3.0448e+03 2.0690e+01]
[1.2630e+01 1.3720e+01 1.6220e+01 2.9620e+01 3.2456e+03 4.0150e+01]
[1.1670e+01 1.2360e+01 1.5180e+01 3.9790e+01 3.1047e+03 2.5880e+01]]

[[3.4800e+00 4.1000e+00 5.0100e+00 8.9700e+00 3.3652e+03 5.4270e+01]
[1.2600e+00 1.1500e+00 9.4000e-01 4.6700e+00 3.0459e+03 2.0710e+01]
[1.3120e+01 1.4210e+01 1.6430e+01 2.9830e+01 3.2467e+03 4.0160e+01]
[1.2230e+01 1.2910e+01 1.5510e+01 3.9890e+01 3.1058e+03 2.5900e+01]]

[[3.9100e+00 4.4500e+00 5.2300e+00 9.1900e+00 3.3663e+03 5.4290e+01]
[1.8300e+00 1.6400e+00 1.1400e+00 4.8700e+00 3.0470e+03 2.0730e+01]
[1.3600e+01 1.4680e+01 1.6650e+01 3.0100e+01 3.2479e+03 4.0190e+01]
[1.3140e+01 1.3570e+01 1.6070e+01 4.0150e+01 3.1069e+03 2.5930e+01]]

time spent until now: 471.9 mins

03:25, 04/21

[pattern_seed, day, sd_R, u_D] = [2, 7, 30, 80]

max(u_D) = 157.3 mean(u_D) = 93.7
0_threshold = 100
means of Order:

89.6 98.6 46.6 141.0 55.2

79.0 112.6 68.9 73.6 77.3

113.8 157.3 101.0 72.1 113.5

85.1 99.5 129.4 81.3 100.2

78.0 96.1 106.4 75.3 91.5

target policy:

0 0 0 1 0
0 1 0 0 0
1 1 1 0 1
0 0 1 0 1
0 0 1 0 0

number of reward locations: 9
0_threshold = 101
number of reward locations: 8
0_threshold = 105
number of reward locations: 7
0_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:50.662
0_threshold = 100
MC for this TARGET:[58.827, 0.341]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.44, -1.66, -2.12]]][[-2.95, -58.03, -7.37]]
std:[11.52, 1.46, 1.0]]][[0.87, 0.0, 0.39]]
sd_MSE:[array([0.58, 0.59, 0.46])][array([0.53, 0. , 0.59])]
MSE:[4.38, 4.89, 5.40]]][[5.46, 3367.48, 34.47]]
RMSE(-DR):[0.0, 0.51, 1.11]]][[5.08, 3363.1, 50.09]]
#####
0_threshold = 101
MC for this TARGET:[55.203, 0.34]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[0.22, 0.03, -0.64]]][[-2.09, -55.2, -4.54]]
std:[11.55, 1.40, 0.99]]][[0.87, 0.0, 0.39]]
sd_MSE:[array([0.34, 0.3 , 0.2 ])]][array([0.38, 0. , 0.37])]
MSE:[2.45, 2.22, 1.39]]][[5.12, 3047.04, 20.76]]
RMSE(-DR):[0.0, -0.23, -1.06]]][[2.67, 3044.59, 18.31]]
#####
0_threshold = 105
MC for this TARGET:[56.997, 0.339]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.4, -3.56, -4.0]]][[-5.45, -57.0, -6.34]]
std:[11.65, 1.61, 1.05]]][[0.83, 0.0, 0.39]]
sd_MSE:[array([1.23, 1.21, 0.9 ])]][array([0.92, 0. , 0.51])]
MSE:[14.20, 15.27, 17.11]]][[38.39, 3249.0, 40.35]]
RMSE(-DR):[0.0, 0.99, 2.82]]][[16.11, 3234.72, 26.07]]
#####
0_threshold = 110
MC for this TARGET:[55.747, 0.337]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.29, -3.38, -3.88]]][[-6.3, -55.75, -5.09]]
std:[11.74, 1.67, 1.16]]][[0.8, 0.0, 0.39]]
sd_MSE:[array([1.12, 1.12, 0.89])][array([1.02, 0. , 0.41])]
MSE:[13.85, 14.21, 16.4]]][[40.33, 3108.06, 26.06]]
RMSE(-DR):[0.0, 0.36, 2.55]]][[26.46, 3094.21, 12.21]]
#####
[[2.9500e+00 3.7300e+00 4.8200e+00 8.5500e+00 3.3617e+03 5.4220e+01]
[2.6000e-01 2.6000e-01 5.0000e-01 4.2100e+00 3.0426e+03 2.0570e+01]
[1.2210e+01 1.3490e+01 1.5810e+01 2.9040e+01 3.2433e+03 4.0120e+01]
[1.0740e+01 1.1620e+01 1.4740e+01 3.9280e+01 3.1025e+03 2.5860e+01]]

[[2.9200e+00 3.6500e+00 4.7500e+00 8.6400e+00 3.3628e+03 5.4220e+01]
[3.2000e-01 3.1000e-01 5.8000e-01 4.3100e+00 3.0426e+03 2.0570e+01]
[1.2280e+01 1.3480e+01 1.5830e+01 2.9160e+01 3.2444e+03 4.0120e+01]
[1.0950e+01 1.1750e+01 1.4860e+01 3.9420e+01 3.1036e+03 2.5860e+01]]

[[2.9400e+00 3.6200e+00 4.8300e+00 8.6900e+00 3.3640e+03 5.4230e+01]
[5.2000e-01 4.8000e-01 6.4000e-01 4.3600e+00 3.0427e+03 2.0670e+01]
[1.2300e+01 1.3400e+01 1.5930e+01 2.9420e+01 3.2444e+03 4.0130e+01]
[1.1190e+01 1.1810e+01 1.4970e+01 3.9580e+01 3.1047e+03 2.5870e+01]]
```

```
[[3.1800e+00 3.8000e+00 4.9000e+00 8.8300e+00 3.3640e+03 5.4250e+01]
[8.5000e-01 7.9000e-01 7.5000e-01 4.4900e+00 3.0448e+03 2.0690e+01]
[1.2630e+01 1.3720e+01 1.6220e+01 2.9620e+01 3.2456e+03 4.0150e+01]
[1.1670e+01 1.2360e+01 1.5180e+01 3.9790e+01 3.1047e+03 2.5880e+01]]

[[3.4900e+00 4.1000e+00 5.0100e+00 8.9700e+00 3.3652e+03 5.4270e+01]
[1.2600e+00 1.1500e+00 9.4000e-01 4.6700e+00 3.0459e+03 2.0710e+01]
[1.3120e+01 1.4210e+01 1.6430e+01 2.9830e+01 3.2467e+03 4.0160e+01]
[1.2230e+01 1.2910e+01 1.5510e+01 3.9890e+01 3.1058e+03 2.5900e+01]]

[[3.9100e+00 4.4500e+00 5.2300e+00 9.1900e+00 3.3663e+03 5.4290e+01]
[1.8300e+00 1.6400e+00 1.1400e+00 4.8700e+00 3.0470e+03 2.0730e+01]
[1.3600e+01 1.4680e+01 1.6650e+01 3.0100e+01 3.2479e+03 4.0190e+01]
[1.3140e+01 1.3570e+01 1.6070e+01 4.0150e+01 3.1069e+03 2.5930e+01]]

[[4.3900e+00 4.8900e+00 5.4900e+00 9.4600e+00 3.3675e+03 5.4470e+01]
[2.4500e+00 2.2200e+00 1.3900e+00 5.1200e+00 3.0470e+03 2.0760e+01]
[1.4280e+01 1.5270e+01 1.7100e+01 3.0390e+01 3.2490e+03 4.0350e+01]
[1.3850e+01 1.4210e+01 1.6400e+01 4.0330e+01 3.1081e+03 2.6060e+01]]
```

time spent until now: 552.2 mins

04:46, 04/21

ubuntu@ip-172-31-8-252:~\$