```
Last login: Thu Apr  2 09:09:00 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-3-230-143-148.compute-1.amazonaws.com
^[[B^C
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-3-230-143-148.compute-1.amazonaws.com
^C
Run-Mac:.ssh mac$ sh -i "Runzhe_Song_0110.pem" ubuntu@ec2-35-175-103-246.compute-1.amazonaws.com
Runzhe_Song_0110.pem: line 1: -----BEGIN: command not found
: command not foundm: line 2: MIIEowIBAAKCAQEA1a8Z6XTRvFtXlJCtVtdLC7EJRch5gY02aSgk0dcjcFaGkrbxuMHcRwtV5GQi
: No such file or directory3: 8pWEeHpJyrLb4fdFA6BreQtO+Lsd5yWlidf/sTj4Uw1nN8WvmBIiNgfLLBC1RjRVUWg2/ePgytH6
: No such file or directory4: hOoPbIAIU/68mhzV3aLAC1hTI4beFFPB9tmSCJOkBzlt6YTL/sMJGaVF4n2tPKBotbzZERHKSmAv
: command not foundm: line 5: 7CL7BmZfUJaHS4xwXjz4IQUHBHw8WOrw7Rxykbnok1GlbRfUCdbt1MxK5TLPU1NEsViChmAmZZEZ
: No such file or directory6: WhiP396F7Kw/TklYOghE7bbkvRFRvXesgSIU4XH8jRmFuIH4kBw4fQIDAQABAoIBACEiH5B1jrf4
: No such file or directory7: KeJddKNMitRrRRbO+Rg/fzpbASqQg2w93mwNdfnCBZf2iAD8Dfcze/D6dUq65SFqqAbh7wg72qjy
: No such file or directory8: yrY2vi5bkcg3Hzoxj9m0C4A2/5sFJlnU1WbnYntW/3/qv8ZzPLczW+TlDWWI0FHNOtQ9Ei4wKeQO
: command not foundm: line 9: W40nV6Ig3bY3of7iHAeKMDbG+cGqaj+5dais4km0TXrvzgN5FETE+0LWGnQU3vONPGbGj40lnsLU
: command not foundm: line 10: RLzWVe+Uv2kpyF40ZIKMjB5qugl3bgTudhMmA03h7FOLMTFvHDGApNFtCLEbPriVZPUgcQnDvlnT
: command not foundm: line 11: Zc35IKjtnu6m3ydJxBHtVp0qXgECgYEA+Z0TrhYSt81tBb7MpJd9KKj3pi2HYL8G29tqUdi9sPGp
: No such file or directory12: O49mR7WIV051MWz7qM291jk5H7ZcQTDPbi1egwurO6nxEmpsvl5oEv0/gzmFjHC+aAJm0a4/qzPj
: command not foundm: line 13: HI1pUW5YUtvmxaFv9MiFb6n2eztaqC2iAlLi+U7MufundQGMYkECgYEA2yaxLihEt5RnYxH7hLAA
: command not foundm: line 14: afSoAHSOmkNuNnb4NjUYpwmPKQYBa4pYU4yHWdjYWipJXAMPuzccsultcHesoIeUnBJtx9iUPpKe
: command not foundm: line 15: rTC+Ktt6BVeYsASD3iWziXEbvyXfNaYVDIusv8b2k7RqItvw4Ka4CCdoS3vSIPc8ViZSXQAnDz0C
: command not foundm: line 16: gYEA2M4ZGNztNx5DcnkN0GwP9rAEIh9rdZmWQIIsr5hc8oKEJkpgYSqImKR54Asz7qN9HsOnHSdM
: No such file or directory17: rAbUNHNNou8QtIrf68nqV9Mx/ERg1P6yQ5BtoXNskPUTMgWoXtVow6TfLGD56XBWCLveQmXSw6iD
: No such file or directory18: Do5MKE6L/FT4NxYgld1ZbMECgYBY2EnSZZ7eMfrsDhlCxLuWOQp3DMnXPmkLCZY9G69/4H86aWXf
: No such file or directory19: /DksvfcNdLzZaPVOqSsSghhPbBTjGJFSj0SvNUVuVJ5c88ICgpVF+K5yRf3vthTj1bMAkGtoRelE
: No such file or directory20: 7azw5htFsFq8rNKuiS4RVWWcYKWGnI8msDS01Fplu68aHQKBgAHgLRNVpnXafJ/ycOaO28pu0Yyu
: command not foundm: line 21: 4u1IOhm6pVMyY6gf0Mz175AUBzdG8+3cqVMDraBj2tte9+XwcSiMHcSwyDSoxM3A0nQcsgf7F2Ja
: No such file or directory22: WvpYfoCAl4KK5zl092z/uUULJKGVy21ZtTQsJtcUbZSnMHN1GoTXBIFIZbkxl5aDoOlc
Runzhe_Song_0110.pem: line 23: -----END: command not found
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-35-175-103-246.compute-1.amazonaws.com
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1063-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Thu Apr  2 15:32:05 UTC 2020

  System load:  5.44               Processes:           370
  Usage of /:   56.3% of 15.45GB   Users logged in:     0
  Memory usage: 0%                 IP address for ens5: 172.31.77.47
  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.


Last login: Thu Apr  2 13:09:10 2020 from 107.13.161.147
ubuntu@ip-172-31-77-47:~$ tmux -t a 0
tmux: unknown option -- t
usage: tmux [-2CluvV] [-c shell-command] [-f file] [-L socket-name]
            [-S socket-path] [command [flags]]
ubuntu@ip-172-31-77-47:~$ tmux -a t 0
tmux: unknown option -- a
usage: tmux [-2CluvV] [-c shell-command] [-f file] [-L socket-name]
            [-S socket-path] [command [flags]]
ubuntu@ip-172-31-77-47:~$ tmux -t 0
tmux: unknown option -- t
usage: tmux [-2CluvV] [-c shell-command] [-f file] [-L socket-name]
            [-S socket-path] [command [flags]]
ubuntu@ip-172-31-77-47:~$ tmux
[detached (from session 0)]
ubuntu@ip-172-31-77-47:~$ tmux a -t 0
```

```
[detached (from session 0)]
ubuntu@ip-172-31-77-47:~$ python EC22.py
11:37, 04/02; num of cores:36

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, simple, u_O_u_D]] = [N
one, 96, 10, 10, None, 0.3, 0.5, 1, [True, False, True, False, 10]]


--------------------------------------
[pattern_seed, lam, day] = [2, 0.01, 7]

max(u_O) =  197.9
O_threshold = 80
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

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
O_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
O_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
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0
```

```
number of reward locations:  6
1 -th region DONE!
^CTraceback (most recent call last):
  File "EC22.py", line 74, in <module>
    print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 60, in simu
    value_reps = parmap(once, range(OPE_rep_times), n_cores)
  File "/home/ubuntu/_uti_basic.py", line 74, in parmap
    sent = [q_in.put((i, x)) for i, x in enumerate(X)]
  File "/home/ubuntu/_uti_basic.py", line 74, in <listcomp>
    sent = [q_in.put((i, x)) for i, x in enumerate(X)]
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
ubuntu@ip-172-31-77-47:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC22.py
11:39, 04/02; num of cores:36

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, simple, u_O_u_D]] = [N
one, 96, 10, 10, None, 0.3, 0.5, 1, [True, False, True, False, 10]]


---------------------------------------
[pattern_seed, lam, day] = [2, 0.01, 7]

max(u_O) =  197.9
O_threshold = 80
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

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
O_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
O_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
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
```

```
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!

--------------------------------------
Value of Behaviour policy:60.786
O_threshold = 80
MC for this TARGET:[70.884, 0.141]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.88, -4.61, -70.88]][[-70.88, -70.88, -10.1]]
std:[[0.0, 0.42, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.88, 4.63, 70.88]][[70.88, 70.88, 10.1]]
MSE(-DR):[[0.0, -66.25, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.371, 0.133]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -3.86, -69.37]][[-69.37, -69.37, -8.59]]
std:[[0.0, 0.41, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[69.37, 3.88, 69.37]][[69.37, 69.37, 8.59]]
MSE(-DR):[[0.0, -65.49, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 100
MC for this TARGET:[68.94, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -4.98, -68.94]][[-68.94, -68.94, -8.15]]
std:[[0.0, 0.39, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[68.94, 5.0, 68.94]][[68.94, 68.94, 8.15]]
MSE(-DR):[[0.0, -63.94, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 110
MC for this TARGET:[70.484, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.48, -8.4, -70.48]][[-70.48, -70.48, -9.7]]
```

std:[[0.0, 0.5, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.48, 8.41, 70.48]][[70.48, 70.48, 9.7]]
MSE(-DR):[[0.0, -62.07, 0.0]][[0.0, 0.0, -60.78]]
***
==========

********** THIS SETTING IS GOOD **********
[[70.88 4.63 70.88 70.88 70.88 10.1]
 [69.37 3.88 69.37 69.37 69.37 8.59]
 [68.94 5.   68.94 68.94 68.94 8.15]
 [70.48 8.41 70.48 70.48 70.48 9.7]]

time spent until now: 1.0 mins

[pattern_seed, lam, day] = [2, 0.001, 7]

max(u_O) = 197.9
O_threshold = 80
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

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
O_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
O_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
O_threshold = 110
target policy:

```
0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
```

```
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!

----------------------------------------
Value of Behaviour policy:60.786
O_threshold = 80
MC for this TARGET:[70.884, 0.141]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.88, -1.23, -70.88]][[-70.88, -70.88, -10.1]]
std:[[0.0, 0.52, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.88, 1.34, 70.88]][[70.88, 70.88, 10.1]]
MSE(-DR):[[0.0, -69.54, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.371, 0.133]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -1.19, -69.37]][[-69.37, -69.37, -8.59]]
std:[[0.0, 0.55, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[69.37, 1.31, 69.37]][[69.37, 69.37, 8.59]]
MSE(-DR):[[0.0, -68.06, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 100
MC for this TARGET:[68.94, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -3.44, -68.94]][[-68.94, -68.94, -8.15]]
std:[[0.0, 0.53, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[68.94, 3.48, 68.94]][[68.94, 68.94, 8.15]]
MSE(-DR):[[0.0, -65.46, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 110
MC for this TARGET:[70.484, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.48, -7.23, -70.48]][[-70.48, -70.48, -9.7]]
std:[[0.0, 0.61, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.48, 7.26, 70.48]][[70.48, 70.48, 9.7]]
MSE(-DR):[[0.0, -63.22, 0.0]][[0.0, 0.0, -60.78]]
***
==============
```

```
******************** THIS SETTING IS GOOD ********************
[[70.88  4.63 70.88 70.88 70.88 10.1 ]
 [69.37  3.88 69.37 69.37 69.37  8.59]
 [68.94  5.   68.94 68.94 68.94  8.15]
 [70.48  8.41 70.48 70.48 70.48  9.7 ]]


[[70.88  1.34 70.88 70.88 70.88 10.1 ]
 [69.37  1.31 69.37 69.37 69.37  8.59]
 [68.94  3.48 68.94 68.94 68.94  8.15]
 [70.48  7.26 70.48 70.48 70.48  9.7 ]]


time spent until now: 2.1 mins


--------------------------------------
[pattern_seed, lam, day] = [2, 0.0001, 7]

max(u_O) =  197.9
O_threshold = 80
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

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
O_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
O_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
O_threshold = 110
target policy:
```

```
0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
```

```
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!

---------------------------------------
Value of Behaviour policy:60.786
O_threshold = 80
MC for this TARGET:[70.884, 0.141]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.88, 0.13, -70.88]][[-70.88, -70.88, -10.1]]
std:[[0.0, 0.63, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.88, 0.64, 70.88]][[70.88, 70.88, 10.1]]
MSE(-DR):[[0.0, -70.24, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.371, 0.133]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -0.28, -69.37]][[-69.37, -69.37, -8.59]]
std:[[0.0, 0.65, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[69.37, 0.71, 69.37]][[69.37, 69.37, 8.59]]
MSE(-DR):[[0.0, -68.66, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 100
MC for this TARGET:[68.94, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -2.88, -68.94]][[-68.94, -68.94, -8.15]]
std:[[0.0, 0.66, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[68.94, 2.95, 68.94]][[68.94, 68.94, 8.15]]
MSE(-DR):[[0.0, -65.99, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 110
MC for this TARGET:[70.484, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.48, -6.53, -70.48]][[-70.48, -70.48, -9.7]]
std:[[0.0, 0.66, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.48, 6.56, 70.48]][[70.48, 70.48, 9.7]]
MSE(-DR):[[0.0, -63.92, 0.0]][[0.0, 0.0, -60.78]]
***
```

==========

********** THIS SETTING IS GOOD **********
[[70.88 4.63 70.88 70.88 70.88 10.1]
 [69.37 3.88 69.37 69.37 69.37 8.59]
 [68.94 5.   68.94 68.94 68.94 8.15]
 [70.48 8.41 70.48 70.48 70.48 9.7]]

[[70.88 1.34 70.88 70.88 70.88 10.1]
 [69.37 1.31 69.37 69.37 69.37 8.59]
 [68.94 3.48 68.94 68.94 68.94 8.15]
 [70.48 7.26 70.48 70.48 70.48 9.7]]

[[70.88 0.64 70.88 70.88 70.88 10.1]
 [69.37 0.71 69.37 69.37 69.37 8.59]
 [68.94 2.95 68.94 68.94 68.94 8.15]
 [70.48 6.56 70.48 70.48 70.48 9.7]]

time spent until now: 3.1 mins

[pattern_seed, lam, day] = [2, 1e-05, 7]

max(u_O) = 197.9
O_threshold = 80
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

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
O_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
O_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
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
```

```
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!

----------------------------------------
Value of Behaviour policy:60.786
O_threshold = 80
MC for this TARGET:[70.884, 0.141]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.88, 0.28, -70.88]][[-70.88, -70.88, -10.1]]
std:[[0.0, 0.71, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.88, 0.76, 70.88]][[70.88, 70.88, 10.1]]
MSE(-DR):[[0.0, -70.12, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.371, 0.133]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -0.09, -69.37]][[-69.37, -69.37, -8.59]]
std:[[0.0, 0.73, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[69.37, 0.74, 69.37]][[69.37, 69.37, 8.59]]
MSE(-DR):[[0.0, -68.63, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 100
MC for this TARGET:[68.94, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -2.56, -68.94]][[-68.94, -68.94, -8.15]]
std:[[0.0, 0.72, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[68.94, 2.66, 68.94]][[68.94, 68.94, 8.15]]
MSE(-DR):[[0.0, -66.28, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 110
```

```
MC for this TARGET:[70.484, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.48, -6.12, -70.48]][[-70.48, -70.48, -9.7]]
std:[[0.0, 0.81, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.48, 6.17, 70.48]][[70.48, 70.48, 9.7]]
MSE(-DR):[[0.0, -64.31, 0.0]][[0.0, 0.0, -60.78]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.88  4.63 70.88 70.88 70.88 10.1 ]
 [69.37  3.88 69.37 69.37 69.37  8.59]
 [68.94  5.   68.94 68.94 68.94  8.15]
 [70.48  8.41 70.48 70.48 70.48  9.7 ]]


[[70.88  1.34 70.88 70.88 70.88 10.1 ]
 [69.37  1.31 69.37 69.37 69.37  8.59]
 [68.94  3.48 68.94 68.94 68.94  8.15]
 [70.48  7.26 70.48 70.48 70.48  9.7 ]]


[[70.88  0.64 70.88 70.88 70.88 10.1 ]
 [69.37  0.71 69.37 69.37 69.37  8.59]
 [68.94  2.95 68.94 68.94 68.94  8.15]
 [70.48  6.56 70.48 70.48 70.48  9.7 ]]


[[70.88  0.76 70.88 70.88 70.88 10.1 ]
 [69.37  0.74 69.37 69.37 69.37  8.59]
 [68.94  2.66 68.94 68.94 68.94  8.15]
 [70.48  6.17 70.48 70.48 70.48  9.7 ]]


time spent until now: 4.2 mins


---------------------------------------
[pattern_seed, lam, day] = [2, 0.01, 10]

max(u_O) =  197.9
O_threshold = 80
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

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
O_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
O_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
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
```

```
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!

--------------------------------------
Value of Behaviour policy:60.792
O_threshold = 80
MC for this TARGET:[70.887, 0.092]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, -4.62, -70.89]][[-70.89, -70.89, -10.09]]
std:[[0.0, 0.38, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.89, 4.64, 70.89]][[70.89, 70.89, 10.09]]
MSE(-DR):[[0.0, -66.25, 0.0]][[0.0, 0.0, -60.8]]
***
==============


O_threshold = 90
MC for this TARGET:[69.373, 0.094]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -3.79, -69.37]][[-69.37, -69.37, -8.58]]
std:[[0.0, 0.37, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[69.37, 3.81, 69.37]][[69.37, 69.37, 8.58]]
MSE(-DR):[[0.0, -65.56, 0.0]][[0.0, 0.0, -60.79]]
***
==============
```

```
O_threshold = 100
MC for this TARGET:[68.936, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -4.94, -68.94]][[-68.94, -68.94, -8.14]]
std:[[0.0, 0.37, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[68.94, 4.95, 68.94]][[68.94, 68.94, 8.14]]
MSE(-DR):[[0.0, -63.99, 0.0]][[0.0, 0.0, -60.8]]
***
==============


O_threshold = 110
MC for this TARGET:[70.474, 0.102]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -8.3, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.39, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.47, 8.31, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -62.16, 0.0]][[0.0, 0.0, -60.79]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.88  4.63 70.88 70.88 70.88 10.1 ]
 [69.37  3.88 69.37 69.37 69.37  8.59]
 [68.94  5.   68.94 68.94 68.94  8.15]
 [70.48  8.41 70.48 70.48 70.48  9.7 ]]


[[70.88  1.34 70.88 70.88 70.88 10.1 ]
 [69.37  1.31 69.37 69.37 69.37  8.59]
 [68.94  3.48 68.94 68.94 68.94  8.15]
 [70.48  7.26 70.48 70.48 70.48  9.7 ]]


[[70.88  0.64 70.88 70.88 70.88 10.1 ]
 [69.37  0.71 69.37 69.37 69.37  8.59]
 [68.94  2.95 68.94 68.94 68.94  8.15]
 [70.48  6.56 70.48 70.48 70.48  9.7 ]]


[[70.88  0.76 70.88 70.88 70.88 10.1 ]
 [69.37  0.74 69.37 69.37 69.37  8.59]
 [68.94  2.66 68.94 68.94 68.94  8.15]
 [70.48  6.17 70.48 70.48 70.48  9.7 ]]


[[70.89  4.64 70.89 70.89 70.89 10.09]
 [69.37  3.81 69.37 69.37 69.37  8.58]
 [68.94  4.95 68.94 68.94 68.94  8.14]
 [70.47  8.31 70.47 70.47 70.47  9.68]]


time spent until now: 6.4 mins


--------------------------------------
[pattern_seed, lam, day] = [2, 0.001, 10]

max(u_O) =  197.9
O_threshold = 80
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
```

```
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
O_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
O_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
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
```

```
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!

--------------------------------------
Value of Behaviour policy:60.792
O_threshold = 80
MC for this TARGET:[70.887, 0.092]
```

```
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, -1.14, -70.89]][[-70.89, -70.89, -10.09]]
std:[[0.0, 0.49, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.89, 1.24, 70.89]][[70.89, 70.89, 10.09]]
MSE(-DR):[[0.0, -69.65, 0.0]][[0.0, 0.0, -60.8]]
***
==============


O_threshold = 90
MC for this TARGET:[69.373, 0.094]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -1.05, -69.37]][[-69.37, -69.37, -8.58]]
std:[[0.0, 0.47, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[69.37, 1.15, 69.37]][[69.37, 69.37, 8.58]]
MSE(-DR):[[0.0, -68.22, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.936, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -3.44, -68.94]][[-68.94, -68.94, -8.14]]
std:[[0.0, 0.45, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[68.94, 3.47, 68.94]][[68.94, 68.94, 8.14]]
MSE(-DR):[[0.0, -65.47, 0.0]][[0.0, 0.0, -60.8]]
***
==============


O_threshold = 110
MC for this TARGET:[70.474, 0.102]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -7.16, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.49, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.47, 7.18, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -63.29, 0.0]][[0.0, 0.0, -60.79]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.88  4.63 70.88 70.88 70.88 10.1 ]
 [69.37  3.88 69.37 69.37 69.37  8.59]
 [68.94  5.   68.94 68.94 68.94  8.15]
 [70.48  8.41 70.48 70.48 70.48  9.7 ]]


[[70.88  1.34 70.88 70.88 70.88 10.1 ]
 [69.37  1.31 69.37 69.37 69.37  8.59]
 [68.94  3.48 68.94 68.94 68.94  8.15]
 [70.48  7.26 70.48 70.48 70.48  9.7 ]]


[[70.88  0.64 70.88 70.88 70.88 10.1 ]
 [69.37  0.71 69.37 69.37 69.37  8.59]
 [68.94  2.95 68.94 68.94 68.94  8.15]
 [70.48  6.56 70.48 70.48 70.48  9.7 ]]


[[70.88  0.76 70.88 70.88 70.88 10.1 ]
 [69.37  0.74 69.37 69.37 69.37  8.59]
 [68.94  2.66 68.94 68.94 68.94  8.15]
 [70.48  6.17 70.48 70.48 70.48  9.7 ]]


[[70.89  4.64 70.89 70.89 70.89 10.09]
 [69.37  3.81 69.37 69.37 69.37  8.58]
 [68.94  4.95 68.94 68.94 68.94  8.14]
 [70.47  8.31 70.47 70.47 70.47  9.68]]


[[70.89  1.24 70.89 70.89 70.89 10.09]
 [69.37  1.15 69.37 69.37 69.37  8.58]
 [68.94  3.47 68.94 68.94 68.94  8.14]
 [70.47  7.18 70.47 70.47 70.47  9.68]]
```

time spent until now: 8.6 mins

[pattern_seed, lam, day] = [2, 0.0001, 10]

max(u_O) = 197.9
O_threshold = 80
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

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
O_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
O_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
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations: 6
1 -th region DONE!
2 -th region DONE!

```
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
```

```
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!

---------------------------------------
Value of Behaviour policy:60.792
O_threshold = 80
MC for this TARGET:[70.887, 0.092]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, 0.37, -70.89]][[-70.89, -70.89, -10.09]]
std:[[0.0, 0.56, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.89, 0.67, 70.89]][[70.89, 70.89, 10.09]]
MSE(-DR):[[0.0, -70.22, 0.0]][[0.0, 0.0, -60.8]]
***
==============


O_threshold = 90
MC for this TARGET:[69.373, 0.094]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -0.15, -69.37]][[-69.37, -69.37, -8.58]]
std:[[0.0, 0.52, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[69.37, 0.54, 69.37]][[69.37, 69.37, 8.58]]
MSE(-DR):[[0.0, -68.83, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.936, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -2.96, -68.94]][[-68.94, -68.94, -8.14]]
std:[[0.0, 0.57, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[68.94, 3.01, 68.94]][[68.94, 68.94, 8.14]]
MSE(-DR):[[0.0, -65.93, 0.0]][[0.0, 0.0, -60.8]]
***
==============


O_threshold = 110
MC for this TARGET:[70.474, 0.102]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -6.61, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.64, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.47, 6.64, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -63.83, 0.0]][[0.0, 0.0, -60.79]]
***
==============


********** THIS SETTING IS GOOD **********
[[70.88  4.63 70.88 70.88 70.88 10.1 ]
 [69.37  3.88 69.37 69.37 69.37  8.59]
 [68.94  5.   68.94 68.94 68.94  8.15]
 [70.48  8.41 70.48 70.48 70.48  9.7 ]]


[[70.88  1.34 70.88 70.88 70.88 10.1 ]
 [69.37  1.31 69.37 69.37 69.37  8.59]
 [68.94  3.48 68.94 68.94 68.94  8.15]
 [70.48  7.26 70.48 70.48 70.48  9.7 ]]
```

```
[[70.88  0.64 70.88 70.88 70.88 10.1 ]
 [69.37  0.71 69.37 69.37 69.37  8.59]
 [68.94  2.95 68.94 68.94 68.94  8.15]
 [70.48  6.56 70.48 70.48 70.48  9.7 ]]


[[70.88  0.76 70.88 70.88 70.88 10.1 ]
 [69.37  0.74 69.37 69.37 69.37  8.59]
 [68.94  2.66 68.94 68.94 68.94  8.15]
 [70.48  6.17 70.48 70.48 70.48  9.7 ]]


[[70.89  4.64 70.89 70.89 70.89 10.09]
 [69.37  3.81 69.37 69.37 69.37  8.58]
 [68.94  4.95 68.94 68.94 68.94  8.14]
 [70.47  8.31 70.47 70.47 70.47  9.68]]


[[70.89  1.24 70.89 70.89 70.89 10.09]
 [69.37  1.15 69.37 69.37 69.37  8.58]
 [68.94  3.47 68.94 68.94 68.94  8.14]
 [70.47  7.18 70.47 70.47 70.47  9.68]]


[[70.89  0.67 70.89 70.89 70.89 10.09]
 [69.37  0.54 69.37 69.37 69.37  8.58]
 [68.94  3.01 68.94 68.94 68.94  8.14]
 [70.47  6.64 70.47 70.47 70.47  9.68]]


time spent until now: 10.8 mins


---------------------------------------
[pattern_seed, lam, day] = [2, 1e-05, 10]

max(u_O) =  197.9
O_threshold = 80
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

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
O_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
O_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
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
```

```
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
5 -th region DONE!
6 -th region DONE!
7 -th region DONE!
8 -th region DONE!
9 -th region DONE!
10 -th region DONE!
11 -th region DONE!
12 -th region DONE!
13 -th region DONE!
14 -th region DONE!
15 -th region DONE!
16 -th region DONE!
17 -th region DONE!
18 -th region DONE!
19 -th region DONE!
20 -th region DONE!
21 -th region DONE!
22 -th region DONE!
23 -th region DONE!
24 -th region DONE!
25 -th region DONE!

--------------------------------------
Value of Behaviour policy:60.792
O_threshold = 80
MC for this TARGET:[70.887, 0.092]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, 0.51, -70.89]][[-70.89, -70.89, -10.09]]
std:[[0.0, 0.61, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.89, 0.8, 70.89]][[70.89, 70.89, 10.09]]
MSE(-DR):[[0.0, -70.09, 0.0]][[0.0, 0.0, -60.8]]
***
==============


O_threshold = 90
MC for this TARGET:[69.373, 0.094]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, 0.07, -69.37]][[-69.37, -69.37, -8.58]]
std:[[0.0, 0.59, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[69.37, 0.59, 69.37]][[69.37, 69.37, 8.58]]
MSE(-DR):[[0.0, -68.78, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.936, 0.097]
```

```
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -2.56, -68.94]][[-68.94, -68.94, -8.14]]
std:[[0.0, 0.61, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[68.94, 2.63, 68.94]][[68.94, 68.94, 8.14]]
MSE(-DR):[[0.0, -66.31, 0.0]][[0.0, 0.0, -60.8]]
***
==============


O_threshold = 110
MC for this TARGET:[70.474, 0.102]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -6.12, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.73, 0.0]][[0.0, 0.0, 0.21]]
MSE:[[70.47, 6.16, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -64.31, 0.0]][[0.0, 0.0, -60.79]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.88  4.63 70.88 70.88 70.88 10.1 ]
 [69.37  3.88 69.37 69.37 69.37  8.59]
 [68.94  5.   68.94 68.94 68.94  8.15]
 [70.48  8.41 70.48 70.48 70.48  9.7 ]]


[[70.88  1.34 70.88 70.88 70.88 10.1 ]
 [69.37  1.31 69.37 69.37 69.37  8.59]
 [68.94  3.48 68.94 68.94 68.94  8.15]
 [70.48  7.26 70.48 70.48 70.48  9.7 ]]


[[70.88  0.64 70.88 70.88 70.88 10.1 ]
 [69.37  0.71 69.37 69.37 69.37  8.59]
 [68.94  2.95 68.94 68.94 68.94  8.15]
 [70.48  6.56 70.48 70.48 70.48  9.7 ]]


[[70.88  0.76 70.88 70.88 70.88 10.1 ]
 [69.37  0.74 69.37 69.37 69.37  8.59]
 [68.94  2.66 68.94 68.94 68.94  8.15]
 [70.48  6.17 70.48 70.48 70.48  9.7 ]]


[[70.89  4.64 70.89 70.89 70.89 10.09]
 [69.37  3.81 69.37 69.37 69.37  8.58]
 [68.94  4.95 68.94 68.94 68.94  8.14]
 [70.47  8.31 70.47 70.47 70.47  9.68]]


[[70.89  1.24 70.89 70.89 70.89 10.09]
 [69.37  1.15 69.37 69.37 69.37  8.58]
 [68.94  3.47 68.94 68.94 68.94  8.14]
 [70.47  7.18 70.47 70.47 70.47  9.68]]


[[70.89  0.67 70.89 70.89 70.89 10.09]
 [69.37  0.54 69.37 69.37 69.37  8.58]
 [68.94  3.01 68.94 68.94 68.94  8.14]
 [70.47  6.64 70.47 70.47 70.47  9.68]]


[[70.89  0.8  70.89 70.89 70.89 10.09]
 [69.37  0.59 69.37 69.37 69.37  8.58]
 [68.94  2.63 68.94 68.94 68.94  8.14]
 [70.47  6.16 70.47 70.47 70.47  9.68]]


time spent until now: 13.1 mins

ubuntu@ip-172-31-77-47:~$
```