```
Last login: Thu Apr  2 00:30:48 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-35-175-103-246.compute-1.amazonaws.com
The authenticity of host 'ec2-35-175-103-246.compute-1.amazonaws.com (35.175.103.246)' can't be established.
ECDSA key fingerprint is SHA256:0j1bzeDeu0CkIMW5oeU8XCCKBuuLUtsvxspuDieuUd0.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-35-175-103-246.compute-1.amazonaws.com,35.175.103.246' (ECDSA) to the list of known hosts.
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 13:09:07 UTC 2020

  System load:  0.81               Processes:           386
  Usage of /:   57.1% 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

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


Last login: Wed Apr  1 20:30:39 2020 from 107.13.161.147
ubuntu@ip-172-31-77-47:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
09:10, 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, pois0, simple, u_O_u_D]] = [None, 36, 10, 10,
None, 0.3, 0.5, 1, [True, False, True, False, 10]]


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

---------------------------------------
Value of Behaviour policy:60.732
O_threshold = 80
MC for this TARGET:[70.884, 0.141]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.37, 0.22, -0.96]][[1.21, 30205.52, -10.15]]
std:[[0.67, 0.65, 0.39]][[0.39, 81102.9, 0.25]]
MSE:[[0.77, 0.69, 1.04]][[1.27, 86545.1, 10.15]]
MSE(-DR):[[0.0, -0.08, 0.27]][[0.5, 86544.33, 9.38]]
***
==============


O_threshold = 90
MC for this TARGET:[69.371, 0.133]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.12, -0.26, -1.03]][[-0.58, 10251.21, -8.64]]
std:[[0.74, 0.73, 0.47]][[0.36, 84750.31, 0.25]]
MSE:[[0.75, 0.77, 1.13]][[0.68, 85368.04, 8.64]]
MSE(-DR):[[0.0, 0.02, 0.38]][[-0.07, 85367.29, 7.89]]
==============


O_threshold = 100
MC for this TARGET:[68.94, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.78, -2.9, -3.38]][[-4.85, -77843.94, -8.21]]
std:[[0.67, 0.68, 0.4]][[0.36, 240260.01, 0.25]]
MSE:[[2.86, 2.98, 3.4]][[4.86, 252556.04, 8.21]]
MSE(-DR):[[0.0, 0.12, 0.54]][[2.0, 252553.18, 5.35]]
***
==============


O_threshold = 110
MC for this TARGET:[70.484, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.56, -6.64, -6.86]][[-8.85, -24358.88, -9.75]]
std:[[0.63, 0.64, 0.4]][[0.35, 175285.73, 0.25]]
MSE:[[6.59, 6.67, 6.87]][[8.86, 176970.17, 9.75]]
MSE(-DR):[[0.0, 0.08, 0.28]][[2.27, 176963.58, 3.16]]
***
==============


[[7.7000e-01 6.9000e-01 1.0400e+00 1.2700e+00 8.6545e+04 1.0150e+01]
 [7.5000e-01 7.7000e-01 1.1300e+00 6.8000e-01 8.5368e+04 8.6400e+00]
 [2.8600e+00 2.9800e+00 3.4000e+00 4.8600e+00 2.5256e+05 8.2100e+00]
```

```
 [6.5900e+00 6.6700e+00 6.8700e+00 8.8600e+00 1.7697e+05 9.7500e+00]]


time spent until now: 89.6 mins


---------------------------------------
[pattern_seed, day, sd_R] = [2, 10, 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
0 -th region DONE!
1 -th region DONE!
2 -th region DONE!
3 -th region DONE!
4 -th region DONE!
^CProcess Process-66:
Process Process-60:
Process Process-59:
Process Process-63:
Process Process-43:
Process Process-50:
```

```
Process Process-39:
Process Process-68:
Process Process-70:
Process Process-71:
Process Process-62:
Process Process-48:
Process Process-54:
Traceback (most recent call last):
  File "EC2.py", line 74, in <module>
Process Process-47:
Process Process-49:
Process Process-51:
    print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 60, in simu
Process Process-61:
Process Process-45:
Process Process-55:
Process Process-40:
    value_reps = parmap(once, range(OPE_rep_times), n_cores)
  File "/home/ubuntu/_uti_basic.py", line 75, in parmap
Process Process-57:
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/_uti_basic.py", line 75, in <listcomp>
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
Process Process-65:
Process Process-67:
Process Process-53:
Process Process-44:
Process Process-72:
Process Process-38:
Process Process-46:
Process Process-37:
Process Process-52:
Process Process-58:
Process Process-42:
Process Process-69:
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/main.py", line 136, in V_DR
    else:
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
  File "/home/ubuntu/main.py", line 51, in getOneRegionValue
    CV_QV = CV_QV, penalty_range = penalty, spatial = True)
  File "/home/ubuntu/main.py", line 266, in computeQV
    return computeQV_basic(tuples_i = tuples_i, R = R, penalty = penalty,
  File "/home/ubuntu/main.py", line 397, in computeQV_basic
    Kg = SA_GRBF(Z, gamma_g)
  File "/home/ubuntu/main.py", line 330, in SA_GRBF
    dim = int(Z.shape[1] // 2 - 1)
KeyboardInterrupt
Traceback (most recent call last):
Traceback (most recent call last):
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
Traceback (most recent call last):
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
Traceback (most recent call last):
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
Traceback (most recent call last):
Traceback (most recent call last):
Traceback (most recent call last):
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
```

```
    self.run()
Traceback (most recent call last):
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/main.py", line 136, in V_DR
    else:
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 136, in V_DR
    else:
Traceback (most recent call last):
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
  File "/home/ubuntu/main.py", line 59, in getOneRegionValue
    #                       epsilon = epsilon)
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
Traceback (most recent call last):
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 136, in V_DR
```

```
  else:
File "/home/ubuntu/main.py", line 59, in getOneRegionValue
  #                       epsilon = epsilon)
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
  self._target(*self._args, **self._kwargs)
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
  self._target(*self._args, **self._kwargs)
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
  self._target(*self._args, **self._kwargs)
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
  self._target(*self._args, **self._kwargs)
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/weight.py", line 301, in train
  self.policy_ratio2: policy_ratio2
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
  self.run()
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
  self.run()
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
  self.run()
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
  self._target(*self._args, **self._kwargs)
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/weight.py", line 301, in train
  self.policy_ratio2: policy_ratio2
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
  run_metadata_ptr)
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
  self._target(*self._args, **self._kwargs)
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
```

```
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
  self._target(*self._args, **self._kwargs)
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 59, in getOneRegionValue
  #                       epsilon = epsilon)
File "/home/ubuntu/weight.py", line 301, in train
  self.policy_ratio2: policy_ratio2
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
  run_metadata_ptr)
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1158, in _run
  self._graph, fetches, feed_dict_tensor, feed_handles=feed_handles)
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/_uti_basic.py", line 62, in fun
  q_out.put((i, f(x)))
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
  run_metadata_ptr)
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
  feed_dict_tensor, options, run_metadata)
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 474, in __init__
  self._fetch_mapper = _FetchMapper.for_fetch(fetches)
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/weight.py", line 301, in train
  self.policy_ratio2: policy_ratio2
File "/home/ubuntu/simu_funs.py", line 58, in once
  inner_parallel = inner_parallel)
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/weight.py", line 301, in train
  self.policy_ratio2: policy_ratio2
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
```

```
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/weight.py", line 301, in train
  self.policy_ratio2: policy_ratio2
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
  feed_dict_tensor, options, run_metadata)
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
  run_metadata)
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
  run_metadata)
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 264, in for_fetch
  return _ListFetchMapper(fetch)
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
  run_metadata_ptr)
File "/home/ubuntu/simu_funs.py", line 201, in simu_once
  w_hidden = w_hidden, Learning_rate = Learning_rate,
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
  run_metadata_ptr)
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
  run_metadata_ptr)
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
  return fn(*args)
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
  return fn(*args)
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 373, in __init__
  self._mappers = [_FetchMapper.for_fetch(fetch) for fetch in fetches]
File "/home/ubuntu/main.py", line 59, in getOneRegionValue
  #                       epsilon = epsilon)
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
  feed_dict_tensor, options, run_metadata)
File "/home/ubuntu/main.py", line 136, in V_DR
  else:
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
  feed_dict_tensor, options, run_metadata)
File "/home/ubuntu/weight.py", line 306, in train
  state_ratio = self.get_density_ratio(S_whole)
File "/home/ubuntu/weight.py", line 301, in train
  self.policy_ratio2: policy_ratio2
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 136, in <listcomp>
  else:
File "/home/ubuntu/main.py", line 112, in getOneRegionValue
  #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
File "/home/ubuntu/main.py", line 240, in getWeight
  batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
File "/home/ubuntu/main.py", line 51, in getOneRegionValue
  CV_QV = CV_QV, penalty_range = penalty, spatial = True)
File "/home/ubuntu/weight.py", line 301, in train
  self.policy_ratio2: policy_ratio2
File "/home/ubuntu/weight.py", line 301, in train
  self.policy_ratio2: policy_ratio2
File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
  feed_dict_tensor, options, run_metadata)
```

```
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 373, in <listcomp>
    self._mappers = [_FetchMapper.for_fetch(fetch) for fetch in fetches]
  File "/home/ubuntu/main.py", line 240, in getWeight
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
  File "/home/ubuntu/weight.py", line 301, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
  File "/home/ubuntu/weight.py", line 301, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/weight.py", line 301, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/weight.py", line 166, in get_density_ratio
    self.state : states
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/main.py", line 59, in getOneRegionValue
    #                       epsilon = epsilon)
  File "/home/ubuntu/main.py", line 240, in getWeight
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
  File "/home/ubuntu/main.py", line 266, in computeQV
    return computeQV_basic(tuples_i = tuples_i, R = R, penalty = penalty,
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 274, in for_fetch
    return _ElementFetchMapper(fetches, contraction_fn)
  File "/home/ubuntu/weight.py", line 301, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/main.py", line 59, in getOneRegionValue
    #                       epsilon = epsilon)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/main.py", line 240, in getWeight
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
  File "/home/ubuntu/weight.py", line 301, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/main.py", line 406, in computeQV_basic
    CKQ_1 = np.hstack((C.dot(KQ), -vec1))
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 302, in __init__
    self._unique_fetches.append(ops.get_default_graph().as_graph_element(
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
KeyboardInterrupt
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/main.py", line 240, in getWeight
```

```
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/weight.py", line 301, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/numpy/linalg/linalg.py", line 403, in solve
    r = gufunc(a, b, signature=signature, extobj=extobj)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/weight.py", line 301, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
```

```
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
Traceback (most recent call last):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
KeyboardInterrupt
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1356, in _do_call
    return fn(*args)
KeyboardInterrupt
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1341, in _run_fn
    options, feed_dict, fetch_list, target_list, run_metadata)
KeyboardInterrupt
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1429, in _call_tf_sessionrun
    run_metadata)
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/main.py", line 136, in V_DR
    else:
KeyboardInterrupt
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
KeyboardInterrupt
  File "/home/ubuntu/main.py", line 59, in getOneRegionValue
    #                       epsilon = epsilon)
  File "/home/ubuntu/main.py", line 240, in getWeight
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
  File "/home/ubuntu/weight.py", line 196, in train
    pi0 = den_b_disc(Ta_tl, n_neigh) * 0.5
  File "/home/ubuntu/_utility.py", line 68, in den_b_disc
    den += binom.pmf(i, N_neigh, 0.5)
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_distn_infrastructure.py", line 3002, in pmf
    goodargs = argsreduce(cond, *((k,)+args))
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
Traceback (most recent call last):
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
Traceback (most recent call last):
  File "/home/ubuntu/main.py", line 136, in V_DR
    else:
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
```

```
  File "/home/ubuntu/main.py", line 59, in getOneRegionValue
    #                       epsilon = epsilon)
  File "/home/ubuntu/main.py", line 240, in getWeight
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
  File "/home/ubuntu/weight.py", line 196, in train
    pi0 = den_b_disc(Ta_tl, n_neigh) * 0.5
  File "/home/ubuntu/_utility.py", line 68, in den_b_disc
    den += binom.pmf(i, N_neigh, 0.5)
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_distn_infrastructure.py", line 2996, in pmf
    cond0 = self._argcheck(*args)
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_discrete_distns.py", line 46, in _argcheck
    return (n >= 0) & (p >= 0) & (p <= 1)
KeyboardInterrupt
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 136, in V_DR
    else:
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
  File "/home/ubuntu/main.py", line 136, in V_DR
    else:
  File "/home/ubuntu/main.py", line 59, in getOneRegionValue
    #                       epsilon = epsilon)
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
  File "/home/ubuntu/main.py", line 240, in getWeight
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
  File "/home/ubuntu/main.py", line 59, in getOneRegionValue
    #                       epsilon = epsilon)
  File "/home/ubuntu/weight.py", line 196, in train
    pi0 = den_b_disc(Ta_tl, n_neigh) * 0.5
  File "/home/ubuntu/main.py", line 240, in getWeight
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
  File "/home/ubuntu/_utility.py", line 68, in den_b_disc
    den += binom.pmf(i, N_neigh, 0.5)
  File "/home/ubuntu/weight.py", line 196, in train
    pi0 = den_b_disc(Ta_tl, n_neigh) * 0.5
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_distn_infrastructure.py", line 2996, in pmf
    cond0 = self._argcheck(*args)
  File "/home/ubuntu/_utility.py", line 68, in den_b_disc
    den += binom.pmf(i, N_neigh, 0.5)
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_discrete_distns.py", line 46, in _argcheck
    return (n >= 0) & (p >= 0) & (p <= 1)
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_distn_infrastructure.py", line 3003, in pmf
    place(output, cond, np.clip(self._pmf(*goodargs), 0, 1))
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_discrete_distns.py", line 58, in _pmf
    return exp(self._logpmf(x, n, p))
KeyboardInterrupt
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_discrete_distns.py", line 54, in _logpmf
    return combiln + special.xlogy(k, p) + special.xlog1py(n-k, -p)
KeyboardInterrupt
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
Traceback (most recent call last):
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/main.py", line 136, in V_DR
    else:
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
  File "/home/ubuntu/main.py", line 112, in getOneRegionValue
    #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
  File "/home/ubuntu/main.py", line 240, in getWeight
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
```

```
  File "/home/ubuntu/weight.py", line 301, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1108, in _run
    feed_dict = nest.flatten_dict_items(feed_dict)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/util/nest.py", line 341, in flatten_dict_items
    if not is_sequence(i):
  File "/home/ubuntu/_uti_basic.py", line 62, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/pywrap_tensorflow_internal.py", line 2459, in IsSequence
    return _pywrap_tensorflow_internal.IsSequence(o)
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 201, in simu_once
    w_hidden = w_hidden, Learning_rate = Learning_rate,
  File "/home/ubuntu/main.py", line 136, in V_DR
    else:
KeyboardInterrupt
  File "/home/ubuntu/main.py", line 136, in <listcomp>
    else:
  File "/home/ubuntu/main.py", line 112, in getOneRegionValue
    #                   epsilon = epsilon, spatial = True, mean_field = False)[0]
  File "/home/ubuntu/main.py", line 240, in getWeight
    batch_size = batch_size, max_iteration = max_iteration, n_neigh = n_neigh,
  File "/home/ubuntu/weight.py", line 301, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1108, in _run
    feed_dict = nest.flatten_dict_items(feed_dict)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/util/nest.py", line 341, in flatten_dict_items
    if not is_sequence(i):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/pywrap_tensorflow_internal.py", line 2459, in IsSequence
    return _pywrap_tensorflow_internal.IsSequence(o)
KeyboardInterrupt
    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 EC2.py
10:46, 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]] = [None, 36, 10, 10,
None, 0.3, 0.5, 1, [True, False, True, False, 10]]


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

----------------------------------------
Value of Behaviour policy:60.732
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, 0.22, -70.88]][[-70.88, -70.88, -10.15]]
std:[[0.0, 0.65, 0.0]][[0.0, 0.0, 0.25]]
MSE:[[70.88, 0.69, 70.88]][[70.88, 70.88, 10.15]]
MSE(-DR):[[0.0, -70.19, 0.0]][[0.0, 0.0, -60.73]]
***
==============


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, -0.26, -69.37]][[-69.37, -69.37, -8.64]]
std:[[0.0, 0.73, 0.0]][[0.0, 0.0, 0.25]]
MSE:[[69.37, 0.77, 69.37]][[69.37, 69.37, 8.64]]
MSE(-DR):[[0.0, -68.6, 0.0]][[0.0, 0.0, -60.73]]
***
==============


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, -2.9, -68.94]][[-68.94, -68.94, -8.21]]
std:[[0.0, 0.68, 0.0]][[0.0, 0.0, 0.25]]
MSE:[[68.94, 2.98, 68.94]][[68.94, 68.94, 8.21]]
MSE(-DR):[[0.0, -65.96, 0.0]][[0.0, 0.0, -60.73]]
***
==============


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, -6.64, -70.48]][[-70.48, -70.48, -9.75]]
std:[[0.0, 0.64, 0.0]][[0.0, 0.0, 0.25]]
MSE:[[70.48, 6.67, 70.48]][[70.48, 70.48, 9.75]]
MSE(-DR):[[0.0, -63.81, 0.0]][[0.0, 0.0, -60.73]]
***
==============
```

```
******************** THIS SETTING IS GOOD ********************
[[70.88  0.69 70.88 70.88 70.88 10.15]
 [69.37  0.77 69.37 69.37 69.37  8.64]
 [68.94  2.98 68.94 68.94 68.94  8.21]
 [70.48  6.67 70.48 70.48 70.48  9.75]]


time spent until now: 0.5 mins


---------------------------------------
[pattern_seed, day, sd_R] = [2, 10, 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
0 -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!
0 -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!
0 -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!
0 -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!

----------------------------------------
Value of Behaviour policy:60.798
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, 0.43, -70.89]][[-70.89, -70.89, -10.09]]
std:[[0.0, 0.65, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.89, 0.78, 70.89]][[70.89, 70.89, 10.09]]
MSE(-DR):[[0.0, -70.11, 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, -0.02, -69.37]][[-69.37, -69.37, -8.58]]
std:[[0.0, 0.59, 0.0]][[0.0, 0.0, 0.23]]
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, -2.86, -68.94]][[-68.94, -68.94, -8.14]]
std:[[0.0, 0.65, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[68.94, 2.93, 68.94]][[68.94, 68.94, 8.14]]
MSE(-DR):[[0.0, -66.01, 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, -6.56, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.72, 0.0]][[0.0, 0.0, 0.23]]
MSE:[[70.47, 6.6, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -63.87, 0.0]][[0.0, 0.0, -60.79]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.88  0.69 70.88 70.88 70.88 10.15]
 [69.37  0.77 69.37 69.37 69.37  8.64]
 [68.94  2.98 68.94 68.94 68.94  8.21]
 [70.48  6.67 70.48 70.48 70.48  9.75]]


[[70.89  0.78 70.89 70.89 70.89 10.09]
 [69.37  0.59 69.37 69.37 69.37  8.58]
 [68.94  2.93 68.94 68.94 68.94  8.14]
 [70.47  6.6  70.47 70.47 70.47  9.68]]


time spent until now: 1.5 mins


----------------------------------------
[pattern_seed, day, sd_R] = [2, 14, 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
0 -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!
0 -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!
0 -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!
0 -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!

----------------------------------------
Value of Behaviour policy:60.775
O_threshold = 80
MC for this TARGET:[70.894, 0.091]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, 0.58, -70.89]][[-70.89, -70.89, -10.12]]
std:[[0.0, 0.54, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.89, 0.79, 70.89]][[70.89, 70.89, 10.12]]
MSE(-DR):[[0.0, -70.1, 0.0]][[0.0, 0.0, -60.77]]
***
==============


O_threshold = 90
MC for this TARGET:[69.377, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.38, -0.0, -69.38]][[-69.38, -69.38, -8.6]]
std:[[0.0, 0.38, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[69.38, 0.38, 69.38]][[69.38, 69.38, 8.6]]
MSE(-DR):[[0.0, -69.0, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 100
MC for this TARGET:[68.925, 0.09]
```

```
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.92, -3.04, -68.92]][[-68.92, -68.92, -8.15]]
std:[[0.0, 0.5, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[68.92, 3.08, 68.92]][[68.92, 68.92, 8.15]]
MSE(-DR):[[0.0, -65.84, 0.0]][[0.0, 0.0, -60.77]]
***
==============


O_threshold = 110
MC for this TARGET:[70.467, 0.083]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -6.61, -70.47]][[-70.47, -70.47, -9.69]]
std:[[0.0, 0.55, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.47, 6.63, 70.47]][[70.47, 70.47, 9.69]]
MSE(-DR):[[0.0, -63.84, 0.0]][[0.0, 0.0, -60.78]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.88  0.69 70.88 70.88 70.88 10.15]
 [69.37  0.77 69.37 69.37 69.37  8.64]
 [68.94  2.98 68.94 68.94 68.94  8.21]
 [70.48  6.67 70.48 70.48 70.48  9.75]]


[[70.89  0.78 70.89 70.89 70.89 10.09]
 [69.37  0.59 69.37 69.37 69.37  8.58]
 [68.94  2.93 68.94 68.94 68.94  8.14]
 [70.47  6.6  70.47 70.47 70.47  9.68]]


[[70.89  0.79 70.89 70.89 70.89 10.12]
 [69.38  0.38 69.38 69.38 69.38  8.6 ]
 [68.92  3.08 68.92 68.92 68.92  8.15]
 [70.47  6.63 70.47 70.47 70.47  9.69]]


time spent until now: 3.5 mins


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

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
0 -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!
0 -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!
0 -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!
0 -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!

----------------------------------------
Value of Behaviour policy:60.709
O_threshold = 80
MC for this TARGET:[70.887, 0.239]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, 0.31, -70.89]][[-70.89, -70.89, -10.18]]
std:[[0.0, 0.99, 0.0]][[0.0, 0.0, 0.32]]
MSE:[[70.89, 1.04, 70.89]][[70.89, 70.89, 10.19]]
MSE(-DR):[[0.0, -69.85, 0.0]][[0.0, 0.0, -60.7]]
***
==============


O_threshold = 90
MC for this TARGET:[69.375, 0.232]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.38, -0.24, -69.38]][[-69.38, -69.38, -8.67]]
std:[[0.0, 1.29, 0.0]][[0.0, 0.0, 0.32]]
MSE:[[69.38, 1.31, 69.38]][[69.38, 69.38, 8.68]]
MSE(-DR):[[0.0, -68.07, 0.0]][[0.0, 0.0, -60.7]]
***
==============


O_threshold = 100
MC for this TARGET:[68.943, 0.229]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.94, -2.94, -68.94]][[-68.94, -68.94, -8.23]]
std:[[0.0, 1.1, 0.0]][[0.0, 0.0, 0.32]]
MSE:[[68.94, 3.14, 68.94]][[68.94, 68.94, 8.24]]
MSE(-DR):[[0.0, -65.8, 0.0]][[0.0, 0.0, -60.7]]
***
==============


O_threshold = 110
MC for this TARGET:[70.487, 0.229]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.49, -6.78, -70.49]][[-70.49, -70.49, -9.78]]
std:[[0.0, 0.99, 0.0]][[0.0, 0.0, 0.32]]
MSE:[[70.49, 6.85, 70.49]][[70.49, 70.49, 9.79]]
MSE(-DR):[[0.0, -63.64, 0.0]][[0.0, 0.0, -60.7]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.88  0.69 70.88 70.88 70.88 10.15]
 [69.37  0.77 69.37 69.37 69.37  8.64]
 [68.94  2.98 68.94 68.94 68.94  8.21]
 [70.48  6.67 70.48 70.48 70.48  9.75]]
```

```
[[70.89  0.78 70.89 70.89 70.89 10.09]
 [69.37  0.59 69.37 69.37 69.37  8.58]
 [68.94  2.93 68.94 68.94 68.94  8.14]
 [70.47  6.6  70.47 70.47 70.47  9.68]]


[[70.89  0.79 70.89 70.89 70.89 10.12]
 [69.38  0.38 69.38 69.38 69.38  8.6 ]
 [68.92  3.08 68.92 68.92 68.92  8.15]
 [70.47  6.63 70.47 70.47 70.47  9.69]]


[[70.89  1.04 70.89 70.89 70.89 10.19]
 [69.38  1.31 69.38 69.38 69.38  8.68]
 [68.94  3.14 68.94 68.94 68.94  8.24]
 [70.49  6.85 70.49 70.49 70.49  9.79]]


time spent until now: 4.1 mins


---------------------------------------
[pattern_seed, day, sd_R] = [2, 10, 20]

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
0 -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!
0 -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!
0 -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!
0 -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!

---------------------------------------
Value of Behaviour policy:60.786
O_threshold = 80
MC for this TARGET:[70.881, 0.169]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.88, 0.33, -70.88]][[-70.88, -70.88, -10.09]]
std:[[0.0, 0.99, 0.0]][[0.0, 0.0, 0.27]]
MSE:[[70.88, 1.04, 70.88]][[70.88, 70.88, 10.09]]
MSE(-DR):[[0.0, -69.84, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 90
MC for this TARGET:[69.368, 0.17]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.37, -0.05, -69.37]][[-69.37, -69.37, -8.58]]
std:[[0.0, 0.9, 0.0]][[0.0, 0.0, 0.27]]
MSE:[[69.37, 0.9, 69.37]][[69.37, 69.37, 8.58]]
MSE(-DR):[[0.0, -68.47, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.931, 0.172]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.93, -2.76, -68.93]][[-68.93, -68.93, -8.14]]
std:[[0.0, 0.92, 0.0]][[0.0, 0.0, 0.27]]
MSE:[[68.93, 2.91, 68.93]][[68.93, 68.93, 8.14]]
MSE(-DR):[[0.0, -66.02, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 110
MC for this TARGET:[70.469, 0.175]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -6.47, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 1.09, 0.0]][[0.0, 0.0, 0.27]]
MSE:[[70.47, 6.56, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -63.91, 0.0]][[0.0, 0.0, -60.79]]
***
==============


********** THIS SETTING IS GOOD **********
[[70.88  0.69 70.88 70.88 70.88 10.15]
 [69.37  0.77 69.37 69.37 69.37  8.64]
 [68.94  2.98 68.94 68.94 68.94  8.21]
 [70.48  6.67 70.48 70.48 70.48  9.75]]


[[70.89  0.78 70.89 70.89 70.89 10.09]
 [69.37  0.59 69.37 69.37 69.37  8.58]
 [68.94  2.93 68.94 68.94 68.94  8.14]
 [70.47  6.6  70.47 70.47 70.47  9.68]]


[[70.89  0.79 70.89 70.89 70.89 10.12]
 [69.38  0.38 69.38 69.38 69.38  8.6 ]
 [68.92  3.08 68.92 68.92 68.92  8.15]
 [70.47  6.63 70.47 70.47 70.47  9.69]]


[[70.89  1.04 70.89 70.89 70.89 10.19]
 [69.38  1.31 69.38 69.38 69.38  8.68]
 [68.94  3.14 68.94 68.94 68.94  8.24]
 [70.49  6.85 70.49 70.49 70.49  9.79]]


[[70.88  1.04 70.88 70.88 70.88 10.09]
```

```
 [69.37  0.9  69.37 69.37 69.37  8.58]
 [68.93  2.91 68.93 68.93 68.93  8.14]
 [70.47  6.56 70.47 70.47 70.47  9.68]]


time spent until now: 5.1 mins


----------------------------------------
[pattern_seed, day, sd_R] = [2, 14, 20]

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
0 -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!
0 -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!
0 -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!
0 -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!
^CProcess Process-185:
Traceback (most recent call last):
  File "EC2.py", line 74, in <module>
    print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 60, in simu
Process Process-181:
    value_reps = parmap(once, range(OPE_rep_times), n_cores)
  File "/home/ubuntu/_uti_basic.py", line 75, in parmap
    [q_in.put((None, None)) for _ in range(nprocs)]
```

```
  File "/home/ubuntu/_uti_basic.py", line 75, in <listcomp>
Process Process-195:
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
Process Process-186:
Process Process-210:
ubuntu@ip-172-31-77-47:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC22.py
10:52, 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, pois0, simple, u_O_u_D]] = [None, 96, 10, 10,
None, 0.3, 0.5, 1, [True, False, True, False, 10]]


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

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
0 -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!
0 -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!
0 -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!
0 -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!

---------------------------------------
Value of Behaviour policy:60.789
O_threshold = 80
MC for this TARGET:[70.894, 0.091]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, -4.58, -70.89]][[-70.89, -70.89, -10.11]]
std:[[0.0, 0.31, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.89, 4.59, 70.89]][[70.89, 70.89, 10.11]]
MSE(-DR):[[0.0, -66.3, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.377, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.38, -3.78, -69.38]][[-69.38, -69.38, -8.59]]
std:[[0.0, 0.28, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[69.38, 3.79, 69.38]][[69.38, 69.38, 8.59]]
MSE(-DR):[[0.0, -65.59, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.925, 0.09]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.92, -4.9, -68.92]][[-68.92, -68.92, -8.14]]
std:[[0.0, 0.26, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[68.92, 4.91, 68.92]][[68.92, 68.92, 8.14]]
MSE(-DR):[[0.0, -64.01, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 110
MC for this TARGET:[70.467, 0.083]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -8.26, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.34, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.47, 8.27, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -62.2, 0.0]][[0.0, 0.0, -60.79]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.89  4.59 70.89 70.89 70.89 10.11]
 [69.38  3.79 69.38 69.38 69.38  8.59]
 [68.92  4.91 68.92 68.92 68.92  8.14]
 [70.47  8.27 70.47 70.47 70.47  9.68]]


time spent until now: 4.8 mins


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

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
0 -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!
0 -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!
0 -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!
0 -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!

---------------------------------------
Value of Behaviour policy:60.789
O_threshold = 80
MC for this TARGET:[70.894, 0.091]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, -0.99, -70.89]][[-70.89, -70.89, -10.11]]
std:[[0.0, 0.4, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.89, 1.07, 70.89]][[70.89, 70.89, 10.11]]
MSE(-DR):[[0.0, -69.82, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.377, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.38, -0.97, -69.38]][[-69.38, -69.38, -8.59]]
std:[[0.0, 0.33, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[69.38, 1.02, 69.38]][[69.38, 69.38, 8.59]]
MSE(-DR):[[0.0, -68.36, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.925, 0.09]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.92, -3.38, -68.92]][[-68.92, -68.92, -8.14]]
```

```
std:[[0.0, 0.35, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[68.92, 3.4, 68.92]][[68.92, 68.92, 8.14]]
MSE(-DR):[[0.0, -65.52, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 110
MC for this TARGET:[70.467, 0.083]
   [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, 0.42, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.47, 7.17, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -63.3, 0.0]][[0.0, 0.0, -60.79]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.89  4.59 70.89 70.89 70.89 10.11]
 [69.38  3.79 69.38 69.38 69.38  8.59]
 [68.92  4.91 68.92 68.92 68.92  8.14]
 [70.47  8.27 70.47 70.47 70.47  9.68]]


[[70.89  1.07 70.89 70.89 70.89 10.11]
 [69.38  1.02 69.38 69.38 69.38  8.59]
 [68.92  3.4  68.92 68.92 68.92  8.14]
 [70.47  7.17 70.47 70.47 70.47  9.68]]


time spent until now: 9.6 mins


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

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
0 -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!
0 -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!
0 -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!
0 -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!

---------------------------------------
Value of Behaviour policy:60.789
O_threshold = 80
MC for this TARGET:[70.894, 0.091]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, 0.55, -70.89]][[-70.89, -70.89, -10.11]]
std:[[0.0, 0.53, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.89, 0.76, 70.89]][[70.89, 70.89, 10.11]]
MSE(-DR):[[0.0, -70.13, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.377, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.38, -0.1, -69.38]][[-69.38, -69.38, -8.59]]
std:[[0.0, 0.41, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[69.38, 0.42, 69.38]][[69.38, 69.38, 8.59]]
MSE(-DR):[[0.0, -68.96, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.925, 0.09]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.92, -3.02, -68.92]][[-68.92, -68.92, -8.14]]
std:[[0.0, 0.46, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[68.92, 3.05, 68.92]][[68.92, 68.92, 8.14]]
MSE(-DR):[[0.0, -65.87, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 110
MC for this TARGET:[70.467, 0.083]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -6.73, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.55, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.47, 6.75, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -63.72, 0.0]][[0.0, 0.0, -60.79]]
***
==============


********** THIS SETTING IS GOOD **********
[[70.89  4.59 70.89 70.89 70.89 10.11]
 [69.38  3.79 69.38 69.38 69.38  8.59]
 [68.92  4.91 68.92 68.92 68.92  8.14]
 [70.47  8.27 70.47 70.47 70.47  9.68]]


[[70.89  1.07 70.89 70.89 70.89 10.11]
 [69.38  1.02 69.38 69.38 69.38  8.59]
 [68.92  3.4  68.92 68.92 68.92  8.14]
 [70.47  7.17 70.47 70.47 70.47  9.68]]


[[70.89  0.76 70.89 70.89 70.89 10.11]
```

```
 [69.38  0.42 69.38 69.38 69.38  8.59]
 [68.92  3.05 68.92 68.92 68.92  8.14]
 [70.47  6.75 70.47 70.47 70.47  9.68]]


time spent until now: 14.4 mins


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

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
0 -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!
0 -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!
0 -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!
0 -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!

---------------------------------------
Value of Behaviour policy:60.789
O_threshold = 80
MC for this TARGET:[70.894, 0.091]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.89, 0.68, -70.89]][[-70.89, -70.89, -10.11]]
std:[[0.0, 0.59, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.89, 0.9, 70.89]][[70.89, 70.89, 10.11]]
MSE(-DR):[[0.0, -69.99, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 90
MC for this TARGET:[69.377, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-69.38, 0.12, -69.38]][[-69.38, -69.38, -8.59]]
std:[[0.0, 0.49, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[69.38, 0.5, 69.38]][[69.38, 69.38, 8.59]]
MSE(-DR):[[0.0, -68.88, 0.0]][[0.0, 0.0, -60.79]]
***
==============


O_threshold = 100
MC for this TARGET:[68.925, 0.09]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-68.92, -2.6, -68.92]][[-68.92, -68.92, -8.14]]
std:[[0.0, 0.55, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[68.92, 2.66, 68.92]][[68.92, 68.92, 8.14]]
MSE(-DR):[[0.0, -66.26, 0.0]][[0.0, 0.0, -60.78]]
***
==============


O_threshold = 110
MC for this TARGET:[70.467, 0.083]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-70.47, -6.17, -70.47]][[-70.47, -70.47, -9.68]]
std:[[0.0, 0.66, 0.0]][[0.0, 0.0, 0.14]]
MSE:[[70.47, 6.21, 70.47]][[70.47, 70.47, 9.68]]
MSE(-DR):[[0.0, -64.26, 0.0]][[0.0, 0.0, -60.79]]
***
==============


********** THIS SETTING IS GOOD **********
[[70.89  4.59 70.89 70.89 70.89 10.11]
 [69.38  3.79 69.38 69.38 69.38  8.59]
 [68.92  4.91 68.92 68.92 68.92  8.14]
 [70.47  8.27 70.47 70.47 70.47  9.68]]


[[70.89  1.07 70.89 70.89 70.89 10.11]
 [69.38  1.02 69.38 69.38 69.38  8.59]
 [68.92  3.4  68.92 68.92 68.92  8.14]
 [70.47  7.17 70.47 70.47 70.47  9.68]]


[[70.89  0.76 70.89 70.89 70.89 10.11]
 [69.38  0.42 69.38 69.38 69.38  8.59]
 [68.92  3.05 68.92 68.92 68.92  8.14]
 [70.47  6.75 70.47 70.47 70.47  9.68]]


[[70.89  0.9  70.89 70.89 70.89 10.11]
 [69.38  0.5  69.38 69.38 69.38  8.59]
 [68.92  2.66 68.92 68.92 68.92  8.14]
 [70.47  6.21 70.47 70.47 70.47  9.68]]


time spent until now: 19.2 mins


---------------------------------------
[pattern_seed, lam] = [4, 0.01]

max(u_O) =  193.8
O_threshold = 80
means of Order:

101.0 115.6 73.8 122.5 87.8

61.8 81.9 119.1 109.9 70.5

119.8 96.9 113.0 109.9 70.3
```

```
110.5 82.9 158.2 123.6 100.9

74.1 101.1 104.4 69.2 193.8

target policy:

1 1 0 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

0 1 1 0 1

number of reward locations:  19
O_threshold = 90
target policy:

1 1 0 1 0

0 0 1 1 0

1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 100
target policy:

1 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 110
target policy:

0 1 0 1 0

0 0 1 0 0

1 0 1 0 0

1 0 1 1 0

0 0 0 0 1

number of reward locations:  9
0 -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!
0 -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!
0 -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!
0 -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!

----------------------------------------
Value of Behaviour policy:65.2
O_threshold = 80
MC for this TARGET:[72.837, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-72.84, -3.55, -72.84]][[-72.84, -72.84, -7.64]]
std:[[0.0, 0.32, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[72.84, 3.56, 72.84]][[72.84, 72.84, 7.64]]
MSE(-DR):[[0.0, -69.28, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 90
MC for this TARGET:[74.173, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-74.17, -3.87, -74.17]][[-74.17, -74.17, -8.97]]
```

```
std:[[0.0, 0.27, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[74.17, 3.88, 74.17]][[74.17, 74.17, 8.97]]
MSE(-DR):[[0.0, -70.29, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 100
MC for this TARGET:[74.655, 0.089]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-74.66, -3.11, -74.66]][[-74.66, -74.66, -9.46]]
std:[[0.0, 0.24, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[74.66, 3.12, 74.66]][[74.66, 74.66, 9.46]]
MSE(-DR):[[0.0, -71.54, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 110
MC for this TARGET:[73.624, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-73.62, -6.81, -73.62]][[-73.62, -73.62, -8.42]]
std:[[0.0, 0.28, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[73.62, 6.82, 73.62]][[73.62, 73.62, 8.42]]
MSE(-DR):[[0.0, -66.8, 0.0]][[0.0, 0.0, -65.2]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.89  4.59 70.89 70.89 70.89 10.11]
 [69.38  3.79 69.38 69.38 69.38  8.59]
 [68.92  4.91 68.92 68.92 68.92  8.14]
 [70.47  8.27 70.47 70.47 70.47  9.68]]


[[70.89  1.07 70.89 70.89 70.89 10.11]
 [69.38  1.02 69.38 69.38 69.38  8.59]
 [68.92  3.4  68.92 68.92 68.92  8.14]
 [70.47  7.17 70.47 70.47 70.47  9.68]]


[[70.89  0.76 70.89 70.89 70.89 10.11]
 [69.38  0.42 69.38 69.38 69.38  8.59]
 [68.92  3.05 68.92 68.92 68.92  8.14]
 [70.47  6.75 70.47 70.47 70.47  9.68]]


[[70.89  0.9  70.89 70.89 70.89 10.11]
 [69.38  0.5  69.38 69.38 69.38  8.59]
 [68.92  2.66 68.92 68.92 68.92  8.14]
 [70.47  6.21 70.47 70.47 70.47  9.68]]


[[72.84  3.56 72.84 72.84 72.84  7.64]
 [74.17  3.88 74.17 74.17 74.17  8.97]
 [74.66  3.12 74.66 74.66 74.66  9.46]
 [73.62  6.82 73.62 73.62 73.62  8.42]]


time spent until now: 24.0 mins


---------------------------------------
[pattern_seed, lam] = [4, 0.001]

max(u_O) =  193.8
O_threshold = 80
means of Order:

101.0 115.6 73.8 122.5 87.8

61.8 81.9 119.1 109.9 70.5

119.8 96.9 113.0 109.9 70.3

110.5 82.9 158.2 123.6 100.9

74.1 101.1 104.4 69.2 193.8

target policy:

1 1 0 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1
```

```
0 1 1 0 1

number of reward locations:  19
O_threshold = 90
target policy:

1 1 0 1 0

0 0 1 1 0

1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 100
target policy:

1 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 110
target policy:

0 1 0 1 0

0 0 1 0 0

1 0 1 0 0

1 0 1 1 0

0 0 0 0 1

number of reward locations:  9
0 -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!
0 -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!
0 -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!
0 -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!

--------------------------------------
Value of Behaviour policy:65.2
O_threshold = 80
MC for this TARGET:[72.837, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-72.84, 0.33, -72.84]][[-72.84, -72.84, -7.64]]
std:[[0.0, 0.46, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[72.84, 0.57, 72.84]][[72.84, 72.84, 7.64]]
MSE(-DR):[[0.0, -72.27, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 90
MC for this TARGET:[74.173, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-74.17, -0.24, -74.17]][[-74.17, -74.17, -8.97]]
std:[[0.0, 0.41, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[74.17, 0.48, 74.17]][[74.17, 74.17, 8.97]]
MSE(-DR):[[0.0, -73.69, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 100
MC for this TARGET:[74.655, 0.089]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-74.66, 0.29, -74.66]][[-74.66, -74.66, -9.46]]
std:[[0.0, 0.34, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[74.66, 0.45, 74.66]][[74.66, 74.66, 9.46]]
MSE(-DR):[[0.0, -74.21, 0.0]][[0.0, 0.0, -65.2]]
```

***
==========

O_threshold = 110
MC for this TARGET:[73.624, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-73.62, -4.54, -73.62]][[-73.62, -73.62, -8.42]]
std:[[0.0, 0.35, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[73.62, 4.55, 73.62]][[73.62, 73.62, 8.42]]
MSE(-DR):[[0.0, -69.07, 0.0]][[0.0, 0.0, -65.2]]
***
==========

******************** THIS SETTING IS GOOD ********************
[[70.89 4.59 70.89 70.89 70.89 10.11]
 [69.38 3.79 69.38 69.38 69.38 8.59]
 [68.92 4.91 68.92 68.92 68.92 8.14]
 [70.47 8.27 70.47 70.47 70.47 9.68]]

[[70.89 1.07 70.89 70.89 70.89 10.11]
 [69.38 1.02 69.38 69.38 69.38 8.59]
 [68.92 3.4 68.92 68.92 68.92 8.14]
 [70.47 7.17 70.47 70.47 70.47 9.68]]

[[70.89 0.76 70.89 70.89 70.89 10.11]
 [69.38 0.42 69.38 69.38 69.38 8.59]
 [68.92 3.05 68.92 68.92 68.92 8.14]
 [70.47 6.75 70.47 70.47 70.47 9.68]]

[[70.89 0.9 70.89 70.89 70.89 10.11]
 [69.38 0.5 69.38 69.38 69.38 8.59]
 [68.92 2.66 68.92 68.92 68.92 8.14]
 [70.47 6.21 70.47 70.47 70.47 9.68]]

[[72.84 3.56 72.84 72.84 72.84 7.64]
 [74.17 3.88 74.17 74.17 74.17 8.97]
 [74.66 3.12 74.66 74.66 74.66 9.46]
 [73.62 6.82 73.62 73.62 73.62 8.42]]

[[72.84 0.57 72.84 72.84 72.84 7.64]
 [74.17 0.48 74.17 74.17 74.17 8.97]
 [74.66 0.45 74.66 74.66 74.66 9.46]
 [73.62 4.55 73.62 73.62 73.62 8.42]]

time spent until now: 28.7 mins

[pattern_seed, lam] = [4, 0.0001]

max(u_O) = 193.8
O_threshold = 80
means of Order:

101.0 115.6 73.8 122.5 87.8

61.8 81.9 119.1 109.9 70.5

119.8 96.9 113.0 109.9 70.3

110.5 82.9 158.2 123.6 100.9

74.1 101.1 104.4 69.2 193.8

target policy:

1 1 0 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

0 1 1 0 1

number of reward locations: 19
O_threshold = 90
target policy:

1 1 0 1 0

```
0 0 1 1 0

1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 100
target policy:

1 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 110
target policy:

0 1 0 1 0

0 0 1 0 0

1 0 1 0 0

1 0 1 1 0

0 0 0 0 1

number of reward locations:  9
0 -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!
0 -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!
0 -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!
0 -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!

--------------------------------------
Value of Behaviour policy:65.2
O_threshold = 80
MC for this TARGET:[72.837, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-72.84, 2.65, -72.84]][[-72.84, -72.84, -7.64]]
std:[[0.0, 0.64, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[72.84, 2.73, 72.84]][[72.84, 72.84, 7.64]]
MSE(-DR):[[0.0, -70.11, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 90
MC for this TARGET:[74.173, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-74.17, 1.51, -74.17]][[-74.17, -74.17, -8.97]]
std:[[0.0, 0.53, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[74.17, 1.6, 74.17]][[74.17, 74.17, 8.97]]
MSE(-DR):[[0.0, -72.57, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 100
MC for this TARGET:[74.655, 0.089]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-74.66, 1.52, -74.66]][[-74.66, -74.66, -9.46]]
std:[[0.0, 0.46, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[74.66, 1.59, 74.66]][[74.66, 74.66, 9.46]]
MSE(-DR):[[0.0, -73.07, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 110
MC for this TARGET:[73.624, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-73.62, -3.71, -73.62]][[-73.62, -73.62, -8.42]]
```

std:[[0.0, 0.4, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[73.62, 3.73, 73.62]][[73.62, 73.62, 8.42]]
MSE(-DR):[[0.0, -69.89, 0.0]][[0.0, 0.0, -65.2]]
***
==========

********** THIS SETTING IS GOOD **********
[[70.89 4.59 70.89 70.89 70.89 10.11]
 [69.38 3.79 69.38 69.38 69.38 8.59]
 [68.92 4.91 68.92 68.92 68.92 8.14]
 [70.47 8.27 70.47 70.47 70.47 9.68]]

[[70.89 1.07 70.89 70.89 70.89 10.11]
 [69.38 1.02 69.38 69.38 69.38 8.59]
 [68.92 3.4 68.92 68.92 68.92 8.14]
 [70.47 7.17 70.47 70.47 70.47 9.68]]

[[70.89 0.76 70.89 70.89 70.89 10.11]
 [69.38 0.42 69.38 69.38 69.38 8.59]
 [68.92 3.05 68.92 68.92 68.92 8.14]
 [70.47 6.75 70.47 70.47 70.47 9.68]]

[[70.89 0.9 70.89 70.89 70.89 10.11]
 [69.38 0.5 69.38 69.38 69.38 8.59]
 [68.92 2.66 68.92 68.92 68.92 8.14]
 [70.47 6.21 70.47 70.47 70.47 9.68]]

[[72.84 3.56 72.84 72.84 72.84 7.64]
 [74.17 3.88 74.17 74.17 74.17 8.97]
 [74.66 3.12 74.66 74.66 74.66 9.46]
 [73.62 6.82 73.62 73.62 73.62 8.42]]

[[72.84 0.57 72.84 72.84 72.84 7.64]
 [74.17 0.48 74.17 74.17 74.17 8.97]
 [74.66 0.45 74.66 74.66 74.66 9.46]
 [73.62 4.55 73.62 73.62 73.62 8.42]]

[[72.84 2.73 72.84 72.84 72.84 7.64]
 [74.17 1.6 74.17 74.17 74.17 8.97]
 [74.66 1.59 74.66 74.66 74.66 9.46]
 [73.62 3.73 73.62 73.62 73.62 8.42]]

time spent until now: 33.5 mins

[pattern_seed, lam] = [4, 1e-05]

max(u_O) = 193.8
O_threshold = 80
means of Order:

101.0 115.6 73.8 122.5 87.8

61.8 81.9 119.1 109.9 70.5

119.8 96.9 113.0 109.9 70.3

110.5 82.9 158.2 123.6 100.9

74.1 101.1 104.4 69.2 193.8

target policy:

1 1 0 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

0 1 1 0 1

number of reward locations: 19
O_threshold = 90
target policy:

1 1 0 1 0

0 0 1 1 0

```
1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 100
target policy:

1 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 110
target policy:

0 1 0 1 0

0 0 1 0 0

1 0 1 0 0

1 0 1 1 0

0 0 0 0 1

number of reward locations:  9
0 -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!
0 -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!
0 -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!
0 -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!

----------------------------------------
Value of Behaviour policy:65.2
O_threshold = 80
MC for this TARGET:[72.837, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-72.84, 2.63, -72.84]][[-72.84, -72.84, -7.64]]
std:[[0.0, 0.73, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[72.84, 2.73, 72.84]][[72.84, 72.84, 7.64]]
MSE(-DR):[[0.0, -70.11, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 90
MC for this TARGET:[74.173, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-74.17, 1.62, -74.17]][[-74.17, -74.17, -8.97]]
std:[[0.0, 0.55, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[74.17, 1.71, 74.17]][[74.17, 74.17, 8.97]]
MSE(-DR):[[0.0, -72.46, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 100
MC for this TARGET:[74.655, 0.089]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-74.66, 1.63, -74.66]][[-74.66, -74.66, -9.46]]
std:[[0.0, 0.5, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[74.66, 1.7, 74.66]][[74.66, 74.66, 9.46]]
MSE(-DR):[[0.0, -72.96, 0.0]][[0.0, 0.0, -65.2]]
***
==============


O_threshold = 110
MC for this TARGET:[73.624, 0.086]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-73.62, -3.34, -73.62]][[-73.62, -73.62, -8.42]]
std:[[0.0, 0.44, 0.0]][[0.0, 0.0, 0.13]]
MSE:[[73.62, 3.37, 73.62]][[73.62, 73.62, 8.42]]
```

```
MSE(-DR):[[0.0, -70.25, 0.0]][[0.0, 0.0, -65.2]]
***
==============


******************** THIS SETTING IS GOOD ********************
[[70.89  4.59 70.89 70.89 70.89 10.11]
 [69.38  3.79 69.38 69.38 69.38  8.59]
 [68.92  4.91 68.92 68.92 68.92  8.14]
 [70.47  8.27 70.47 70.47 70.47  9.68]]


[[70.89  1.07 70.89 70.89 70.89 10.11]
 [69.38  1.02 69.38 69.38 69.38  8.59]
 [68.92  3.4  68.92 68.92 68.92  8.14]
 [70.47  7.17 70.47 70.47 70.47  9.68]]


[[70.89  0.76 70.89 70.89 70.89 10.11]
 [69.38  0.42 69.38 69.38 69.38  8.59]
 [68.92  3.05 68.92 68.92 68.92  8.14]
 [70.47  6.75 70.47 70.47 70.47  9.68]]


[[70.89  0.9  70.89 70.89 70.89 10.11]
 [69.38  0.5  69.38 69.38 69.38  8.59]
 [68.92  2.66 68.92 68.92 68.92  8.14]
 [70.47  6.21 70.47 70.47 70.47  9.68]]


[[72.84  3.56 72.84 72.84 72.84  7.64]
 [74.17  3.88 74.17 74.17 74.17  8.97]
 [74.66  3.12 74.66 74.66 74.66  9.46]
 [73.62  6.82 73.62 73.62 73.62  8.42]]


[[72.84  0.57 72.84 72.84 72.84  7.64]
 [74.17  0.48 74.17 74.17 74.17  8.97]
 [74.66  0.45 74.66 74.66 74.66  9.46]
 [73.62  4.55 73.62 73.62 73.62  8.42]]


[[72.84  2.73 72.84 72.84 72.84  7.64]
 [74.17  1.6  74.17 74.17 74.17  8.97]
 [74.66  1.59 74.66 74.66 74.66  9.46]
 [73.62  3.73 73.62 73.62 73.62  8.42]]


[[72.84  2.73 72.84 72.84 72.84  7.64]
 [74.17  1.71 74.17 74.17 74.17  8.97]
 [74.66  1.7  74.66 74.66 74.66  9.46]
 [73.62  3.37 73.62 73.62 73.62  8.42]]


time spent until now: 38.3 mins

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