```
Last login: Wed Apr  1 00:22:25 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-3-221-160-139.compute-1.amazonaws.com
The authenticity of host 'ec2-3-221-160-139.compute-1.amazonaws.com (3.221.160.139)' can't be established.
ECDSA key fingerprint is SHA256:9dJlbEcNfvRLPyZQlCu3st4dx1S5gCF6eay5PnTAogI.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-221-160-139.compute-1.amazonaws.com,3.221.160.139' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1060-aws x86_64)

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

  System information as of Wed Apr  1 15:16:29 UTC 2020


Last login: Thu Mar  5 21:23:34 2020 from 107.13.161.147
ubuntu@ip-172-31-5-213:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
Traceback (most recent call last):
  File "EC2.py", line 5, in <module>
    from simu_funs import *
  File "/home/ubuntu/simu_funs.py", line 105
    printB("MSE(-DR):" + str([mse_rel[:3]]) + str([mse_rel[3:6]])
         ^
SyntaxError: invalid syntax
ubuntu@ip-172-31-5-213:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
11:18, 04/01; num of cores:16

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A,  simple, M_in_R, u_O_u_D, mean_reversion, pois0] = [672, 5, 5, None, 0.2, 1, 1, Fa
lse, True, 5, False, False]


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

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1
```

```
number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:77.545
O_threshold = 80
MC for this TARGET:[87.1, 0.04]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.44, -1.65, 0.11]][[1.71, -87.1, -9.55]]
std:[[0.13, 0.11, 0.15]][[0.13, 0.0, 0.04]]
MSE:[[1.45, 1.65, 0.19]][[1.71, 87.1, 9.55]]
MSE(-DR):[[0.0, 0.2, -1.26]][[0.26, 85.65, 8.1]]
**
==============


O_threshold = 90
MC for this TARGET:[85.289, 0.039]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.59, -0.8, -0.07]][[2.7, -85.29, -7.74]]
std:[[0.02, 0.02, 0.2]][[0.16, 0.0, 0.04]]
MSE:[[0.59, 0.8, 0.21]][[2.7, 85.29, 7.74]]
MSE(-DR):[[0.0, 0.21, -0.38]][[2.11, 84.7, 7.15]]
**
==============


O_threshold = 100
MC for this TARGET:[89.618, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.44, -2.72, -4.0]][[-0.45, -89.62, -12.07]]
std:[[0.12, 0.14, 0.15]][[0.24, 0.0, 0.04]]
MSE:[[2.44, 2.72, 4.0]][[0.51, 89.62, 12.07]]
MSE(-DR):[[0.0, 0.28, 1.56]][[-1.93, 87.18, 9.63]]
==============


O_threshold = 110
MC for this TARGET:[86.85, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.3, -3.53, -3.72]][[-3.05, -86.85, -9.3]]
std:[[0.0, 0.02, 0.01]][[0.17, 0.0, 0.04]]
MSE:[[3.3, 3.53, 3.72]][[3.05, 86.85, 9.3]]
MSE(-DR):[[0.0, 0.23, 0.42]][[-0.25, 83.55, 6.0]]
==============


O_threshold = 120
MC for this TARGET:[88.239, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.29, -8.47, -7.79]][[-7.6, -88.24, -10.69]]
std:[[0.13, 0.18, 0.17]][[0.24, 0.0, 0.04]]
MSE:[[8.29, 8.47, 7.79]][[7.6, 88.24, 10.69]]
MSE(-DR):[[0.0, 0.18, -0.5]][[-0.69, 79.95, 2.4]]
==============


[[ 1.45  1.65  0.19  1.71 87.1   9.55]
 [ 0.59  0.8   0.21  2.7  85.29  7.74]
 [ 2.44  2.72  4.    0.51 89.62 12.07]
```

```
 [ 3.3   3.53  3.72  3.05 86.85  9.3 ]
 [ 8.29  8.47  7.79  7.6  88.24 10.69]]
time spent until now: 6.2 mins


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

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1
```

```
number of reward locations:  8
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

---------------------------------------
Value of Behaviour policy:77.545
O_threshold = 80
MC for this TARGET:[87.1, 0.04]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.42, -1.65, 0.11]][[1.71, -87.1, -9.55]]
std:[[0.13, 0.11, 0.2]][[0.13, 0.0, 0.04]]
MSE:[[1.43, 1.65, 0.23]][[1.71, 87.1, 9.55]]
MSE(-DR):[[0.0, 0.22, -1.2]][[0.28, 85.67, 8.12]]
**
==============


O_threshold = 90
MC for this TARGET:[85.289, 0.039]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.61, -0.8, -0.08]][[2.7, -85.29, -7.74]]
std:[[0.02, 0.02, 0.16]][[0.19, 0.0, 0.04]]
MSE:[[0.61, 0.8, 0.18]][[2.71, 85.29, 7.74]]
MSE(-DR):[[0.0, 0.19, -0.43]][[2.1, 84.68, 7.13]]
**
==============


O_threshold = 100
MC for this TARGET:[89.618, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.45, -2.72, -4.04]][[-0.45, -89.62, -12.07]]
std:[[0.12, 0.14, 0.18]][[0.24, 0.0, 0.04]]
MSE:[[2.45, 2.72, 4.04]][[0.51, 89.62, 12.07]]
MSE(-DR):[[0.0, 0.27, 1.59]][[-1.94, 87.17, 9.62]]
==============


O_threshold = 110
MC for this TARGET:[86.85, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.33, -3.53, -3.76]][[-3.05, -86.85, -9.3]]
std:[[0.04, 0.02, 0.03]][[0.16, 0.0, 0.04]]
MSE:[[3.33, 3.53, 3.76]][[3.05, 86.85, 9.3]]
MSE(-DR):[[0.0, 0.2, 0.43]][[-0.28, 83.52, 5.97]]
==============


O_threshold = 120
MC for this TARGET:[88.239, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.33, -8.47, -7.81]][[-7.61, -88.24, -10.69]]
std:[[0.11, 0.18, 0.16]][[0.24, 0.0, 0.04]]
MSE:[[8.33, 8.47, 7.81]][[7.61, 88.24, 10.69]]
MSE(-DR):[[0.0, 0.14, -0.52]][[-0.72, 79.91, 2.36]]
==============


[[ 1.45  1.65  0.19  1.71 87.1   9.55]
 [ 0.59  0.8   0.21  2.7  85.29  7.74]
 [ 2.44  2.72  4.    0.51 89.62 12.07]
 [ 3.3   3.53  3.72  3.05 86.85  9.3 ]
 [ 8.29  8.47  7.79  7.6  88.24 10.69]]
[[ 1.43  1.65  0.23  1.71 87.1   9.55]
 [ 0.61  0.8   0.18  2.71 85.29  7.74]
 [ 2.45  2.72  4.04  0.51 89.62 12.07]
 [ 3.33  3.53  3.76  3.05 86.85  9.3 ]
 [ 8.33  8.47  7.81  7.61 88.24 10.69]]
time spent until now: 12.1 mins


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

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6
```

```
target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:77.545
O_threshold = 80
MC for this TARGET:[87.1, 0.04]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.42, -1.65, 0.09]][[1.71, -87.1, -9.55]]
std:[[0.12, 0.11, 0.2]][[0.12, 0.0, 0.04]]
MSE:[[1.43, 1.65, 0.22]][[1.71, 87.1, 9.55]]
MSE(-DR):[[0.0, 0.22, -1.21]][[0.28, 85.67, 8.12]]
**
==============


O_threshold = 90
MC for this TARGET:[85.289, 0.039]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.61, -0.8, -0.05]][[2.71, -85.29, -7.74]]
```

```
std:[[0.01, 0.02, 0.18]][[0.2, 0.0, 0.04]]
MSE:[[0.61, 0.8, 0.19]][[2.72, 85.29, 7.74]]
MSE(-DR):[[0.0, 0.19, -0.42]][[2.11, 84.68, 7.13]]
**
==============


O_threshold = 100
MC for this TARGET:[89.618, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.45, -2.72, -4.0]][[-0.44, -89.62, -12.07]]
std:[[0.11, 0.14, 0.18]][[0.24, 0.0, 0.04]]
MSE:[[2.45, 2.72, 4.0]][[0.5, 89.62, 12.07]]
MSE(-DR):[[0.0, 0.27, 1.55]][[-1.95, 87.17, 9.62]]
==============


O_threshold = 110
MC for this TARGET:[86.85, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.34, -3.53, -3.73]][[-3.05, -86.85, -9.3]]
std:[[0.02, 0.02, 0.04]][[0.17, 0.0, 0.04]]
MSE:[[3.34, 3.53, 3.73]][[3.05, 86.85, 9.3]]
MSE(-DR):[[0.0, 0.19, 0.39]][[-0.29, 83.51, 5.96]]
==============


O_threshold = 120
MC for this TARGET:[88.239, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-8.29, -8.47, -7.79]][[-7.61, -88.24, -10.69]]
std:[[0.11, 0.18, 0.14]][[0.23, 0.0, 0.04]]
MSE:[[8.29, 8.47, 7.79]][[7.61, 88.24, 10.69]]
MSE(-DR):[[0.0, 0.18, -0.5]][[-0.68, 79.95, 2.4]]
==============


[[ 1.45  1.65  0.19  1.71 87.1   9.55]
 [ 0.59  0.8   0.21  2.7  85.29  7.74]
 [ 2.44  2.72  4.    0.51 89.62 12.07]
 [ 3.3   3.53  3.72  3.05 86.85  9.3 ]
 [ 8.29  8.47  7.79  7.6  88.24 10.69]]
[[ 1.43  1.65  0.23  1.71 87.1   9.55]
 [ 0.61  0.8   0.18  2.71 85.29  7.74]
 [ 2.45  2.72  4.04  0.51 89.62 12.07]
 [ 3.33  3.53  3.76  3.05 86.85  9.3 ]
 [ 8.33  8.47  7.81  7.61 88.24 10.69]]
[[ 1.43  1.65  0.22  1.71 87.1   9.55]
 [ 0.61  0.8   0.19  2.72 85.29  7.74]
 [ 2.45  2.72  4.    0.5  89.62 12.07]
 [ 3.34  3.53  3.73  3.05 86.85  9.3 ]
 [ 8.29  8.47  7.79  7.61 88.24 10.69]]
time spent until now: 18.0 mins


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

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1
```

```
0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
1 -th target; ^CTraceback (most recent call last):
  File "EC2.py", line 70, in <module>
Process Process-746:
Process Process-738:
Process Process-751:
    print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 62, in simu
    value_reps = rep_seeds(once, OPE_rep_times)
  File "/home/ubuntu/_uti_basic.py", line 119, in rep_seeds
    return list(map(fun, range(rep_times)))
  File "/home/ubuntu/simu_funs.py", line 58, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 202, in simu_once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 131, in V_DR
    r = arr(parmap(getOneRegionValue, range(N), 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>
    [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-745:
Process Process-742:
Process Process-739:
Process Process-747:
Process Process-741:
Process Process-737:
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 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)))
Traceback (most recent call last):
```

```
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  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/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  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/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  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 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  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 1173, in _run
    feed_dict_tensor, options, run_metadata)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  File "/home/ubuntu/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
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/.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 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/weight.py", line 297, 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/.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 1173, in _run
    feed_dict_tensor, options, run_metadata)
  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 85, in getOneRegionValue
    spatial = False)
  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 1350, in _do_run
    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/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  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 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)
  File "/home/ubuntu/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  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 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 950, in run
    run_metadata_ptr)
  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 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 1173, in _run
    feed_dict_tensor, options, 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)
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/main.py", line 85, in getOneRegionValue
    spatial = False)
  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/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  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/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  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 950, in run
    run_metadata_ptr)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
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/_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 1350, in _do_run
    run_metadata)
  File "/home/ubuntu/main.py", line 85, in getOneRegionValue
    spatial = False)
  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 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  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/weight.py", line 297, in train
    self.policy_ratio2: policy_ratio2
  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 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)
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 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 1429, in _call_tf_sessionrun
    run_metadata)
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/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, 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 1145, in _run
    not subfeed_t.get_shape().is_compatible_with(np_val.shape)):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/tensor_shape.py", line 1081, in is_compatible_with
    other = as_shape(other)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/tensor_shape.py", line 1204, in as_shape
    return TensorShape(shape)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/tensor_shape.py", line 774, in __init__
    self._dims = [as_dimension(d) for d in dims_iter]
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/tensor_shape.py", line 774, in <listcomp>
    self._dims = [as_dimension(d) for d in dims_iter]
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/tensor_shape.py", line 716, in as_dimension
    return Dimension(value)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/tensor_shape.py", line 185, in __init__
    self._value = int(value)
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/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 283, in train
    subsamples = np.random.choice(N, batch_size)
  File "mtrand.pyx", line 1152, in mtrand.RandomState.choice
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/numpy/core/fromnumeric.py", line 2772, in prod
    initial=initial)
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/numpy/core/fromnumeric.py", line 73, in _wrapreduction
    if type(obj) is not mu.ndarray:
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/main.py", line 85, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, 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 1145, in _run
    not subfeed_t.get_shape().is_compatible_with(np_val.shape)):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/ops.py", line 512, in get_shape
    return self.shape
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/ops.py", line 434, in shape
    if self._shape_val is None:
KeyboardInterrupt
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
ubuntu@ip-172-31-5-213:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
11:38, 04/01; num of cores:16

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A,  simple, M_in_R, u_O_u_D, mean_reversion, pois0] = [672, 10, 10, None, 0.2, 1, 1,
False, True, 10, False, False]


---------------------------------------
[pattern_seed, sd_R] = [0, 10]

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0
```

```
0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:75.539
O_threshold = 80
MC for this TARGET:[85.046, 0.073]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.73, -0.98, 0.49]][[3.18, -85.05, -9.51]]
std:[[0.14, 0.14, 0.15]][[0.19, 0.0, 0.13]]
MSE:[[0.74, 0.99, 0.51]][[3.19, 85.05, 9.51]]
MSE(-DR):[[0.0, 0.25, -0.23]][[2.45, 84.31, 8.77]]
**
==============


O_threshold = 90
MC for this TARGET:[82.899, 0.072]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.2, -0.03, 0.41]][[4.02, -82.9, -7.36]]
std:[[0.21, 0.21, 0.07]][[0.1, 0.0, 0.13]]
MSE:[[0.29, 0.21, 0.42]][[4.02, 82.9, 7.36]]
MSE(-DR):[[0.0, -0.08, 0.13]][[3.73, 82.61, 7.07]]
***
==============


O_threshold = 100
MC for this TARGET:[87.054, 0.072]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.8, -2.07, -3.4]][[0.6, -87.05, -11.52]]
std:[[0.1, 0.08, 0.18]][[0.1, 0.0, 0.13]]
MSE:[[1.8, 2.07, 3.4]][[0.61, 87.05, 11.52]]
MSE(-DR):[[0.0, 0.27, 1.6]][[-1.19, 85.25, 9.72]]
==============


O_threshold = 110
MC for this TARGET:[84.62, 0.073]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.56, -3.74, -4.2]][[-3.49, -84.62, -9.08]]
std:[[0.78, 0.74, 0.44]][[0.15, 0.0, 0.13]]
MSE:[[3.64, 3.81, 4.22]][[3.49, 84.62, 9.08]]
MSE(-DR):[[0.0, 0.17, 0.58]][[-0.15, 80.98, 5.44]]
```

==========

O_threshold = 120
MC for this TARGET:[84.968, 0.074]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.74, -7.87, -7.58]][[-7.49, -84.97, -9.43]]
std:[[0.66, 0.63, 0.4]][[0.21, 0.0, 0.13]]
MSE:[[7.77, 7.9, 7.59]][[7.49, 84.97, 9.43]]
MSE(-DR):[[0.0, 0.13, -0.18]][[-0.28, 77.2, 1.66]]
==========

[[ 0.74  0.99  0.51  3.19 85.05  9.51]
 [ 0.29  0.21  0.42  4.02 82.9   7.36]
 [ 1.8   2.07  3.4   0.61 87.05 11.52]
 [ 3.64  3.81  4.22  3.49 84.62  9.08]
 [ 7.77  7.9   7.59  7.49 84.97  9.43]]
time spent until now: 5.9 mins

---------------------------------------
[pattern_seed, sd_R] = [0, 10]

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

```
0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
1 -th target; 2 -th target; ^CTraceback (most recent call last):
  File "EC2.py", line 70, in <module>
Process Process-276:
    print_flag_target = False
  File "/home/ubuntu/simu_funs.py", line 62, in simu
    value_reps = rep_seeds(once, OPE_rep_times)
  File "/home/ubuntu/_uti_basic.py", line 119, in rep_seeds
Process Process-277:
    return list(map(fun, range(rep_times)))
  File "/home/ubuntu/simu_funs.py", line 58, in once
Process Process-282:
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 202, in simu_once
Process Process-278:
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 131, in V_DR
    r = arr(parmap(getOneRegionValue, range(N), 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>
    [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
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/main.py", line 59, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, 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 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 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
Process Process-288:
Process Process-280:
Process Process-283:
Traceback (most recent call last):
Process Process-286:
  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/main.py", line 59, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 237, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  File "/home/ubuntu/weight.py", line 297, 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)
Process Process-273:
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1173, in _run
```

```
    feed_dict_tensor, options, run_metadata)
ubuntu@ip-172-31-5-213:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
11:45, 04/01; num of cores:16

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A,  simple, M_in_R, u_O_u_D, mean_reversion, poisO] = [672, 10, 10, None, 0.2, 1, 1,
False, True, 10, False, False]


----------------------------------------
[pattern_seed, sd_R] = [0, 0.5]

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0
```

```
0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:75.526
O_threshold = 80
MC for this TARGET:[85.049, 0.019]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.65, -0.9, 0.51]][[3.08, -85.05, -9.52]]
std:[[0.01, 0.02, 0.21]][[0.19, 0.0, 0.13]]
MSE:[[0.65, 0.9, 0.55]][[3.09, 85.05, 9.52]]
MSE(-DR):[[0.0, 0.25, -0.1]][[2.44, 84.4, 8.87]]
**
==============


O_threshold = 90
MC for this TARGET:[82.903, 0.019]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.11, -0.13, 0.33]][[3.88, -82.9, -7.38]]
std:[[0.06, 0.06, 0.21]][[0.21, 0.0, 0.13]]
MSE:[[0.13, 0.14, 0.39]][[3.89, 82.9, 7.38]]
MSE(-DR):[[0.0, 0.01, 0.26]][[3.76, 82.77, 7.25]]
***
==============


O_threshold = 100
MC for this TARGET:[87.058, 0.017]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.81, -2.08, -3.27]][[0.47, -87.06, -11.53]]
std:[[0.11, 0.12, 0.2]][[0.16, 0.0, 0.13]]
MSE:[[1.81, 2.08, 3.28]][[0.5, 87.06, 11.53]]
MSE(-DR):[[0.0, 0.27, 1.47]][[-1.31, 85.25, 9.72]]
==============


O_threshold = 110
MC for this TARGET:[84.623, 0.018]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.67, -3.88, -4.24]][[-3.56, -84.62, -9.1]]
std:[[0.46, 0.44, 0.26]][[0.11, 0.0, 0.13]]
MSE:[[3.7, 3.9, 4.25]][[3.56, 84.62, 9.1]]
MSE(-DR):[[0.0, 0.2, 0.55]][[-0.14, 80.92, 5.4]]
==============


O_threshold = 120
MC for this TARGET:[84.972, 0.016]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.8, -7.93, -7.65]][[-7.54, -84.97, -9.45]]
std:[[0.38, 0.35, 0.13]][[0.13, 0.0, 0.13]]
MSE:[[7.81, 7.94, 7.65]][[7.54, 84.97, 9.45]]
MSE(-DR):[[0.0, 0.13, -0.16]][[-0.27, 77.16, 1.64]]
==============


[[ 0.65  0.9   0.55  3.09 85.05  9.52]
 [ 0.13  0.14  0.39  3.89 82.9   7.38]
 [ 1.81  2.08  3.28  0.5  87.06 11.53]
 [ 3.7   3.9   4.25  3.56 84.62  9.1 ]
 [ 7.81  7.94  7.65  7.54 84.97  9.45]]
time spent until now: 5.9 mins


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

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:
```

```
1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

---------------------------------------
Value of Behaviour policy:75.532
O_threshold = 80
MC for this TARGET:[85.047, 0.039]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.69, -0.94, 0.49]][[3.14, -85.05, -9.51]]
std:[[0.06, 0.08, 0.2]][[0.18, 0.0, 0.13]]
MSE:[[0.69, 0.94, 0.53]][[3.15, 85.05, 9.51]]
MSE(-DR):[[0.0, 0.25, -0.16]][[2.46, 84.36, 8.82]]
**
==============


O_threshold = 90
MC for this TARGET:[82.901, 0.039]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.13, -0.08, 0.34]][[3.94, -82.9, -7.37]]
std:[[0.12, 0.13, 0.13]][[0.15, 0.0, 0.13]]
MSE:[[0.18, 0.15, 0.36]][[3.94, 82.9, 7.37]]
```

```
MSE(-DR):[[0.0, -0.03, 0.18]][[3.76, 82.72, 7.19]]
***
==============


O_threshold = 100
MC for this TARGET:[87.056, 0.038]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.81, -2.07, -3.33]][[0.53, -87.06, -11.52]]
std:[[0.06, 0.1, 0.13]][[0.13, 0.0, 0.13]]
MSE:[[1.81, 2.07, 3.33]][[0.55, 87.06, 11.52]]
MSE(-DR):[[0.0, 0.26, 1.52]][[-1.26, 85.25, 9.71]]
==============


O_threshold = 110
MC for this TARGET:[84.622, 0.039]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.62, -3.81, -4.16]][[-3.55, -84.62, -9.09]]
std:[[0.61, 0.58, 0.33]][[0.13, 0.0, 0.13]]
MSE:[[3.67, 3.85, 4.17]][[3.55, 84.62, 9.09]]
MSE(-DR):[[0.0, 0.18, 0.5]][[-0.12, 80.95, 5.42]]
==============


O_threshold = 120
MC for this TARGET:[84.97, 0.039]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.74, -7.9, -7.55]][[-7.51, -84.97, -9.44]]
std:[[0.51, 0.49, 0.3]][[0.17, 0.0, 0.13]]
MSE:[[7.76, 7.92, 7.56]][[7.51, 84.97, 9.44]]
MSE(-DR):[[0.0, 0.16, -0.2]][[-0.25, 77.21, 1.68]]
==============


[[ 0.65  0.9   0.55  3.09 85.05  9.52]
 [ 0.13  0.14  0.39  3.89 82.9   7.38]
 [ 1.81  2.08  3.28  0.5  87.06 11.53]
 [ 3.7   3.9   4.25  3.56 84.62  9.1 ]
 [ 7.81  7.94  7.65  7.54 84.97  9.45]]
[[ 0.69  0.94  0.53  3.15 85.05  9.51]
 [ 0.18  0.15  0.36  3.94 82.9   7.37]
 [ 1.81  2.07  3.33  0.55 87.06 11.52]
 [ 3.67  3.85  4.17  3.55 84.62  9.09]
 [ 7.76  7.92  7.56  7.51 84.97  9.44]]
time spent until now: 11.7 mins


----------------------------------------
[pattern_seed, sd_R] = [0, 10]

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0
```

```
0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:75.539
O_threshold = 80
MC for this TARGET:[85.046, 0.073]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.71, -0.98, 0.51]][[3.13, -85.05, -9.51]]
std:[[0.12, 0.14, 0.19]][[0.16, 0.0, 0.13]]
MSE:[[0.72, 0.99, 0.54]][[3.13, 85.05, 9.51]]
MSE(-DR):[[0.0, 0.27, -0.18]][[2.41, 84.33, 8.79]]
**
==============


O_threshold = 90
MC for this TARGET:[82.899, 0.072]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.2, -0.03, 0.43]][[4.0, -82.9, -7.36]]
std:[[0.2, 0.21, 0.07]][[0.11, 0.0, 0.13]]
MSE:[[0.28, 0.21, 0.44]][[4.0, 82.9, 7.36]]
MSE(-DR):[[0.0, -0.07, 0.16]][[3.72, 82.62, 7.08]]
***
==============


O_threshold = 100
MC for this TARGET:[87.054, 0.072]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.77, -2.07, -3.36]][[0.58, -87.05, -11.52]]
std:[[0.11, 0.08, 0.2]][[0.12, 0.0, 0.13]]
MSE:[[1.77, 2.07, 3.37]][[0.59, 87.05, 11.52]]
MSE(-DR):[[0.0, 0.3, 1.6]][[-1.18, 85.28, 9.75]]
==============


O_threshold = 110
MC for this TARGET:[84.62, 0.073]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.52, -3.74, -4.11]][[-3.5, -84.62, -9.08]]
std:[[0.78, 0.74, 0.45]][[0.16, 0.0, 0.13]]
MSE:[[3.61, 3.81, 4.13]][[3.5, 84.62, 9.08]]
MSE(-DR):[[0.0, 0.2, 0.52]][[-0.11, 81.01, 5.47]]
```

==========


O_threshold = 120
MC for this TARGET:[84.968, 0.074]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.73, -7.87, -7.56]][[-7.5, -84.97, -9.43]]
std:[[0.66, 0.63, 0.43]][[0.21, 0.0, 0.13]]
MSE:[[7.76, 7.9, 7.57]][[7.5, 84.97, 9.43]]
MSE(-DR):[[0.0, 0.14, -0.19]][[-0.26, 77.21, 1.67]]
==========


[[ 0.65  0.9   0.55  3.09 85.05  9.52]
 [ 0.13  0.14  0.39  3.89 82.9   7.38]
 [ 1.81  2.08  3.28  0.5  87.06 11.53]
 [ 3.7   3.9   4.25  3.56 84.62  9.1 ]
 [ 7.81  7.94  7.65  7.54 84.97  9.45]]
[[ 0.69  0.94  0.53  3.15 85.05  9.51]
 [ 0.18  0.15  0.36  3.94 82.9   7.37]
 [ 1.81  2.07  3.33  0.55 87.06 11.52]
 [ 3.67  3.85  4.17  3.55 84.62  9.09]
 [ 7.76  7.92  7.56  7.51 84.97  9.44]]
[[ 0.72  0.99  0.54  3.13 85.05  9.51]
 [ 0.28  0.21  0.44  4.   82.9   7.36]
 [ 1.77  2.07  3.37  0.59 87.05 11.52]
 [ 3.61  3.81  4.13  3.5  84.62  9.08]
 [ 7.76  7.9   7.57  7.5  84.97  9.43]]
time spent until now: 17.6 mins


---------------------------------------
[pattern_seed, sd_R] = [0, 20]

max(u_O) =  156.6
O_threshold = 80
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 100
target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18

```
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
O_threshold = 120
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 0 0 0

0 1 0 0 0

0 0 0 0 1

number of reward locations:  8
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE
1 -th target; 2 -th target; 3 -th target; 4 -th target; 5 -th target; one rep DONE

----------------------------------------
Value of Behaviour policy:75.554
O_threshold = 80
MC for this TARGET:[85.042, 0.143]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.78, -1.06, 0.55]][[3.22, -85.04, -9.49]]
std:[[0.23, 0.27, 0.16]][[0.15, 0.0, 0.13]]
MSE:[[0.81, 1.09, 0.57]][[3.22, 85.04, 9.49]]
MSE(-DR):[[0.0, 0.28, -0.24]][[2.41, 84.23, 8.68]]
**
==============


O_threshold = 90
MC for this TARGET:[82.896, 0.143]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.26, 0.06, 0.5]][[4.14, -82.9, -7.34]]
std:[[0.33, 0.38, 0.05]][[0.01, 0.0, 0.13]]
MSE:[[0.42, 0.38, 0.5]][[4.14, 82.9, 7.34]]
MSE(-DR):[[0.0, -0.04, 0.08]][[3.72, 82.48, 6.92]]
***
==============


O_threshold = 100
MC for this TARGET:[87.05, 0.143]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.75, -2.06, -3.49]][[0.73, -87.05, -11.5]]
std:[[0.05, 0.04, 0.15]][[0.05, 0.0, 0.13]]
MSE:[[1.75, 2.06, 3.49]][[0.73, 87.05, 11.5]]
MSE(-DR):[[0.0, 0.31, 1.74]][[-1.02, 85.3, 9.75]]
==============


O_threshold = 110
MC for this TARGET:[84.616, 0.143]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.38, -3.6, -3.91]][[-3.46, -84.62, -9.06]]
std:[[1.12, 1.06, 0.7]][[0.17, 0.0, 0.13]]
MSE:[[3.56, 3.75, 3.97]][[3.46, 84.62, 9.06]]
MSE(-DR):[[0.0, 0.19, 0.41]][[-0.1, 81.06, 5.5]]
==============


O_threshold = 120
MC for this TARGET:[84.964, 0.145]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.61, -7.81, -7.43]][[-7.49, -84.96, -9.41]]
std:[[0.98, 0.93, 0.72]][[0.28, 0.0, 0.13]]
MSE:[[7.67, 7.87, 7.46]][[7.5, 84.96, 9.41]]
MSE(-DR):[[0.0, 0.2, -0.21]][[-0.17, 77.29, 1.74]]
==============


[[ 0.65  0.9   0.55  3.09 85.05  9.52]
 [ 0.13  0.14  0.39  3.89 82.9   7.38]
 [ 1.81  2.08  3.28  0.5  87.06 11.53]
 [ 3.7   3.9   4.25  3.56 84.62  9.1 ]
```

```
 [ 7.81  7.94  7.65  7.54 84.97  9.45]]
[[ 0.69  0.94  0.53  3.15 85.05  9.51]
 [ 0.18  0.15  0.36  3.94 82.9   7.37]
 [ 1.81  2.07  3.33  0.55 87.06 11.52]
 [ 3.67  3.85  4.17  3.55 84.62  9.09]
 [ 7.76  7.92  7.56  7.51 84.97  9.44]]
[[ 0.72  0.99  0.54  3.13 85.05  9.51]
 [ 0.28  0.21  0.44  4.   82.9   7.36]
 [ 1.77  2.07  3.37  0.59 87.05 11.52]
 [ 3.61  3.81  4.13  3.5  84.62  9.08]
 [ 7.76  7.9   7.57  7.5  84.97  9.43]]
[[ 0.81  1.09  0.57  3.22 85.04  9.49]
 [ 0.42  0.38  0.5   4.14 82.9   7.34]
 [ 1.75  2.06  3.49  0.73 87.05 11.5 ]
 [ 3.56  3.75  3.97  3.46 84.62  9.06]
 [ 7.67  7.87  7.46  7.5  84.96  9.41]]
time spent until now: 23.5 mins


---------------------------------------
[pattern_seed, sd_R] = [1, 0.5]

max(u_O) =  141.0
O_threshold = 80
means of Order:

137.7 88.0 89.5 80.3 118.3

62.8 141.0 85.4 106.0 94.6

133.3 65.9 93.3 92.1 124.8

79.8 96.1 83.5 100.3 111.8

79.8 125.1 119.1 110.0 119.1

target policy:

1 1 1 1 1

0 1 1 1 1

1 0 1 1 1

0 1 1 1 1

0 1 1 1 1

number of reward locations:  21
O_threshold = 90
target policy:

1 0 0 0 1

0 1 0 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

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

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 1 1

0 1 1 1 1

number of reward locations:  12
O_threshold = 110
target policy:

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1
```