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

max(u_O) =  145.8
O_threshold = 125
number of reward locations:  2
target 1 in 1 DONE!

----------------------------------------
Value of Behaviour policy:55.239
O_threshold = 125
MC for this TARGET:[68.265, 0.065]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-13.91, -13.91, -13.96]][[-20.03, -68.26, -13.03]]
std:[[0.74, 0.73, 0.48]][[0.37, 0.0, 0.23]]
MSE:[[13.93, 13.93, 13.97]][[20.03, 68.26, 13.03]]
MSE(-DR):[[0.0, 0.0, 0.04]][[6.1, 54.33, -0.9]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[13.93 13.93 13.97 20.03 68.26 13.03]]


time spent until now: 10.3 mins

12:31, 04/12

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

max(u_O) =  145.8
O_threshold = 125
number of reward locations:  2
^CProcess Process-25:
Process Process-17:
Traceback (most recent call last):
Process Process-31:
Process Process-28:
  File "EC2.py", line 83, in <module>
Process Process-24:
Process Process-29:
Process Process-19:
Process Process-30:
Process Process-21:
    with_MF = with_MF, with_NO_MARL = with_NO_MARL, with_IS = with_IS,
  File "/home/ubuntu/simu_funs.py", line 63, in simu
    value_reps = parmap(once, range(OPE_rep_times), n_cores)
  File "/home/ubuntu/_uti_basic.py", line 80, in parmap
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/_uti_basic.py", line 80, 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
ubuntu@ip-172-31-0-69:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
12:38, 04/12; num of cores:16
2_u_O_u_D_20

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 20, 0.5, 1.5, 20, None]

thre_range, sd_R_range, day_range:  [[100, 110], [0, 20], [7]]

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
target 1 in 2 DONE!
target 2 in 2 DONE!

----------------------------------------
Value of Behaviour policy:49.429
O_threshold = 100
MC for this TARGET:[54.842, 0.069]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.35, -1.47, -1.87]][[-2.93, -54.84, -5.41]]
std:[[0.63, 0.61, 0.43]][[0.31, 0.0, 0.2]]
MSE:[[1.49, 1.59, 1.92]][[2.95, 54.84, 5.41]]
MSE(-DR):[[0.0, 0.1, 0.43]][[1.46, 53.35, 3.92]]
***
==============
O_threshold = 110
MC for this TARGET:[52.191, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[-2.61, -2.65, -3.28]][[-6.03, -52.19, -2.76]]
std:[[0.58, 0.55, 0.44]][[0.31, 0.0, 0.2]]
MSE:[[2.67, 2.71, 3.31]][[6.04, 52.19, 2.77]]
MSE(-DR):[[0.0, 0.04, 0.64]][[3.37, 49.52, 0.1]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.49  1.59  1.92  2.95 54.84  5.41]
 [ 2.67  2.71  3.31  6.04 52.19  2.77]]


time spent until now: 20.5 mins

12:59, 04/12

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
target 1 in 2 DONE!
target 2 in 2 DONE!

---------------------------------------
Value of Behaviour policy:49.401
O_threshold = 100
MC for this TARGET:[54.873, 0.241]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.33, -1.44, -2.03]][[-2.98, -54.87, -5.47]]
std:[[1.0, 1.03, 0.62]][[0.34, 0.0, 0.23]]
MSE:[[1.66, 1.77, 2.12]][[3.0, 54.87, 5.47]]
MSE(-DR):[[0.0, 0.11, 0.46]][[1.34, 53.21, 3.81]]
***
==============
O_threshold = 110
MC for this TARGET:[52.223, 0.236]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.48, -2.54, -3.39]][[-6.12, -52.22, -2.82]]
std:[[1.01, 1.03, 0.8]][[0.39, 0.0, 0.23]]
MSE:[[2.68, 2.74, 3.48]][[6.13, 52.22, 2.83]]
MSE(-DR):[[0.0, 0.06, 0.8]][[3.45, 49.54, 0.15]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.49  1.59  1.92  2.95 54.84  5.41]
 [ 2.67  2.71  3.31  6.04 52.19  2.77]]


[[ 1.66  1.77  2.12  3.   54.87  5.47]
 [ 2.68  2.74  3.48  6.13 52.22  2.83]]


time spent until now: 41.3 mins

13:20, 04/12
ubuntu@ip-172-31-0-69:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
13:21, 04/12; num of cores:16
2_u_O_u_D_20

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 20, 0.5, 1.5, 20, None]

thre_range, sd_R_range, day_range:  [[95], [0, 20], [7]]

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

max(u_O) =  145.8
O_threshold = 95
number of reward locations:  12
target 1 in 1 DONE!

---------------------------------------
Value of Behaviour policy:49.429
O_threshold = 95
MC for this TARGET:[55.996, 0.069]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.82, 1.62, 0.66]][[1.22, -56.0, -6.57]]
std:[[0.58, 0.58, 0.39]][[0.25, 0.0, 0.2]]
MSE:[[1.91, 1.72, 0.77]][[1.25, 56.0, 6.57]]
MSE(-DR):[[0.0, -0.19, -1.14]][[-0.66, 54.09, 4.66]]
==============
[[ 1.91  1.72  0.77  1.25 56.    6.57]]


time spent until now: 10.3 mins
```

```
13:31, 04/12

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

max(u_O) =  145.8
O_threshold = 95
number of reward locations:  12
^CProcess Process-29:
Process Process-28:
Traceback (most recent call last):
  File "EC2.py", line 87, in <module>
Process Process-26:
    with_MF = with_MF, with_NO_MARL = with_NO_MARL, with_IS = with_IS,
  File "/home/ubuntu/simu_funs.py", line 63, in simu
    value_reps = parmap(once, range(OPE_rep_times), n_cores)
  File "/home/ubuntu/_uti_basic.py", line 80, in parmap
Process Process-19:
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/_uti_basic.py", line 80, in <listcomp>
Process Process-31:
    [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-20:
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
Traceback (most recent call last):
Process Process-23:
  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 67, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 61, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 213, in simu_once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 158, in V_DR
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 158, in <listcomp>
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 80, in getOneRegionValue
    CV_QV = CV_QV, penalty_range = penalty, spatial = True)
  File "/home/ubuntu/main.py", line 303, in computeQV
    validation_set = valid_tuples)
  File "/home/ubuntu/main.py", line 433, in computeQV_basic
    left = (ECKQ1.T.dot(ECKQ1) + np.vstack((np.hstack((T * lam * KQ, zeros((2 * T, 1)))), zeros((1, 2 * T + 1))))) # Left part of (\hat{
\alpha}, \hat{\eta})
KeyboardInterrupt
Process Process-24:
ubuntu@ip-172-31-0-69:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
13:41, 04/12; num of cores:16
QV_tuning

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 20, 0.5, 1.5, 20, None]

thre_range, sd_R_range, day_range:  [[80, 90], [0, 20], [7]]

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

max(u_O) =  145.8
O_threshold = 80
number of reward locations:  20
O_threshold = 90
number of reward locations:  14
target 1 in 2 DONE!
target 2 in 2 DONE!

--------------------------------------
Value of Behaviour policy:49.429
O_threshold = 80
MC for this TARGET:[62.196, 0.052]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.41, 2.05, 1.03]][[6.16, -62.2, -12.77]]
std:[[0.57, 0.58, 0.37]][[0.33, 0.0, 0.2]]
MSE:[[2.48, 2.13, 1.09]][[6.17, 62.2, 12.77]]
MSE(-DR):[[0.0, -0.35, -1.39]][[3.69, 59.72, 10.29]]
**
==============
O_threshold = 90
MC for this TARGET:[58.335, 0.061]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.57, 1.28, -0.3]][[1.59, -58.34, -8.91]]
std:[[0.37, 0.37, 0.32]][[0.3, 0.0, 0.2]]
MSE:[[1.61, 1.33, 0.44]][[1.62, 58.34, 8.91]]
MSE(-DR):[[0.0, -0.28, -1.17]][[0.01, 56.73, 7.3]]
**
```

```
==============
[[ 2.48  2.13  1.09  6.17 62.2  12.77]
 [ 1.61  1.33  0.44  1.62 58.34  8.91]]


time spent until now: 20.6 mins

14:01, 04/12

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

max(u_O) =  145.8
O_threshold = 80
number of reward locations:  20
O_threshold = 90
number of reward locations:  14
target 1 in 2 DONE!
^CProcess Process-27:
Process Process-24:
Process Process-28:
Traceback (most recent call last):
  File "EC2.py", line 87, in <module>
Process Process-19:
Process Process-25:
Process Process-31:
    with_MF = with_MF, with_NO_MARL = with_NO_MARL, with_IS = with_IS)
  File "/home/ubuntu/simu_funs.py", line 63, in simu
Process Process-30:
    value_reps = parmap(once, range(OPE_rep_times), n_cores)
  File "/home/ubuntu/_uti_basic.py", line 80, in parmap
    p.start()
  File "/home/ubuntu/_uti_basic.py", line 80, in <listcomp>
    p.start()
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
Process Process-23:
Process Process-26:
Traceback (most recent call last):
    if not self._sem.acquire(block, timeout):
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/_uti_basic.py", line 67, in fun
    i, x = q_in.get()
  File "/home/ubuntu/simu_funs.py", line 61, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 213, in simu_once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 158, in V_DR
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 158, in <listcomp>
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 114, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 262, in getWeight
    epsilon = epsilon, spatial = spatial, mean_field = mean_field)
  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 1173, in _run
    feed_dict_tensor, options, run_metadata)
Process Process-21:
ubuntu@ip-172-31-0-69:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
14:15, 04/12; num of cores:16
QV_tuning

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 20, 0.5, 1.5, 20, None]

thre_range, sd_R_range, day_range:  [[100, 110], [0, 20], [7]]

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
True True [0.01, 0.01] 680.8029368514186
True True [0.0001, 0.01] 408.68299740119903
True True [0.01, 0.01] 630.5566880942781
True True [0.01, 0.01] 362.20174424299694
True True [0.001, 0.01] 314.7110697526123
True True [0.01, 0.01] 485.7765241566265
True True [0.0001, 0.01] 517.6544627234454
True True [0.001, 0.01] 526.8510577368103
```

```
False True [0.01, 0.01] 1287.7035049856102
False True [0.0001, 0.0001] 994.6883799580428
False True [0.0001, 0.0001] 1043.6202152015212
True True [0.01, 0.01] 284.8783522075937
False True [0.01, 0.01] 1351.968159863979
False True [0.0001, 0.01] 1194.235631601693
True True [0.01, 0.01] 444.0973997672539
False True [0.0001, 0.0001] 616.4872293114607
False True [0.001, 0.0001] 653.6784327249756
True True [0.01, 0.01] 209.62791125324463
True True [0.01, 0.01] 32.697814545686484
False True [0.001, 0.0001] 1192.46852622704
True True [0.01, 0.01] 30.81134253056524
False True [0.001, 0.0001] 1025.2956749607295
False True [0.0001, 0.0001] 508.7464758575O5
True True [0.01, 0.01] 76.12685018174871
True True [0.01, 0.01] 30.95425762381462
True True [0.01, 0.01] 45.87594482933613
True True [0.01, 0.01] 53.620792524032304
False True [0.001, 0.0001] 1011.8519873606297
False True [0.01, 0.01] 1290.3793787807863
True True [0.01, 0.01] 104.79489866109134
True True [0.01, 0.01] 66.47528184687567
True True [0.01, 0.01] 111.48462465103844
False True [0.01, 0.01] 1288.5095464468432
True True [0.01, 0.01] 49.48476416059076
False True [0.0001, 0.0001] 549.901096446393
True True [0.01, 0.01] 74.70007277917625
False True [0.01, 0.01] 593.4344850419633
True True [0.01, 0.01] 40.64645529893535
True True [0.01, 0.01] 38.74387847356467
False True [0.0001, 0.0001] 495.1521553599148
False True [0.01, 0.01] 652.3080765613697
False True [0.01, 0.01] 652.2559621151589
False True [0.01, 0.01] 583.4844566329332
True True [0.01, 0.01] 115.64980856709313
False True [0.0001, 0.01] 554.1906506203394
True True [0.01, 0.01] 61.28161185207294
False True [0.01, 0.01] 505.8293113721268
False True [0.01, 0.01] 609.9215898216828
True True [0.01, 0.01] 127.39884466515846
False True [0.001, 0.0001] 560.5553974770173
False True [0.01, 0.01] 583.6877943637568
True True [0.0001, 0.001] 128.5999685377276
True True [0.0001, 0.01] 138.9800923307041
False True [0.001, 0.0001] 503.3170810216661
False True [0.0001, 0.0001] 478.1283661986577
True True [0.0001, 0.01] 120.91468519892481
True True [0.01, 0.01] 77.82145060492374
True True [0.01, 0.01] 304.9798863768027
True True [0.0001, 0.01] 105.35499702076521
True True [0.01, 0.01] 202.61028653004288
False True [0.01, 0.01] 654.4067725564194
False True [0.01, 0.01] 581.2156494864932
True True [0.01, 0.01] 241.3950475020508
True True [0.0001, 0.01] 236.66096851762836
False True [0.0001, 0.01] 617.15541354885
True True [0.01, 0.01] 105.94771031264858
False True [0.0001, 0.0001] 313.2462873267549
True True [0.01, 0.01] 117.054436241656
False True [0.0001, 0.0001] 87.69376057805265
True True [0.01, 0.01] 152.56460928525863
True True [0.01, 0.01] 74.20919329470725
False True [0.0001, 0.01] 380.3304132618996
False True [0.0001, 0.0001] 214.72747279927745
False True [0.001, 0.0001] 433.5704138870526
True True [0.01, 0.01] 253.98105269207872
False True [0.0001, 0.0001] 222.53448244919252
False True [0.0001, 0.0001] 287.763460926935
False True [0.0001, 0.0001] 189.2904308034963
True True [0.01, 0.01] 115.48589845127681
False True [0.0001, 0.0001] 268.08345899468856
False True [0.0001, 0.0001] 104.45843217248789
True True [0.01, 0.01] 101.93184407831217
False True [0.01, 0.01] 399.8551250548444
False True [0.01, 0.01] 364.08713496149375
target 2 in 2 DONE!
False True [0.01, 0.01] 528.40540465294
False True [0.01, 0.01] 540.9102720434602
False True [0.01, 0.01] 465.360037850563
False True [0.0001, 0.0001] 188.8396097107634

--------------------------------------
Value of Behaviour policy:49.429
O_threshold = 100
MC for this TARGET:[54.842, 0.069]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.11, -3.29, -1.86]][[-2.71, -54.84, -5.41]]
std:[[0.42, 0.52, 0.4]][[0.31, 0.0, 0.2]]
```

```
MSE:[[2.15, 3.33, 1.9]][[2.73, 54.84, 5.41]]
MSE(-DR):[[0.0, 1.18, -0.25]][[0.58, 52.69, 3.26]]
**
==============
O_threshold = 110
MC for this TARGET:[52.191, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.74, -3.52, -3.29]][[-5.58, -52.19, -2.76]]
std:[[0.5, 0.52, 0.45]][[0.3, 0.0, 0.2]]
MSE:[[2.79, 3.56, 3.32]][[5.59, 52.19, 2.77]]
MSE(-DR):[[0.0, 0.77, 0.53]][[2.8, 49.4, -0.02]]
***
==============
[[ 2.15  3.33  1.9   2.73 54.84  5.41]
 [ 2.79  3.56  3.32  5.59 52.19  2.77]]


time spent until now: 25.0 mins

14:40, 04/12

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
True True [0.01, 0.01] 525.5347741552695
True True [0.01, 0.01] 624.7862150544042
True True [0.01, 0.01] 631.932699504631
True True [0.01, 0.01] 880.7806618012767
True True [0.01, 0.01] 689.160617154415
True True [0.001, 0.0001] 635.7852512767427
True True [0.01, 0.01] 593.603826268494
True True [0.01, 0.01] 404.2415733298936
True True [0.01, 0.01] 161.7343207439318
True True [0.0001, 0.01] 774.7815616747572
True True [0.01, 0.01] 597.079497183111
True True [0.0001, 0.01] 445.86371701401
True True [0.01, 0.01] 831.472918888381
True True [0.01, 0.01] 408.3613410768038
True True [0.01, 0.01] 333.71996981979134
True True [0.01, 0.01] 283.3745796352207
False True [0.01, 0.01] 831.0179395135709
False True [0.01, 0.01] 580.8384292507985
False True [0.001, 0.0001] 62.58139965487662
False True [0.001, 0.0001] 235.27747411857803
False True [0.0001, 0.0001] 192.3304215334818
False True [0.01, 0.01] 716.5217253694146
False True [0.0001, 0.0001] 442.7416665564586
False True [0.0001, 0.0001] 98.9508422462438
False True [0.01, 0.01] 481.830104019601
False True [0.01, 0.0001] 417.30159975617227
False True [0.01, 0.01] 690.6910010577974
False True [0.01, 0.01] 514.9479914589234
False True [0.01, 0.0001] 573.0479128785066
False True [0.0001, 0.0001] 335.92755551330794
False True [0.01, 0.01] 678.0900554057064
False True [0.0001, 0.01] 646.8526476226743
True True [0.01, 0.01] 448.514921986943
True True [0.01, 0.001] 95.99938165290132
True True [0.01, 0.01] 266.9461336261806
True True [0.0001, 0.001] 176.02336135557897
True True [0.01, 0.01] 232.28934689659633
True True [0.001, 0.01] 144.66641906775112
True True [0.001, 0.01] 192.23323233813267
True True [0.01, 0.01] 169.4387910885576
True True [0.0001, 0.01] 161.01274059670627
True True [0.01, 0.01] 327.00194549840535
True True [0.01, 0.01] 395.556786525591
True True [0.0001, 0.001] 162.4043011732272
True True [0.01, 0.01] 264.0174576405618
True True [0.01, 0.01] 121.94749446365009
True True [0.0001, 0.01] 211.939283462592
True True [0.01, 0.01] 174.39334567062986
False True [0.0001, 0.0001] 952.3988626726193
False True [0.001, 0.0001] 1329.695262069618
False True [0.0001, 0.0001] 933.4050184011735
False True [0.01, 0.01] 1045.5584493620522
False True [0.01, 0.01] 917.6022846836076
False True [0.0001, 0.01] 906.6365345112449
False True [0.0001, 0.0001] 1111.6612817935484
False True [0.01, 0.01] 1328.5256761955436
False True [0.01, 0.01] 909.948839600207
False True [0.01, 0.01] 1201.6175753389166
False True [0.0001, 0.01] 1027.161113901652
False True [0.01, 0.01] 800.0791955289792
```

```
True True [0.01, 0.01] 389.3239877248638
True True [0.01, 0.01] 600.1413168692854
False True [0.001, 0.0001] 420.5917104098501
False True [0.01, 0.01] 231.5218876945987
True True [0.01, 0.01] 108.92337866517454
True True [0.01, 0.01] 493.3032951708805
False True [0.0001, 0.0001] 233.2945523518328
False True [0.0001, 0.0001] 242.01233728193756
False True [0.01, 0.01] 678.101522601372
True True [0.0001, 0.01] 189.9325656084826
False True [0.001, 0.0001] 114.96921483400408
True True [0.0001, 0.01] 126.42984651714761
False True [0.01, 0.01] 387.30165661745957
True True [0.01, 0.01] 366.4366849034996
True True [0.01, 0.01] 352.39063908904296
False True [0.01, 0.0001] 229.64668050185253
target 2 in 2 DONE!
False True [0.001, 0.0001] 107.3299076292994
False True [0.001, 0.0001] 248.20066650300308
False True [0.001, 0.0001] 80.23377329053747
False True [0.001, 0.0001] 43.810594565127985
True True [0.01, 0.01] 108.21524890632905
True True [0.01, 0.01] 221.70828909882292
False True [0.0001, 0.0001] 182.55022979605894
False True [0.01, 0.01] 512.8871514664368
False True [0.01, 0.01] 695.6776941615094
False True [0.001, 0.0001] 330.72936378386044
False True [0.0001, 0.01] 662.3561245838342
False True [0.0001, 0.0001] 440.39316002092227

--------------------------------------
Value of Behaviour policy:49.401
O_threshold = 100
MC for this TARGET:[54.873, 0.241]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.37, -3.57, -2.04]][[-2.87, -54.87, -5.47]]
std:[[0.67, 0.58, 0.64]][[0.37, 0.0, 0.23]]
MSE:[[2.46, 3.62, 2.14]][[2.89, 54.87, 5.47]]
MSE(-DR):[[0.0, 1.16, -0.32]][[0.43, 52.41, 3.01]]
**
==============
O_threshold = 110
MC for this TARGET:[52.223, 0.236]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.79, -3.58, -3.38]][[-5.81, -52.22, -2.82]]
std:[[0.73, 0.58, 0.81]][[0.45, 0.0, 0.23]]
MSE:[[2.88, 3.63, 3.48]][[5.83, 52.22, 2.83]]
MSE(-DR):[[0.0, 0.75, 0.6]][[2.95, 49.34, -0.05]]
***
==============
[[ 2.15  3.33  1.9   2.73 54.84  5.41]
 [ 2.79  3.56  3.32  5.59 52.19  2.77]]


[[ 2.46  3.62  2.14  2.89 54.87  5.47]
 [ 2.88  3.63  3.48  5.83 52.22  2.83]]


time spent until now: 50.0 mins

15:05, 04/12
ubuntu@ip-172-31-0-69:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
15:07, 04/12; num of cores:16
QV_tuning

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 20, 0.5, 1.5, 20, None]

thre_range, sd_R_range, day_range:  [[100, 110], [0, 20], [7]]

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
True True [0.1, 0.1] 212.90267862766157
True True [0.1, 0.1] 155.47534894148765
True True [0.1, 0.1] 192.14412896527796
True True [0.001, 0.1] 199.17739191114214
True True [0.1, 0.1] 157.81857267653052
True True [0.1, 0.1] 167.8286299045977
True True [0.1, 0.1] 140.99407524237836
True True [0.1, 0.1] 155.1090493906437
True True [0.001, 0.1] 239.58511717634138
True True [0.1, 0.1] 137.8740261652483
True True [0.1, 0.1] 170.00525239828033
True True [0.001, 0.1] 177.82822345523115
```

```
True True [0.1, 0.01] 75.51989839751016
False True [0.1, 0.1] 408.6691198992729
False True [0.1, 0.1] 392.6989910443459
False True [0.1, 0.1] 415.8930650965123
True True [0.01, 0.1] 141.23313696465007
False True [0.1, 0.1] 439.63273260353816
True True [0.1, 0.01] 42.821417598606544
False True [0.1, 0.1] 408.13683247458994
True True [0.01, 0.1] 29.089462491466463
True True [0.1, 0.1] 57.69692135420757
False True [0.1, 0.1] 430.4974951982149
True True [0.1, 0.001] 45.0624590994101
True True [0.1, 0.01] 20.47947270828911
True True [0.1, 0.01] 25.290331283998047
True True [0.01, 0.1] 26.835692038124485
False True [0.1, 0.1] 381.0167019551725
True True [0.1, 0.1] 44.27263938098592
True True [0.1, 0.01] 35.13022792946317
True True [0.1, 0.1] 33.60996881009237
False True [0.1, 0.1] 475.6243487206228
True True [0.001, 0.1] 50.61358489250388
False True [0.1, 0.1] 233.3166138456894
True True [0.1, 0.01] 23.41888005739899
True True [0.1, 0.01] 28.282697958856517
False True [0.1, 0.1] 214.29740278357664
True True [0.1, 0.01] 18.908389392664297
False True [0.1, 0.1] 421.2028721412075
False True [0.1, 0.1] 200.93292827349208
False True [0.1, 0.1] 233.82848596987094
False True [0.1, 0.1] 213.06639364333873
False True [0.1, 0.1] 188.6382576838023
True True [0.1, 0.01] 22.07255703244421
False True [0.1, 0.1] 206.773710703048
False True [0.1, 0.1] 227.0503618149835
False True [0.1, 0.1] 237.7479707809142
True True [0.01, 0.1] 26.911415318130224
False True [0.1, 0.1] 213.1009849296966
False True [0.1, 0.1] 208.69198673907172
True True [0.001, 0.1] 76.52084936954907
False True [0.1, 0.1] 243.46465428592956
False True [0.1, 0.1] 225.49839706625124
True True [0.1, 0.01] 90.94894665142779
True True [0.1, 0.01] 16.628780302690238
True True [0.1, 0.01] 66.61708636951155
True True [0.1, 0.01] 88.23263613943948
True True [0.001, 0.1] 78.20287772751905
True True [0.001, 0.1] 116.12205147859532
False True [0.1, 0.1] 182.42260931137383
True True [0.001, 0.1] 126.5422082973901
True True [0.1, 0.01] 111.61043838663886
True True [0.001, 0.1] 90.92757295684561
False True [0.1, 0.1] 186.56911517865086
True True [0.001, 0.1] 74.42795054358704
False True [0.1, 0.1] 214.26100887859087
True True [0.1, 0.01] 61.57988980731813
True True [0.001, 0.1] 86.01896219840428
True True [0.1, 0.01] 63.085144932939706
False True [0.1, 0.1] 243.0248011644061
False True [0.1, 0.1] 175.23661278641487
False True [0.1, 0.1] 197.39294697323723
False True [0.1, 0.1] 192.60789248390557
False True [0.1, 0.1] 196.7001480362485
False True [0.1, 0.1] 234.76946860842037
False True [0.001, 0.1] 222.5505042812751
False True [0.1, 0.1] 239.23920216801613
True True [0.1, 0.01] 151.60326800787652
False True [0.1, 0.1] 197.21829541061845
True True [0.01, 0.1] 104.69874926661925
False True [0.1, 0.1] 185.0836432326024
False True [0.1, 0.1] 220.8621083835225
False True [0.1, 0.1] 229.05826737649997
False True [0.1, 0.1] 234.58821594723582
True True [0.001, 0.1] 57.804128393332775
False True [0.1, 0.1] 189.54096655662698
False True [0.1, 0.1] 222.51501782778135
False True [0.1, 0.1] 170.5364774667645
target 2 in 2 DONE!

--------------------------------------
Value of Behaviour policy:49.429
O_threshold = 100
MC for this TARGET:[54.842, 0.069]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.19, -5.65, -1.87]][[-2.92, -54.84, -5.41]]
std:[[0.39, 0.28, 0.41]][[0.28, 0.0, 0.2]]
MSE:[[2.22, 5.66, 1.91]][[2.93, 54.84, 5.41]]
MSE(-DR):[[0.0, 3.44, -0.31]][[0.71, 52.62, 3.19]]
**
==============
```

```
O_threshold = 110
MC for this TARGET:[52.191, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.65, -3.66, -3.28]][[-5.56, -52.19, -2.76]]
std:[[0.51, 0.28, 0.45]][[0.27, 0.0, 0.2]]
MSE:[[2.7, 3.67, 3.31]][[5.57, 52.19, 2.77]]
MSE(-DR):[[0.0, 0.97, 0.61]][[2.87, 49.49, 0.07]]
***
==============
[[ 2.22  5.66  1.91  2.93 54.84  5.41]
 [ 2.7   3.67  3.31  5.57 52.19  2.77]]


time spent until now: 24.7 mins

15:31, 04/12

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
True True [0.001, 0.1] 261.19907705945934
True True [0.1, 0.1] 246.77466230342006
True True [0.1, 0.1] 277.27090579703633
True True [0.1, 0.1] 269.66248093762323
True True [0.1, 0.1] 279.08593454867537
True True [0.01, 0.1] 197.49454185740498
True True [0.1, 0.1] 187.86216296026467
True True [0.01, 0.1] 199.8510896694141
True True [0.001, 0.1] 193.8511161477109
True True [0.1, 0.1] 183.01496856070685
True True [0.1, 0.1] 219.4915895430756
True True [0.1, 0.1] 190.15540836153102
True True [0.001, 0.1] 193.77861589875008
True True [0.1, 0.1] 218.18529019492536
True True [0.1, 0.1] 127.5428454970841
True True [0.001, 0.1] 113.51597318620293
False True [0.001, 0.001] 196.85478928725632
False True [0.1, 0.1] 281.34385913934204
False True [0.1, 0.1] 341.8785156298028
False True [0.1, 0.1] 350.30582299738523
False True [0.1, 0.1] 268.97930589341536
False True [0.1, 0.1] 294.30579952624805
False True [0.1, 0.1] 196.0139572932421
False True [0.1, 0.1] 295.6936084917163
False True [0.1, 0.1] 280.98497327503867
False True [0.1, 0.1] 202.79375494614195
False True [0.1, 0.1] 148.27432269911606
False True [0.1, 0.1] 245.0063275103406
False True [0.1, 0.1] 331.70141529149424
False True [0.1, 0.1] 191.37900968772198
False True [0.1, 0.1] 187.61728435800816
False True [0.1, 0.1] 268.1626204606277
True True [0.1, 0.01] 157.6592400506011
True True [0.1, 0.1] 146.02943564054797
True True [0.1, 0.1] 152.46228756627025
True True [0.01, 0.001] 95.99938165290132
True True [0.1, 0.01] 103.68114737535501
True True [0.1, 0.1] 76.73527910744573
True True [0.1, 0.1] 113.46247739716216
True True [0.1, 0.01] 86.52310417682038
True True [0.001, 0.1] 115.29129065264023
True True [0.1, 0.1] 86.64266882101985
True True [0.01, 0.1] 126.76650058187256
True True [0.1, 0.1] 105.09459026819498
True True [0.01, 0.1] 139.113328594529
True True [0.1, 0.01] 69.63408629793022
True True [0.01, 0.1] 78.01494408660974
True True [0.1, 0.1] 115.8255181251042
False True [0.1, 0.1] 476.52625428838553
False True [0.1, 0.1] 358.28012938232075
False True [0.1, 0.1] 596.1695718304721
False True [0.1, 0.1] 682.9677671094371
False True [0.1, 0.1] 352.4869455156506
False True [0.1, 0.1] 433.95229730563443
False True [0.1, 0.1] 329.82023776948006
False True [0.1, 0.1] 301.1061935560968
False True [0.1, 0.1] 406.0203027659187
False True [0.1, 0.1] 419.22797244408684
False True [0.1, 0.1] 497.5684723528525
False True [0.1, 0.1] 565.7682760047604
False True [0.1, 0.1] 406.6090298320341
False True [0.1, 0.1] 432.93513866475416
False True [0.1, 0.1] 401.1266524918189
False True [0.1, 0.1] 492.1179968111363
```

```
^CProcess Process-28:
Process Process-26:
Process Process-25:
Traceback (most recent call last):
Process Process-32:
  File "EC2.py", line 88, in <module>
Process Process-23:
Process Process-30:
    dim_S_plus_Ts = 3 + 3, epsilon = 1e-6, # Fixed
  File "/home/ubuntu/simu_funs.py", line 63, in simu
    value_reps = parmap(once, range(OPE_rep_times), n_cores)
  File "/home/ubuntu/_uti_basic.py", line 83, in parmap
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/_uti_basic.py", line 83, in <listcomp>
Traceback (most recent call last):
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
  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 70, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 61, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 213, in simu_once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 158, in V_DR
    r = arr([getOneRegionValue(i) for i in range(N)])
    if not self._sem.acquire(block, timeout):
  File "/home/ubuntu/main.py", line 158, in <listcomp>
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 80, in getOneRegionValue
    CV_QV = CV_QV, penalty_range = penalty, spatial = True)
  File "/home/ubuntu/main.py", line 303, in computeQV
    n_neigh = n_neigh, spatial = spatial, mean_field = mean_field,
KeyboardInterrupt
  File "/home/ubuntu/main.py", line 480, in computeQV_basic
    count = 0
  File "/home/ubuntu/main.py", line 480, in <listcomp>
    count = 0
KeyboardInterrupt
Process Process-21:
Process Process-31:
Process Process-18:
Process Process-19:
ubuntu@ip-172-31-0-69:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
Traceback (most recent call last):
  File "EC2.py", line 57, in <module>
    from simu_funs import *
  File "/home/ubuntu/simu_funs.py", line 4, in <module>
    from main import *
  File "/home/ubuntu/main.py", line 279
    global count = 0
                 ^
SyntaxError: invalid syntax
ubuntu@ip-172-31-0-69:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
15:33, 04/12; num of cores:16
QV_tuning

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 20, 0.5, 1.5, 20, None]

thre_range, sd_R_range, day_range:  [[100, 110], [0, 20], [7]]

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  6
False True [0.1, 10.0] 8.918129068305046
False True [0.1, 1.0] 50.747949521520624
False True [0.1, 10.0] 15.833823623103314
False True [0.1, 10.0] 50.03023958916105
False True [0.1, 1.0] 53.73583785043404
False True [1.0, 0.1] 42.51116085957887
False True [0.1, 10.0] 55.25508917072625
False True [0.1, 10.0] 21.495922774172183
False True [0.1, 1.0] 41.041949273778904
False True [0.1, 1.0] 57.24407511725116
False True [0.1, 1.0] 114.65601451703995
False True [1.0, 0.1] 67.16365057499436
False True [0.1, 1.0] 94.90499679897931
False True [1.0, 0.1] 37.777989436821386
False True [0.1, 10.0] 43.082970613839485
False True [0.1, 1.0] 60.529334823065554
False True [0.1, 10.0] 102.71494739271004
```

```
False True [1.0, 0.1] 35.059556217090694
False True [0.1, 1.0] 178.67360560628745
False True [0.1, 1.0] 118.97171836593358
False True [0.1, 10.0] 68.7898363926511
False True [0.1, 1.0] 241.74649181984233
False True [0.1, 1.0] 77.25173758111228
False True [0.1, 10.0] 57.18246493863398
False True [0.1, 1.0] 68.43102712035117
False True [0.1, 1.0] 95.14035016240346
False True [0.1, 10.0] 125.44806994853798
False True [0.1, 1.0] 86.48331955274281
False True [0.1, 10.0] 94.21778460629216
False True [1.0, 0.1] 16.12062903991746
False True [0.1, 10.0] 49.63872981475241
False True [0.1, 1.0] 144.39629646689406
target 1 in 2 DONE!
False True [0.1, 10.0] 27.060642370187967
False True [0.1, 10.0] 43.59624354451957
False True [0.1, 10.0] 11.857969469655508
False True [0.1, 10.0] 30.016272461174157
False True [0.1, 10.0] 42.63953240686675
False True [0.1, 10.0] 13.65917360864835
False True [0.1, 10.0] 23.63824011375958
False True [0.1, 10.0] 12.197121779068326
False True [0.1, 10.0] 43.8378704633925
False True [0.1, 1.0] 67.21609646909128
False True [0.1, 10.0] 23.026049085014634
False True [0.1, 10.0] 20.324926571443225
False True [0.1, 10.0] 67.12633260479123
False True [0.1, 10.0] 38.342725910592534
False True [0.1, 1.0] 87.92546909107779
False True [0.1, 10.0] 33.81136044306206
False True [0.1, 10.0] 55.18997085939871
False True [0.1, 10.0] 57.764856639544966
False True [0.1, 10.0] 15.312499598809055
False True [0.1, 1.0] 107.14584191506505
False True [0.1, 10.0] 61.158924051101195
False True [0.1, 10.0] 56.726576682135594
False True [0.1, 1.0] 113.37659994650872
False True [0.1, 1.0] 78.83303263060291
False True [0.1, 10.0] 73.86715878092272
False True [0.1, 1.0] 73.74833532246404
False True [0.1, 10.0] 30.918892127244938
False True [0.1, 10.0] 44.22606080074154
False True [0.1, 1.0] 127.12578043667187
False True [0.1, 10.0] 74.73455240594662
False True [0.1, 1.0] 103.87390133589864
False True [0.1, 10.0] 84.94448038803384
False True [1.0, 0.1] 33.54656282042497
False True [0.1, 1.0] 73.84457850571735
False True [0.1, 10.0] 48.314800587800114
False True [0.1, 10.0] 38.73982368466477
False True [0.1, 10.0] 31.03384004740924
False True [0.1, 10.0] 34.00044110342321
False True [1.0, 0.1] 36.88928384179255
False True [0.1, 1.0] 107.68022874872445
False True [0.1, 10.0] 87.57887247686385
False True [0.1, 10.0] 33.82824467732234
False True [0.1, 10.0] 76.85628874961401
False True [0.1, 10.0] 21.489436635702248
False True [0.1, 1.0] 98.19391431293434
False True [0.1, 10.0] 39.08131490815628
False True [0.1, 10.0] 15.220643017227747
False True [0.1, 1.0] 91.95312415838733
target 2 in 2 DONE!

--------------------------------------
Value of Behaviour policy:49.429
O_threshold = 100
MC for this TARGET:[54.842, 0.069]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.31, -5.8, -1.87]][[-2.85, -54.84, -5.41]]
std:[[0.36, 0.34, 0.42]][[0.32, 0.0, 0.2]]
MSE:[[2.34, 5.81, 1.92]][[2.87, 54.84, 5.41]]
MSE(-DR):[[0.0, 3.47, -0.42]][[0.53, 52.5, 3.07]]
**
==============
O_threshold = 110
MC for this TARGET:[52.191, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.22, -3.1, -3.3]][[-5.61, -52.19, -2.76]]
std:[[0.44, 0.27, 0.45]][[0.33, 0.0, 0.2]]
MSE:[[3.25, 3.11, 3.33]][[5.62, 52.19, 2.77]]
MSE(-DR):[[0.0, -0.14, 0.08]][[2.37, 48.94, -0.48]]
***
==============
[[ 2.34  5.81  1.92  2.87 54.84  5.41]
 [ 3.25  3.11  3.33  5.62 52.19  2.77]]
```