```
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 67, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 60, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 212, in simu_once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 149, in V_DR
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 149, in <listcomp>
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 84, in getOneRegionValue
    epsilon = epsilon)
  File "/home/ubuntu/main.py", line 238, in getWeight
    w_hidden = w_hidden, Learning_rate = lr, reg_weight = reg_weight)
  File "/home/ubuntu/weight.py", line 101, in __init__
    self.sess.run(tf.global_variables_initializer())
  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 1339, in _run_fn
    self._extend_graph()
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/client/session.py", line 1374, in _extend_graph
    tf_session.ExtendSession(self._session)
KeyboardInterrupt
ubuntu@ip-172-31-67-47:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
20:10, 04/06; num of cores:16

final sd_R trend for[0, 10, 20]80 the same

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, pois0, u_O_u_D], sd_R_range, t_func] = [None,
16, None, None, None, 30, 0.5, 1, [True, False, True, 10], [0, 10, 20], None]


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

max(u_O) =  168.1
O_threshold = 90
number of reward locations:  21
O_threshold = 100
number of reward locations:  18
O_threshold = 110
number of reward locations:  15
O_threshold = 120
number of reward locations:  10
O_threshold = 130
number of reward locations:  6
target 1 in 5 DONE!
target 2 in 5 DONE!
target 3 in 5 DONE!
target 4 in 5 DONE!
target 5 in 5 DONE!

----------------------------------------
Value of Behaviour policy:71.28
O_threshold = 90
MC for this TARGET:[81.187, 0.076]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.76, 1.59, 1.22]][[4.97, -81.19, -9.91]]
std:[[0.83, 0.84, 0.42]][[0.26, 0.0, 0.14]]
MSE:[[1.95, 1.8, 1.29]][[4.98, 81.19, 9.91]]
MSE(-DR):[[0.0, -0.15, -0.66]][[3.03, 79.24, 7.96]]
**
==============
O_threshold = 100
MC for this TARGET:[85.436, 0.073]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.26, -0.49, -2.39]][[1.57, -85.44, -14.16]]
std:[[0.51, 0.52, 0.23]][[0.26, 0.0, 0.14]]
MSE:[[0.57, 0.71, 2.4]][[1.59, 85.44, 14.16]]
MSE(-DR):[[0.0, 0.14, 1.83]][[1.02, 84.87, 13.59]]
***
==============
O_threshold = 110
MC for this TARGET:[85.779, 0.07]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.19, -1.46, -3.26]][[-0.53, -85.78, -14.5]]
std:[[0.4, 0.41, 0.25]][[0.3, 0.0, 0.14]]
```

```
MSE:[[1.26, 1.52, 3.27]][[0.61, 85.78, 14.5]]
MSE(-DR):[[0.0, 0.26, 2.01]][[-0.65, 84.52, 13.24]]
==============
O_threshold = 120
MC for this TARGET:[79.988, 0.07]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.37, -2.51, -2.81]][[-3.2, -79.99, -8.71]]
std:[[0.48, 0.5, 0.28]][[0.33, 0.0, 0.14]]
MSE:[[2.42, 2.56, 2.82]][[3.22, 79.99, 8.71]]
MSE(-DR):[[0.0, 0.14, 0.4]][[0.8, 77.57, 6.29]]
***
==============
O_threshold = 130
MC for this TARGET:[84.356, 0.066]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-10.32, -10.34, -9.69]][[-13.55, -84.36, -13.08]]
std:[[0.67, 0.67, 0.33]][[0.35, 0.0, 0.14]]
MSE:[[10.34, 10.36, 9.7]][[13.55, 84.36, 13.08]]
MSE(-DR):[[0.0, 0.02, -0.64]][[3.21, 74.02, 2.74]]
**
==============
[[ 1.95  1.8   1.29  4.98 81.19  9.91]
 [ 0.57  0.71  2.4   1.59 85.44 14.16]
 [ 1.26  1.52  3.27  0.61 85.78 14.5 ]
 [ 2.42  2.56  2.82  3.22 79.99  8.71]
 [10.34 10.36  9.7  13.55 84.36 13.08]]


time spent until now: 49.3 mins


---------------------------------------
[pattern_seed, day, sd_R] = [0, 7, 10]

max(u_O) =  168.1
O_threshold = 90
number of reward locations:  21
O_threshold = 100
number of reward locations:  18
O_threshold = 110
number of reward locations:  15
O_threshold = 120
number of reward locations:  10
O_threshold = 130
number of reward locations:  6
^CProcess Process-26:
Process Process-30:
Process Process-17:
Process Process-19:
Process Process-31:
Traceback (most recent call last):
  File "EC2.py", line 87, in <module>
Process Process-18:
    dim_S_plus_Ts = 3 + 3, epsilon = 1e-6, # Fixed
  File "/home/ubuntu/simu_funs.py", line 62, 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)]
Process Process-22:
  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
Process Process-32:
    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()
Process Process-25:
  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 60, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 212, in simu_once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/main.py", line 149, in V_DR
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 149, in <listcomp>
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 78, in getOneRegionValue
    CV_QV = CV_QV, penalty_range = penalty, spatial = True)
Traceback (most recent call last):
  File "/home/ubuntu/main.py", line 294, in computeQV
    validation_set = valid_tuples)
  File "/home/ubuntu/main.py", line 329, in computeQV_basic
    Z = np.array([np.concatenate((a[0], a[3], [a[1]], [a[4]])) for a in tuples_i]) # T * p. [S, Ts, A, Ta]
  File "/home/ubuntu/main.py", line 329, in <listcomp>
```

```
    Z = np.array([np.concatenate((a[0], a[3], [a[1]], [a[4]])) for a in tuples_i]) # T * p. [S, Ts, A, Ta]
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
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 60, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 212, in simu_once
    inner_parallel = inner_parallel)
Process Process-27:
  File "/home/ubuntu/main.py", line 149, in V_DR
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 149, in <listcomp>
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 108, in getOneRegionValue
    spatial = False)
  File "/home/ubuntu/main.py", line 253, 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 1109, in _run
    for feed, feed_val in feed_dict.items():
KeyboardInterrupt
ubuntu@ip-172-31-67-47:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
21:07, 04/06; num of cores:16

final sd_R trend for[5, 15] the same

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, pois0, u_O_u_D], sd_R_range, t_func] = [None,
16, None, None, None, 30, 0.5, 1, [True, False, True, 10], [5, 15], None]


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

max(u_O) =  168.8
O_threshold = 80
number of reward locations:  15
O_threshold = 85
number of reward locations:  14
O_threshold = 90
number of reward locations:  12
O_threshold = 95
number of reward locations:  12
O_threshold = 100
number of reward locations:  9
target 1 in 5 DONE!
target 2 in 5 DONE!
target 3 in 5 DONE!
target 4 in 5 DONE!
target 5 in 5 DONE!

----------------------------------------
Value of Behaviour policy:57.689
O_threshold = 80
MC for this TARGET:[68.363, 0.083]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.24, -0.46, -1.06]][[1.09, -68.36, -10.67]]
std:[[0.52, 0.53, 0.32]][[0.31, 0.0, 0.2]]
MSE:[[0.57, 0.7, 1.11]][[1.13, 68.36, 10.67]]
MSE(-DR):[[0.0, 0.13, 0.54]][[0.56, 67.79, 10.1]]
***
==============
O_threshold = 85
MC for this TARGET:[68.378, 0.088]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.03, -0.26, -1.29]][[0.05, -68.38, -10.69]]
std:[[0.44, 0.43, 0.31]][[0.34, 0.0, 0.2]]
MSE:[[0.44, 0.5, 1.33]][[0.34, 68.38, 10.69]]
MSE(-DR):[[0.0, 0.06, 0.89]][[-0.1, 67.94, 10.25]]
==============
O_threshold = 90
MC for this TARGET:[66.722, 0.092]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.13, -0.32, -0.79]][[-0.52, -66.72, -9.03]]
std:[[0.42, 0.42, 0.27]][[0.27, 0.0, 0.2]]
MSE:[[0.44, 0.53, 0.83]][[0.59, 66.72, 9.03]]
MSE(-DR):[[0.0, 0.09, 0.39]][[0.15, 66.28, 8.59]]
***
==============
O_threshold = 95
MC for this TARGET:[66.722, 0.092]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.13, -0.32, -0.81]][[-0.51, -66.72, -9.03]]
```

```
std:[[0.43, 0.42, 0.29]][[0.27, 0.0, 0.2]]
MSE:[[0.45, 0.53, 0.86]][[0.58, 66.72, 9.03]]
MSE(-DR):[[0.0, 0.08, 0.41]][[0.13, 66.27, 8.58]]
***
==============
O_threshold = 100
MC for this TARGET:[66.956, 0.097]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.76, -2.88, -3.2]][[-4.27, -66.96, -9.27]]
std:[[0.47, 0.48, 0.29]][[0.27, 0.0, 0.2]]
MSE:[[2.8, 2.92, 3.21]][[4.28, 66.96, 9.27]]
MSE(-DR):[[0.0, 0.12, 0.41]][[1.48, 64.16, 6.47]]
***
==============
[[ 0.57  0.7   1.11  1.13 68.36 10.67]
 [ 0.44  0.5   1.33  0.34 68.38 10.69]
 [ 0.44  0.53  0.83  0.59 66.72  9.03]
 [ 0.45  0.53  0.86  0.58 66.72  9.03]
 [ 2.8   2.92  3.21  4.28 66.96  9.27]]


time spent until now: 49.3 mins


--------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 15]

max(u_O) =  168.8
O_threshold = 80
number of reward locations:  15
O_threshold = 85
number of reward locations:  14
O_threshold = 90
number of reward locations:  12
O_threshold = 95
number of reward locations:  12
O_threshold = 100
number of reward locations:  9
target 1 in 5 DONE!
target 2 in 5 DONE!
target 3 in 5 DONE!
target 4 in 5 DONE!
target 5 in 5 DONE!

--------------------------------------
Value of Behaviour policy:57.675
O_threshold = 80
MC for this TARGET:[68.379, 0.181]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.33, -0.56, -1.07]][[1.01, -68.38, -10.7]]
std:[[0.93, 0.95, 0.48]][[0.42, 0.0, 0.21]]
MSE:[[0.99, 1.1, 1.17]][[1.09, 68.38, 10.7]]
MSE(-DR):[[0.0, 0.11, 0.18]][[0.1, 67.39, 9.71]]
***
==============
O_threshold = 85
MC for this TARGET:[68.394, 0.187]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.13, -0.37, -1.31]][[-0.01, -68.39, -10.72]]
std:[[0.85, 0.83, 0.56]][[0.44, 0.0, 0.21]]
MSE:[[0.86, 0.91, 1.42]][[0.44, 68.39, 10.72]]
MSE(-DR):[[0.0, 0.05, 0.56]][[-0.42, 67.53, 9.86]]
==============
O_threshold = 90
MC for this TARGET:[66.737, 0.191]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.34, -0.52, -0.91]][[-0.56, -66.74, -9.06]]
std:[[0.87, 0.86, 0.58]][[0.35, 0.0, 0.21]]
MSE:[[0.93, 1.0, 1.08]][[0.66, 66.74, 9.06]]
MSE(-DR):[[0.0, 0.07, 0.15]][[-0.27, 65.81, 8.13]]
==============
O_threshold = 95
MC for this TARGET:[66.737, 0.191]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.32, -0.52, -0.89]][[-0.56, -66.74, -9.06]]
std:[[0.87, 0.86, 0.61]][[0.35, 0.0, 0.21]]
MSE:[[0.93, 1.0, 1.08]][[0.66, 66.74, 9.06]]
MSE(-DR):[[0.0, 0.07, 0.15]][[-0.27, 65.81, 8.13]]
==============
O_threshold = 100
MC for this TARGET:[66.972, 0.196]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.7, -2.84, -3.14]][[-4.29, -66.97, -9.3]]
std:[[0.84, 0.87, 0.47]][[0.36, 0.0, 0.21]]
MSE:[[2.83, 2.97, 3.17]][[4.31, 66.97, 9.3]]
MSE(-DR):[[0.0, 0.14, 0.34]][[1.48, 64.14, 6.47]]
***
==============
[[ 0.57  0.7   1.11  1.13 68.36 10.67]
```

```
 [ 0.44  0.5   1.33  0.34 68.38 10.69]
 [ 0.44  0.53  0.83  0.59 66.72  9.03]
 [ 0.45  0.53  0.86  0.58 66.72  9.03]
 [ 2.8   2.92  3.21  4.28 66.96  9.27]]


[[ 0.99  1.1   1.17  1.09 68.38 10.7 ]
 [ 0.86  0.91  1.42  0.44 68.39 10.72]
 [ 0.93  1.    1.08  0.66 66.74  9.06]
 [ 0.93  1.    1.08  0.66 66.74  9.06]
 [ 2.83  2.97  3.17  4.31 66.97  9.3 ]]


time spent until now: 98.6 mins

ubuntu@ip-172-31-67-47:~$
```