```
  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-10:
ubuntu@ip-172-31-13-78:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
09:14, 04/13; num of cores:16
try_diff_seed

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, 0, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[140, 180], [0, 20, 40], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]]

---------------------------------------
[pattern_seed, day, sd_R] = [None, 7, 0]

max(u_O) =  200
O_threshold = 140
means of Order:

80 80 80 80 80

80 150 80 150 80

80 80 200 80 80

80 150 80 150 80

80 80 80 80 80

target policy:

0 0 0 0 0

0 1 0 1 0

0 0 1 0 0

0 1 0 1 0

0 0 0 0 0

number of reward locations:  5
O_threshold = 180
number of reward locations:  1
target 1 in 2 DONE!
target 2 in 2 DONE!

---------------------------------------
Value of Behaviour policy:56.382
O_threshold = 140
MC for this TARGET:[63.66, 0.087]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.16, -4.29, -4.91]][[-5.02, -63.66, -7.28]]
std:[[0.36, 0.4, 0.28]][[0.34, 0.0, 0.23]]
MSE:[[4.18, 4.31, 4.92]][[5.03, 63.66, 7.28]]
MSE(-DR):[[0.0, 0.13, 0.74]][[0.85, 59.48, 3.1]]
***
==============
O_threshold = 180
MC for this TARGET:[73.771, 0.051]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-18.75, -18.75, -18.63]][[-26.07, -73.77, -17.39]]
std:[[0.94, 1.0, 0.49]][[0.41, 0.0, 0.23]]
MSE:[[18.77, 18.78, 18.64]][[26.07, 73.77, 17.39]]
MSE(-DR):[[0.0, 0.01, -0.13]][[7.3, 55.0, -1.38]]
**
==============
[[ 4.18  4.31  4.92  5.03 63.66  7.28]
 [18.77 18.78 18.64 26.07 73.77 17.39]]


time spent until now: 20.4 mins

09:34, 04/13

---------------------------------------
[pattern_seed, day, sd_R] = [None, 7, 20]

max(u_O) =  200
```

O_threshold = 140
means of Order:

80 80 80 80 80

80 150 80 150 80

80 80 200 80 80

80 150 80 150 80

80 80 80 80 80

target policy:

0 0 0 0 0

0 1 0 1 0

0 0 1 0 0

0 1 0 1 0

0 0 0 0 0

number of reward locations:  5
O_threshold = 180
number of reward locations:  1
^CProcess Process-29:
Traceback (most recent call last):
  File "EC2.py", line 100, in <module>
Process Process-25:
Process Process-23:
    with_MF = with_MF, with_NO_MARL = with_NO_MARL, with_IS = with_IS)
Process Process-24:
  File "/home/ubuntu/simu_funs.py", line 72, in simu
Process Process-21:
    value_reps = parmap(once, range(OPE_rep_times), n_cores)
  File "/home/ubuntu/_uti_basic.py", line 83, in parmap
Process Process-17:
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/_uti_basic.py", line 83, in <listcomp>
Process Process-26:
Process Process-28:
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
Process Process-20:
Process Process-18:
Process Process-22:
Process Process-27:
ubuntu@ip-172-31-13-78:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
09:36, 04/13; num of cores:16
try_diff_seed

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, 0, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[140, 180], [0, 20, 40], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]]

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

max(u_O) =  150
O_threshold = 140
means of Order:

100 80 100 80 100

80 120 80 120 80

100 80 150 80 100

80 120 80 120 80

100 80 100 80 100

target policy:

0 0 0 0 0

0 0 0 0 0

0 0 1 0 0

0 0 0 0 0

0 0 0 0 0

```
KeyboardInterrupt
ubuntu@ip-172-31-13-78:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
09:37, 04/13; num of cores:16
try_diff_seed

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, 0, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[99, 119, 149], [0, 20, 40], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]]


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

max(u_O) =  150
O_threshold = 99
means of Order:

100 80 100 80 100

80 120 80 120 80

100 80 150 80 100

80 120 80 120 80

100 80 100 80 100

target policy:

1 0 1 0 1

0 1 0 1 0

1 0 1 0 1

0 1 0 1 0

1 0 1 0 1

number of reward locations:  13
O_threshold = 119
number of reward locations:  5
O_threshold = 149
number of reward locations:  1
target 1 in 3 DONE!
target 2 in 3 DONE!
target 3 in 3 DONE!


---------------------------------------
Value of Behaviour policy:55.348
O_threshold = 99
MC for this TARGET:[66.505, 0.088]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.27, 0.06, -1.26]][[0.14, -66.5, -11.16]]
std:[[0.43, 0.43, 0.3]][[0.32, 0.0, 0.2]]
MSE:[[0.51, 0.43, 1.3]][[0.35, 66.5, 11.16]]
MSE(-DR):[[0.0, -0.08, 0.79]][[-0.16, 65.99, 10.65]]
==============
O_threshold = 119
MC for this TARGET:[55.146, 0.055]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.29, -2.28, -2.46]][[-3.0, -55.15, 0.2]]
std:[[0.34, 0.33, 0.34]][[0.24, 0.0, 0.2]]
MSE:[[2.32, 2.3, 2.48]][[3.01, 55.15, 0.28]]
MSE(-DR):[[0.0, -0.02, 0.16]][[0.69, 52.83, -2.04]]
***
==============
O_threshold = 149
MC for this TARGET:[70.312, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-16.89, -16.88, -17.09]][[-23.96, -70.31, -14.96]]
std:[[1.0, 1.04, 0.54]][[0.36, 0.0, 0.2]]
MSE:[[16.92, 16.91, 17.1]][[23.96, 70.31, 14.96]]
MSE(-DR):[[0.0, -0.01, 0.18]][[7.04, 53.39, -1.96]]
***
==============
[[ 0.51  0.43  1.3   0.35 66.5  11.16]
 [ 2.32  2.3   2.48  3.01 55.15  0.28]
 [16.92 16.91 17.1  23.96 70.31 14.96]]


time spent until now: 29.0 mins

10:06, 04/13

---------------------------------------
[pattern_seed, day, sd_R] = [0, 7, 20]

max(u_O) =  150
O_threshold = 99
```

means of Order:

100 80 100 80 100

80 120 80 120 80

100 80 150 80 100

80 120 80 120 80

100 80 100 80 100

target policy:

1 0 1 0 1

0 1 0 1 0

1 0 1 0 1

0 1 0 1 0

1 0 1 0 1

number of reward locations:  13
O_threshold = 119
number of reward locations:  5
O_threshold = 149
number of reward locations:  1
target 1 in 3 DONE!
target 2 in 3 DONE!
target 3 in 3 DONE!

_____

Value of Behaviour policy:55.32
O_threshold = 99
MC for this TARGET:[66.537, 0.254]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.68, 0.45, −1.13]][[0.25, −66.54, −11.22]]
std:[[0.76, 0.77, 0.57]][[0.52, 0.0, 0.3]]
MSE:[[1.02, 0.89, 1.27]][[0.58, 66.54, 11.22]]
MSE(−DR):[[0.0, −0.13, 0.25]][[−0.44, 65.52, 10.2]]
==============
O_threshold = 119
MC for this TARGET:[55.177, 0.252]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[−2.1, −2.14, −2.44]][[−3.09, −55.18, 0.14]]
std:[[0.92, 0.89, 0.72]][[0.48, 0.0, 0.3]]
MSE:[[2.29, 2.32, 2.54]][[3.13, 55.18, 0.33]]
MSE(−DR):[[0.0, 0.03, 0.25]][[0.84, 52.89, −1.96]]
***
==============
O_threshold = 149
MC for this TARGET:[70.343, 0.251]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[−16.93, −17.0, −17.18]][[−24.1, −70.34, −15.02]]
std:[[1.11, 1.14, 0.75]][[0.58, 0.0, 0.3]]
MSE:[[16.97, 17.04, 17.2]][[24.11, 70.34, 15.02]]
MSE(−DR):[[0.0, 0.07, 0.23]][[7.14, 53.37, −1.95]]
***
==============
[[ 0.51  0.43  1.3   0.35 66.5  11.16]
 [ 2.32  2.3   2.48  3.01 55.15  0.28]
 [16.92 16.91 17.1  23.96 70.31 14.96]]


[[ 1.02  0.89  1.27  0.58 66.54 11.22]
 [ 2.29  2.32  2.54  3.13 55.18  0.33]
 [16.97 17.04 17.2  24.11 70.34 15.02]]


time spent until now: 58.2 mins

10:35, 04/13

_____
[pattern_seed, day, sd_R] = [0, 7, 40]

max(u_O) =  150
O_threshold = 99
means of Order:

100 80 100 80 100

80 120 80 120 80

100 80 150 80 100

80 120 80 120 80

```
100 80 100 80 100

target policy:

1 0 1 0 1

0 1 0 1 0

1 0 1 0 1

0 1 0 1 0

1 0 1 0 1

number of reward locations:  13
O_threshold = 119
number of reward locations:  5
O_threshold = 149
number of reward locations:  1
target 1 in 3 DONE!
^CProcess Process-42:
Process Process-33:
Process Process-44:
Process Process-47:
Traceback (most recent call last):
  File "EC2.py", line 101, in <module>
Process Process-41:
Process Process-37:
    with_MF = with_MF, with_NO_MARL = with_NO_MARL, with_IS = with_IS)
  File "/home/ubuntu/simu_funs.py", line 83, in simu
    value_reps = parmap(once, range(OPE_rep_times), n_cores)
Process Process-36:
  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>
    [q_in.put((None, None)) for _ in range(nprocs)]
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/queues.py", line 82, in put
Traceback (most recent call last):
    if not self._sem.acquire(block, timeout):
Process Process-34:
KeyboardInterrupt
Process Process-48:
Process Process-38:
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 70, in fun
    q_out.put((i, f(x)))
  File "/home/ubuntu/simu_funs.py", line 81, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 233, in simu_once
    inner_parallel = inner_parallel)
Process Process-40:
  File "/home/ubuntu/main.py", line 158, in V_DR
    r = arr([getOneRegionValue(i) for i in range(N)])
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
ubuntu@ip-172-31-13-78:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
10:47, 04/13; num of cores:16
try_diff_seed

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, 0, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[99, 119, 149], [0, 20, 40], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]]

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

max(u_O) =  150
O_threshold = 99
means of Order:

100 80 100 80 100

80 120 80 120 80

100 80 150 80 100

80 120 80 120 80

100 80 100 80 100

target policy:

1 0 1 0 1
```

```
0 1 0 1 0

1 0 1 0 1

0 1 0 1 0

1 0 1 0 1

number of reward locations:  13
O_threshold = 119
number of reward locations:  5
O_threshold = 149
number of reward locations:  1
target 1 in 3 DONE!
target 2 in 3 DONE!
target 3 in 3 DONE!

_____
Value of Behaviour policy:50.084
O_threshold = 99
MC for this TARGET:[59.95, 0.067]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.06, -1.27, -2.41]][[-1.45, -59.95, -9.87]]
std:[[0.35, 0.34, 0.28]][[0.33, 0.0, 0.16]]
MSE:[[1.12, 1.31, 2.43]][[1.49, 59.95, 9.87]]
MSE(-DR):[[0.0, 0.19, 1.31]][[0.37, 58.83, 8.75]]
***
==============
O_threshold = 119
MC for this TARGET:[48.785, 0.061]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.22, -2.2, -2.55]][[-3.51, -48.78, 1.3]]
std:[[0.31, 0.33, 0.31]][[0.22, 0.0, 0.16]]
MSE:[[2.24, 2.22, 2.57]][[3.52, 48.78, 1.31]]
MSE(-DR):[[0.0, -0.02, 0.33]][[1.28, 46.54, -0.93]]
***
==============
O_threshold = 149
MC for this TARGET:[57.916, 0.046]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-14.69, -14.62, -15.19]][[-21.8, -57.92, -7.83]]
std:[[0.81, 0.84, 0.47]][[0.29, 0.0, 0.16]]
MSE:[[14.71, 14.64, 15.2]][[21.8, 57.92, 7.83]]
MSE(-DR):[[0.0, -0.07, 0.49]][[7.09, 43.21, -6.88]]
***
==============
****************** THIS SETTING IS GOOD ******************
[[ 1.12  1.31  2.43  1.49 59.95  9.87]
 [ 2.24  2.22  2.57  3.52 48.78  1.31]
 [14.71 14.64 15.2  21.8  57.92  7.83]]


time spent until now: 29.0 mins

11:16, 04/13

_____
[pattern_seed, day, sd_R] = [0, 7, 20]

max(u_O) =  150
O_threshold = 99
means of Order:

100 80 100 80 100

80 120 80 120 80

100 80 150 80 100

80 120 80 120 80

100 80 100 80 100

target policy:

1 0 1 0 1

0 1 0 1 0

1 0 1 0 1

0 1 0 1 0

1 0 1 0 1

number of reward locations:  13
O_threshold = 119
number of reward locations:  5
```