```
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/ops/gradients_impl.py", line 158, in gradients
    unconnected_gradients)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/ops/gradients_util.py", line 731, in _GradientsHelper
    lambda: grad_fn(op, *out_grads))
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/ops/gradients_util.py", line 403, in _MaybeCompile
    return grad_fn()  # Exit early
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/ops/gradients_util.py", line 731, in <lambda>
    lambda: grad_fn(op, *out_grads))
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/ops/math_grad.py", line 142, in _MeanGrad
    output_shape = array_ops.shape(op.outputs[0])
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/ops/array_ops.py", line 330, in shape
    return shape_internal(input, name, optimize=True, out_type=out_type)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/ops/array_ops.py", line 357, in shape_internal
    return constant(input_shape.as_list(), out_type, name=name)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/constant_op.py", line 246, in constant
    allow_broadcast=True)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/constant_op.py", line 290, in _constant_impl
    name=name).outputs[0]
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/util/deprecation.py", line 507, in new_func
    return func(*args, **kwargs)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/ops.py", line 3616, in create_op
    op_def=op_def)
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/ops.py", line 2040, in __init__
    self._graph._add_op(self)  # pylint: disable=protected-access
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/ops.py", line 3240, in _add_op
    with self._lock:
KeyboardInterrupt
Process Process-1:
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 84, in once
    inner_parallel = inner_parallel)
  File "/home/ubuntu/simu_funs.py", line 252, 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)
  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
ubuntu@ip-172-31-6-10:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
13:40, 04/13; num of cores:96
sd_u_O_30

Basic setting:[rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, u_O_u_D, t_func] = [16, None, None, 40, 0.5, 1.5, 0, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[95, 100, 110, 120, 125, 140], [0, 20, 40], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]]

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

max(u_O) =  191.7
O_threshold = 95
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:
```

```
0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 0 1 0 0

number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
number of reward locations:  3
O_threshold = 140
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!


----------------------------------------
Value of Behaviour policy:50.343
O_threshold = 95
MC for this TARGET:[62.565, 0.076]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.25, -0.48, -1.6]][[-0.78, -62.56, -12.22]]
std:[[0.42, 0.42, 0.35]][[0.3, 0.0, 0.3]]
MSE:[[0.49, 0.64, 1.64]][[0.84, 62.56, 12.22]]
MSE(-DR):[[0.0, 0.15, 1.15]][[0.35, 62.07, 11.73]]
***
==============
O_threshold = 100
MC for this TARGET:[61.987, 0.078]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.19, -1.39, -2.28]][[-2.67, -61.99, -11.64]]
std:[[0.59, 0.6, 0.44]][[0.33, 0.0, 0.3]]
MSE:[[1.33, 1.51, 2.32]][[2.69, 61.99, 11.64]]
MSE(-DR):[[0.0, 0.18, 0.99]][[1.36, 60.66, 10.31]]
***
==============
O_threshold = 110
MC for this TARGET:[61.518, 0.077]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.08, -3.25, -4.27]][[-5.17, -61.52, -11.17]]
std:[[0.59, 0.63, 0.39]][[0.31, 0.0, 0.3]]
MSE:[[3.14, 3.31, 4.29]][[5.18, 61.52, 11.17]]
MSE(-DR):[[0.0, 0.17, 1.15]][[2.04, 58.38, 8.03]]
***
==============
O_threshold = 120
MC for this TARGET:[59.981, 0.073]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.95, -3.09, -3.82]][[-5.85, -59.98, -9.64]]
std:[[0.62, 0.62, 0.47]][[0.32, 0.0, 0.3]]
MSE:[[3.01, 3.15, 3.85]][[5.86, 59.98, 9.64]]
MSE(-DR):[[0.0, 0.14, 0.84]][[2.85, 56.97, 6.63]]
***
==============
O_threshold = 125
MC for this TARGET:[60.214, 0.078]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.25, -7.31, -7.3]][[-12.62, -60.21, -9.87]]
std:[[1.01, 1.04, 0.45]][[0.33, 0.0, 0.3]]
MSE:[[7.32, 7.38, 7.31]][[12.62, 60.21, 9.87]]
MSE(-DR):[[0.0, 0.06, -0.01]][[5.3, 52.89, 2.55]]
**
==============
O_threshold = 140
MC for this TARGET:[60.214, 0.078]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.26, -7.31, -7.35]][[-12.63, -60.21, -9.87]]
std:[[1.02, 1.04, 0.44]][[0.33, 0.0, 0.3]]
MSE:[[7.33, 7.38, 7.36]][[12.63, 60.21, 9.87]]
MSE(-DR):[[0.0, 0.05, 0.03]][[5.3, 52.88, 2.54]]
***
==============
[[ 0.49  0.64  1.64  0.84 62.56 12.22]
 [ 1.33  1.51  2.32  2.69 61.99 11.64]
 [ 3.14  3.31  4.29  5.18 61.52 11.17]
 [ 3.01  3.15  3.85  5.86 59.98  9.64]
```

```
 [ 7.32  7.38  7.31 12.62 60.21  9.87]
 [ 7.33  7.38  7.36 12.63 60.21  9.87]]


time spent until now: 17.0 mins
```

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

max(u_O) =  191.7
O_threshold = 95
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 0 1 0 0

number of reward locations:  11
O_threshold = 100
number of reward locations:  9
O_threshold = 110
number of reward locations:  7
O_threshold = 120
number of reward locations:  6
O_threshold = 125
number of reward locations:  3
O_threshold = 140
number of reward locations:  3
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!
target 1 in 1 DONE!


----------------------------------------
Value of Behaviour policy:50.315
O_threshold = 95
MC for this TARGET:[62.596, 0.239]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.14, -0.41, -1.5]][[-0.92, -62.6, -12.28]]
std:[[0.98, 0.96, 0.6]][[0.66, 0.0, 0.28]]
MSE:[[0.99, 1.04, 1.62]][[1.13, 62.6, 12.28]]
MSE(-DR):[[0.0, 0.05, 0.63]][[0.14, 61.61, 11.29]]
***
==============
O_threshold = 100
MC for this TARGET:[62.018, 0.241]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.95, -1.16, -2.21]][[-2.75, -62.02, -11.7]]
std:[[0.98, 0.98, 0.47]][[0.62, 0.0, 0.28]]
MSE:[[1.36, 1.52, 2.26]][[2.82, 62.02, 11.7]]
MSE(-DR):[[0.0, 0.16, 0.9]][[1.46, 60.66, 10.34]]
***
==============
O_threshold = 110
MC for this TARGET:[61.549, 0.239]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.11, -3.36, -4.13]][[-5.18, -61.55, -11.23]]
std:[[1.14, 1.18, 0.55]][[0.52, 0.0, 0.28]]
MSE:[[3.31, 3.56, 4.17]][[5.21, 61.55, 11.23]]
MSE(-DR):[[0.0, 0.25, 0.86]][[1.9, 58.24, 7.92]]
***
==============
O_threshold = 120
MC for this TARGET:[60.013, 0.24]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.01, -3.2, -3.8]][[-5.94, -60.01, -9.7]]
std:[[1.12, 1.16, 0.66]][[0.48, 0.0, 0.28]]
```

MSE:[[3.21, 3.4, 3.86]][[5.96, 60.01, 9.7]]
MSE(-DR):[[0.0, 0.19, 0.65]][[2.75, 56.8, 6.49]]
***
==============
O_threshold = 125
MC for this TARGET:[60.246, 0.241]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.76, -7.81, -7.51]][[-12.72, -60.25, -9.93]]
std:[[1.58, 1.66, 0.66]][[0.5, 0.0, 0.28]]
MSE:[[7.92, 7.98, 7.54]][[12.73, 60.25, 9.93]]
MSE(-DR):[[0.0, 0.06, -0.38]][[4.81, 52.33, 2.01]]
**
==============
O_threshold = 140
MC for this TARGET:[60.246, 0.241]
    [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.77, -7.81, -7.54]][[-12.72, -60.25, -9.93]]
std:[[1.53, 1.66, 0.71]][[0.52, 0.0, 0.28]]
MSE:[[7.92, 7.98, 7.57]][[12.73, 60.25, 9.93]]
MSE(-DR):[[0.0, 0.06, -0.35]][[4.81, 52.33, 2.01]]
**
==============
[[0.49 0.64 1.64 0.84 62.56 12.22]
 [1.33 1.51 2.32 2.69 61.99 11.64]
 [3.14 3.31 4.29 5.18 61.52 11.17]
 [3.01 3.15 3.85 5.86 59.98 9.64]
 [7.32 7.38 7.31 12.62 60.21 9.87]
 [7.33 7.38 7.36 12.63 60.21 9.87]]

[[0.99 1.04 1.62 1.13 62.6 12.28]
 [1.36 1.52 2.26 2.82 62.02 11.7]
 [3.31 3.56 4.17 5.21 61.55 11.23]
 [3.21 3.4   3.86 5.96 60.01 9.7]
 [7.92 7.98 7.54 12.73 60.25 9.93]
 [7.92 7.98 7.57 12.73 60.25 9.93]]

time spent until now: 34.2 mins

14:14, 04/13

[pattern_seed, day, sd_R] = [2, 7, 40]

max(u_0) = 191.7
O_threshold = 95
means of Order:

83.3 97.7 14.6 165.6 28.3

66.3 120.1 50.2 57.7 63.6

122.1 191.7 101.7 55.3 121.6

76.2 99.2 147.0 70.1 100.4

64.9 93.7 110.3 60.4 86.4

target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 0 1 0 0

number of reward locations: 11
O_threshold = 100
number of reward locations: 9
O_threshold = 110
number of reward locations: 7
O_threshold = 120
number of reward locations: 6
O_threshold = 125
number of reward locations: 3
O_threshold = 140
number of reward locations: 3