```
Last login: Sun Apr 12 16:06:04 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe.pem" ubuntu@ec2-54-237-166-41.compute-1.amazonaws.com
Warning: Permanently added the ED25519 host key for IP address '54.237.166.41' 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 Sun Apr 12 21:33:47 UTC 2020

  System load:  0.79               Processes:           223
  Usage of /:   28.0% of 30.96GB   Users logged in:     0
  Memory usage: 1%                 IP address for ens5: 172.31.9.175
  Swap usage:   0%

 * Kubernetes 1.18 GA is now available! See https://microk8s.io for docs or
   install it with:

     sudo snap install microk8s --channel=1.18 --classic

 * Multipass 1.1 adds proxy support for developers behind enterprise
   firewalls. Rapid prototyping for cloud operations just got easier.

     https://multipass.run/

 * Canonical Livepatch is available for installation.
   - Reduce system reboots and improve kernel security. Activate at:
     https://ubuntu.com/livepatch

89 packages can be updated.
39 updates are security updates.


Last login: Fri Apr  3 19:45:17 2020 from 107.13.161.147
export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
ubuntu@ip-172-31-9-175:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
17:35, 04/12; num of cores:16
median_u_O_u_D

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, 10, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[100, 105, 110, 115], [0, 15, 30], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]]

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
False True [5e-05, 5e-05] 320.3050051101101
False True [0.0001, 0.0001] 508.56655105962375
False True [0.0001, 5e-05] 99.55072569973943
False True [0.0001, 0.0001] 509.43386479571916
False True [0.0001, 0.0001] 541.2532230017397
False True [0.0001, 0.0001] 1210.0546694132918
False True [5e-05, 5e-05] 301.67714562350653
False True [0.0001, 5e-05] 615.0733941911267
False True [5e-05, 5e-05] 841.4094652913697
False True [0.0001, 0.0001] 649.2964429776496
False True [0.0001, 0.0001] 754.3219590823256
False True [0.0001, 0.0001] 1119.7830629855157
False True [0.0001, 0.0001] 777.4401302376071
False True [0.0001, 0.0001] 719.4635172625684
False True [0.0001, 5e-05] 274.5955651369494
False True [5e-05, 0.0001] 1077.7434669958834
^CProcess Process-14:
Process Process-10:
Traceback (most recent call last):
  File "EC2.py", line 92, in <module>
Process Process-15:
Process Process-2:
Process Process-16:
Process Process-8:
Process Process-3:
    with_MF = with_MF, with_NO_MARL = with_NO_MARL, with_IS = with_IS)
Process Process-5:
  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
Process Process-7:
    [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
Process Process-6:
Process Process-1:
Process Process-13:
Process Process-12:
Process Process-9:
Process Process-4:
Traceback (most recent call last):
    if not self._sem.acquire(block, timeout):
  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)
KeyboardInterrupt
  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)])
  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 306, in computeQV
    validation_set = valid_tuples)
  File "/home/ubuntu/main.py", line 491, in computeQV_basic
    QSA1 = alpha.T.dot(SA_GRBF(Z = Z_tilde, gamma = gamma_q, Z2 = SA_t1)).T
  File "/home/ubuntu/main.py", line 358, in SA_GRBF
    K = GRBF(Z[:,:(l - 2)], Z2[:,:(l - 2)], gamma) + nonsingular
KeyboardInterrupt
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 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)])
  File "/home/ubuntu/main.py", line 158, in <listcomp>
    r = arr([getOneRegionValue(i) for i in range(N)])
  File "/home/ubuntu/main.py", line 87, in getOneRegionValue
    epsilon = epsilon)
  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
Traceback (most recent call last):
ubuntu@ip-172-31-9-175:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
17:40, 04/12; num of cores:16
median_u_O_u_D

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, 10, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[100, 105, 110, 115], [0, 15, 30], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]]

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
```

```
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:52.865
O_threshold = 100
MC for this TARGET:[58.154, 0.081]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.24, -0.33, -1.1]][[-1.31, -58.15, -5.29]]
std:[[0.6, 0.59, 0.41]][[0.32, 0.0, 0.23]]
MSE:[[0.65, 0.68, 1.17]][[1.35, 58.15, 5.29]]
MSE(-DR):[[0.0, 0.03, 0.52]][[0.7, 57.5, 4.64]]
***
==============
O_threshold = 105
MC for this TARGET:[57.708, 0.073]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.18, -2.24, -3.07]][[-4.14, -57.71, -4.84]]
std:[[0.63, 0.61, 0.4]][[0.32, 0.0, 0.23]]
MSE:[[2.27, 2.32, 3.1]][[4.15, 57.71, 4.85]]
MSE(-DR):[[0.0, 0.05, 0.83]][[1.88, 55.44, 2.58]]
***
==============
O_threshold = 110
MC for this TARGET:[56.697, 0.063]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.12, -2.15, -2.93]][[-5.27, -56.7, -3.83]]
std:[[0.52, 0.5, 0.41]][[0.33, 0.0, 0.23]]
MSE:[[2.18, 2.21, 2.96]][[5.28, 56.7, 3.84]]
MSE(-DR):[[0.0, 0.03, 0.78]][[3.1, 54.52, 1.66]]
***
==============
O_threshold = 115
MC for this TARGET:[58.647, 0.054]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.17, -7.14, -7.54]][[-13.2, -58.65, -5.78]]
std:[[0.74, 0.74, 0.5]][[0.32, 0.0, 0.23]]
MSE:[[7.21, 7.18, 7.56]][[13.2, 58.65, 5.78]]
MSE(-DR):[[0.0, -0.03, 0.35]][[5.99, 51.44, -1.43]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 0.65  0.68  1.17  1.35 58.15  5.29]
 [ 2.27  2.32  3.1   4.15 57.71  4.85]
 [ 2.18  2.21  2.96  5.28 56.7   3.84]
 [ 7.21  7.18  7.56 13.2  58.65  5.78]]


time spent until now: 39.0 mins

18:19, 04/12

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

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:52.843
O_threshold = 100
MC for this TARGET:[58.178, 0.188]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.22, -0.31, -1.2]][[-1.37, -58.18, -5.33]]
std:[[0.74, 0.76, 0.56]][[0.31, 0.0, 0.21]]
MSE:[[0.77, 0.82, 1.32]][[1.4, 58.18, 5.33]]
MSE(-DR):[[0.0, 0.05, 0.55]][[0.63, 57.41, 4.56]]
***
==============
O_threshold = 105
MC for this TARGET:[57.732, 0.182]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.94, -2.04, -3.19]][[-4.24, -57.73, -4.89]]
std:[[0.92, 0.91, 0.57]][[0.32, 0.0, 0.21]]
```

```
MSE:[[2.15, 2.23, 3.24]][[4.25, 57.73, 4.89]]
MSE(-DR):[[0.0, 0.08, 1.09]][[2.1, 55.58, 2.74]]
***
==============
O_threshold = 110
MC for this TARGET:[56.721, 0.18]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.02, -2.08, -3.0]][[-5.33, -56.72, -3.88]]
std:[[0.73, 0.71, 0.67]][[0.37, 0.0, 0.21]]
MSE:[[2.15, 2.2, 3.07]][[5.34, 56.72, 3.89]]
MSE(-DR):[[0.0, 0.05, 0.92]][[3.19, 54.57, 1.74]]
***
==============
O_threshold = 115
MC for this TARGET:[58.671, 0.183]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.24, -7.23, -7.57]][[-13.25, -58.67, -5.83]]
std:[[1.0, 0.98, 0.7]][[0.38, 0.0, 0.21]]
MSE:[[7.31, 7.3, 7.6]][[13.26, 58.67, 5.83]]
MSE(-DR):[[0.0, -0.01, 0.29]][[5.95, 51.36, -1.48]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 0.65  0.68  1.17  1.35 58.15  5.29]
 [ 2.27  2.32  3.1   4.15 57.71  4.85]
 [ 2.18  2.21  2.96  5.28 56.7   3.84]
 [ 7.21  7.18  7.56 13.2  58.65  5.78]]


[[ 0.77  0.82  1.32  1.4  58.18  5.33]
 [ 2.15  2.23  3.24  4.25 57.73  4.89]
 [ 2.15  2.2   3.07  5.34 56.72  3.89]
 [ 7.31  7.3   7.6  13.26 58.67  5.83]]


time spent until now: 78.0 mins

18:58, 04/12

----------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 30]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

----------------------------------------
Value of Behaviour policy:52.822
O_threshold = 100
MC for this TARGET:[58.202, 0.352]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.19, -0.31, -1.3]][[-1.43, -58.2, -5.38]]
std:[[1.33, 1.36, 0.89]][[0.54, 0.0, 0.3]]
MSE:[[1.34, 1.39, 1.58]][[1.53, 58.2, 5.39]]
MSE(-DR):[[0.0, 0.05, 0.24]][[0.19, 56.86, 4.05]]
***
==============
O_threshold = 105
MC for this TARGET:[57.755, 0.348]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.77, -1.88, -3.31]][[-4.36, -57.76, -4.93]]
std:[[1.65, 1.68, 0.86]][[0.55, 0.0, 0.3]]
MSE:[[2.42, 2.52, 3.42]][[4.39, 57.76, 4.94]]
MSE(-DR):[[0.0, 0.1, 1.0]][[1.97, 55.34, 2.52]]
***
==============
O_threshold = 110
MC for this TARGET:[56.744, 0.347]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.95, -2.03, -3.05]][[-5.39, -56.74, -3.92]]
std:[[1.32, 1.28, 0.99]][[0.63, 0.0, 0.3]]
MSE:[[2.35, 2.4, 3.21]][[5.43, 56.74, 3.93]]
MSE(-DR):[[0.0, 0.05, 0.86]][[3.08, 54.39, 1.58]]
***
==============
O_threshold = 115
MC for this TARGET:[58.695, 0.353]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.27, -7.26, -7.62]][[-13.3, -58.7, -5.87]]
```

```
std:[[1.72, 1.7, 1.3]][[0.67, 0.0, 0.3]]
MSE:[[7.47, 7.46, 7.73]][[13.32, 58.7, 5.88]]
MSE(-DR):[[0.0, -0.01, 0.26]][[5.85, 51.23, -1.59]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 0.65  0.68  1.17  1.35 58.15  5.29]
 [ 2.27  2.32  3.1   4.15 57.71  4.85]
 [ 2.18  2.21  2.96  5.28 56.7   3.84]
 [ 7.21  7.18  7.56 13.2  58.65  5.78]]


[[ 0.77  0.82  1.32  1.4  58.18  5.33]
 [ 2.15  2.23  3.24  4.25 57.73  4.89]
 [ 2.15  2.2   3.07  5.34 56.72  3.89]
 [ 7.31  7.3   7.6  13.26 58.67  5.83]]


[[ 1.34  1.39  1.58  1.53 58.2   5.39]
 [ 2.42  2.52  3.42  4.39 57.76  4.94]
 [ 2.35  2.4   3.21  5.43 56.74  3.93]
 [ 7.47  7.46  7.73 13.32 58.7   5.88]]


time spent until now: 117.2 mins

19:37, 04/12
ubuntu@ip-172-31-9-175:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
19:47, 04/12; num of cores:16
median_u_O_u_D_other_C

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, 10, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[90, 95, 120], [0, 15, 30], [7], [[0.0001, 5e-05], [0.0001, 5e-05]]]

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

max(u_O) =  145.8
O_threshold = 90
number of reward locations:  14
O_threshold = 95
number of reward locations:  12
O_threshold = 120
number of reward locations:  3
target 1 in 3 DONE!
target 2 in 3 DONE!
target 3 in 3 DONE!

---------------------------------------
Value of Behaviour policy:52.865
O_threshold = 90
MC for this TARGET:[62.904, 0.077]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.61, 1.33, -0.48]][[1.91, -62.9, -10.04]]
std:[[0.3, 0.3, 0.36]][[0.3, 0.0, 0.23]]
MSE:[[1.64, 1.36, 0.6]][[1.93, 62.9, 10.04]]
MSE(-DR):[[0.0, -0.28, -1.04]][[0.29, 61.26, 8.4]]
**
==============
O_threshold = 95
MC for this TARGET:[59.142, 0.082]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.59, 2.4, 1.12]][[2.48, -59.14, -6.28]]
std:[[0.48, 0.47, 0.38]][[0.27, 0.0, 0.23]]
MSE:[[2.63, 2.45, 1.18]][[2.49, 59.14, 6.28]]
MSE(-DR):[[0.0, -0.18, -1.45]][[-0.14, 56.51, 3.65]]
==============
O_threshold = 120
MC for this TARGET:[58.647, 0.054]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.15, -7.14, -7.54]][[-13.2, -58.65, -5.78]]
std:[[0.74, 0.74, 0.48]][[0.32, 0.0, 0.23]]
MSE:[[7.19, 7.18, 7.56]][[13.2, 58.65, 5.78]]
MSE(-DR):[[0.0, -0.01, 0.37]][[6.01, 51.46, -1.41]]
***
==============
[[ 1.64  1.36  0.6   1.93 62.9  10.04]
 [ 2.63  2.45  1.18  2.49 59.14  6.28]
 [ 7.19  7.18  7.56 13.2  58.65  5.78]]


time spent until now: 29.2 mins

20:16, 04/12

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

```
max(u_O) =  145.8
O_threshold = 90
number of reward locations:  14
O_threshold = 95
number of reward locations:  12
O_threshold = 120
number of reward locations:  3
target 1 in 3 DONE!
target 2 in 3 DONE!
target 3 in 3 DONE!

----------------------------------------
Value of Behaviour policy:52.843
O_threshold = 90
MC for this TARGET:[62.928, 0.188]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.85, 1.6, -0.3]][[2.0, -62.93, -10.08]]
std:[[0.5, 0.5, 0.45]][[0.42, 0.0, 0.21]]
MSE:[[1.92, 1.68, 0.54]][[2.04, 62.93, 10.08]]
MSE(-DR):[[0.0, -0.24, -1.38]][[0.12, 61.01, 8.16]]
**
==============
O_threshold = 95
MC for this TARGET:[59.165, 0.186]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.77, 2.59, 1.24]][[2.54, -59.16, -6.32]]
std:[[0.51, 0.5, 0.44]][[0.35, 0.0, 0.21]]
MSE:[[2.82, 2.64, 1.32]][[2.56, 59.16, 6.32]]
MSE(-DR):[[0.0, -0.18, -1.5]][[-0.26, 56.34, 3.5]]
==============
O_threshold = 120
MC for this TARGET:[58.671, 0.183]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.22, -7.23, -7.56]][[-13.25, -58.67, -5.83]]
std:[[1.01, 0.98, 0.69]][[0.38, 0.0, 0.21]]
MSE:[[7.29, 7.3, 7.59]][[13.26, 58.67, 5.83]]
MSE(-DR):[[0.0, 0.01, 0.3]][[5.97, 51.38, -1.46]]
***
==============
[[ 1.64  1.36  0.6   1.93 62.9  10.04]
 [ 2.63  2.45  1.18  2.49 59.14  6.28]
 [ 7.19  7.18  7.56 13.2  58.65  5.78]]


[[ 1.92  1.68  0.54  2.04 62.93 10.08]
 [ 2.82  2.64  1.32  2.56 59.16  6.32]
 [ 7.29  7.3   7.59 13.26 58.67  5.83]]


time spent until now: 58.5 mins

20:45, 04/12

----------------------------------------
[pattern_seed, day, sd_R] = [2, 7, 30]

max(u_O) =  145.8
O_threshold = 90
number of reward locations:  14
O_threshold = 95
number of reward locations:  12
O_threshold = 120
number of reward locations:  3
target 1 in 3 DONE!
target 2 in 3 DONE!
target 3 in 3 DONE!

----------------------------------------
Value of Behaviour policy:52.822
O_threshold = 90
MC for this TARGET:[62.951, 0.354]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.15, 1.88, -0.1]][[2.08, -62.95, -10.13]]
std:[[0.86, 0.9, 0.69]][[0.71, 0.0, 0.3]]
MSE:[[2.32, 2.08, 0.7]][[2.2, 62.95, 10.13]]
MSE(-DR):[[0.0, -0.24, -1.62]][[-0.12, 60.63, 7.81]]
==============
O_threshold = 95
MC for this TARGET:[59.189, 0.349]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[2.99, 2.8, 1.38]][[2.6, -59.19, -6.37]]
std:[[0.99, 0.91, 0.73]][[0.61, 0.0, 0.3]]
MSE:[[3.15, 2.94, 1.56]][[2.67, 59.19, 6.38]]
MSE(-DR):[[0.0, -0.21, -1.59]][[-0.48, 56.04, 3.23]]
==============
O_threshold = 120
MC for this TARGET:[58.695, 0.353]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.25, -7.26, -7.62]][[-13.3, -58.7, -5.87]]
```

```
std:[[1.74, 1.7, 1.27]][[0.68, 0.0, 0.3]]
MSE:[[7.46, 7.46, 7.73]][[13.32, 58.7, 5.88]]
MSE(-DR):[[0.0, 0.0, 0.27]][[5.86, 51.24, -1.58]]
***
==============
[[ 1.64  1.36  0.6   1.93 62.9  10.04]
 [ 2.63  2.45  1.18  2.49 59.14  6.28]
 [ 7.19  7.18  7.56 13.2  58.65  5.78]]


[[ 1.92  1.68  0.54  2.04 62.93 10.08]
 [ 2.82  2.64  1.32  2.56 59.16  6.32]
 [ 7.29  7.3   7.59 13.26 58.67  5.83]]


[[ 2.32  2.08  0.7   2.2  62.95 10.13]
 [ 3.15  2.94  1.56  2.67 59.19  6.38]
 [ 7.46  7.46  7.73 13.32 58.7   5.88]]


time spent until now: 87.8 mins

21:14, 04/12
ubuntu@ip-172-31-9-175:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
21:59, 04/12; num of cores:16
median_u_O_u_D_with_T_45

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, 10, None]

[thre_range, sd_R_range, day_range, penalty_range]:  [[100, 105, 110, 115], [15], [4, 5], [[0.0001, 5e-05], [0.0001, 5e-05]]]

----------------------------------------
[pattern_seed, day, sd_R] = [2, 4, 15]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

----------------------------------------
Value of Behaviour policy:52.766
O_threshold = 100
MC for this TARGET:[58.163, 0.249]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.74, -0.84, -1.5]][[-1.38, -58.16, -5.4]]
std:[[1.24, 1.26, 0.83]][[0.67, 0.0, 0.4]]
MSE:[[1.44, 1.51, 1.71]][[1.53, 58.16, 5.41]]
MSE(-DR):[[0.0, 0.07, 0.27]][[0.09, 56.72, 3.97]]
***
==============
O_threshold = 105
MC for this TARGET:[57.717, 0.24]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.03, -3.08, -3.48]][[-4.23, -57.72, -4.95]]
std:[[1.21, 1.22, 1.05]][[0.61, 0.0, 0.4]]
MSE:[[3.26, 3.31, 3.63]][[4.27, 57.72, 4.97]]
MSE(-DR):[[0.0, 0.05, 0.37]][[1.01, 54.46, 1.71]]
***
==============
O_threshold = 110
MC for this TARGET:[56.709, 0.241]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.9, -2.95, -3.32]][[-5.34, -56.71, -3.94]]
std:[[1.31, 1.31, 1.06]][[0.57, 0.0, 0.4]]
MSE:[[3.18, 3.23, 3.49]][[5.37, 56.71, 3.96]]
MSE(-DR):[[0.0, 0.05, 0.31]][[2.19, 53.53, 0.78]]
***
==============
O_threshold = 115
MC for this TARGET:[58.656, 0.251]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.38, -7.34, -7.8]][[-13.32, -58.66, -5.89]]
std:[[1.59, 1.58, 0.96]][[0.64, 0.0, 0.4]]
MSE:[[7.55, 7.51, 7.86]][[13.34, 58.66, 5.9]]
MSE(-DR):[[0.0, -0.04, 0.31]][[5.79, 51.11, -1.65]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.44  1.51  1.71  1.53 58.16  5.41]
 [ 3.26  3.31  3.63  4.27 57.72  4.97]
```

```
 [ 3.18  3.23  3.49  5.37 56.71  3.96]
 [ 7.55  7.51  7.86 13.34 58.66  5.9 ]]


time spent until now: 34.3 mins

22:34, 04/12

---------------------------------------
[pattern_seed, day, sd_R] = [2, 5, 15]

max(u_O) =  145.8
O_threshold = 100
number of reward locations:  9
O_threshold = 105
number of reward locations:  7
O_threshold = 110
number of reward locations:  6
O_threshold = 115
number of reward locations:  3
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!

---------------------------------------
Value of Behaviour policy:52.847
O_threshold = 100
MC for this TARGET:[58.167, 0.206]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.52, -0.62, -1.43]][[-1.48, -58.17, -5.32]]
std:[[0.86, 0.84, 0.44]][[0.49, 0.0, 0.24]]
MSE:[[1.0, 1.04, 1.5]][[1.56, 58.17, 5.33]]
MSE(-DR):[[0.0, 0.04, 0.5]][[0.56, 57.17, 4.33]]
***
==============
O_threshold = 105
MC for this TARGET:[57.719, 0.202]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.29, -2.39, -3.26]][[-4.33, -57.72, -4.87]]
std:[[0.95, 0.93, 0.57]][[0.51, 0.0, 0.24]]
MSE:[[2.48, 2.56, 3.31]][[4.36, 57.72, 4.88]]
MSE(-DR):[[0.0, 0.08, 0.83]][[1.88, 55.24, 2.4]]
***
==============
O_threshold = 110
MC for this TARGET:[56.708, 0.206]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.12, -2.15, -3.11]][[-5.43, -56.71, -3.86]]
std:[[0.91, 0.88, 0.65]][[0.54, 0.0, 0.24]]
MSE:[[2.31, 2.32, 3.18]][[5.46, 56.71, 3.87]]
MSE(-DR):[[0.0, 0.01, 0.87]][[3.15, 54.4, 1.56]]
***
==============
O_threshold = 115
MC for this TARGET:[58.645, 0.211]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.9, -6.86, -7.73]][[-13.29, -58.64, -5.8]]
std:[[1.17, 1.21, 0.69]][[0.4, 0.0, 0.24]]
MSE:[[7.0, 6.97, 7.76]][[13.3, 58.64, 5.8]]
MSE(-DR):[[0.0, -0.03, 0.76]][[6.3, 51.64, -1.2]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.44  1.51  1.71  1.53 58.16  5.41]
 [ 3.26  3.31  3.63  4.27 57.72  4.97]
 [ 3.18  3.23  3.49  5.37 56.71  3.96]
 [ 7.55  7.51  7.86 13.34 58.66  5.9 ]]


[[ 1.    1.04  1.5   1.56 58.17  5.33]
 [ 2.48  2.56  3.31  4.36 57.72  4.88]
 [ 2.31  2.32  3.18  5.46 56.71  3.87]
 [ 7.    6.97  7.76 13.3  58.64  5.8 ]]


time spent until now: 69.6 mins

23:09, 04/12
ubuntu@ip-172-31-9-175:~$
```