```
Last login: Mon Apr  6 12:19:29 on ttys000
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-3-226-243-168.compute-1.amazonaws.com
ssh: connect to host ec2-3-226-243-168.compute-1.amazonaws.com port 22: Connection refused
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-3-226-243-168.compute-1.amazonaws.com
The authenticity of host 'ec2-3-226-243-168.compute-1.amazonaws.com (3.226.243.168)' can't be established.
ECDSA key fingerprint is SHA256:p4BMNyZAtG1UYhuF8Y0ZIMrQ2AWCCUGbJKWJfUzwBU0.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-226-243-168.compute-1.amazonaws.com,3.226.243.168' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1063-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Mon Apr  6 18:02:47 UTC 2020

  System load:  1.04               Processes:           379
  Usage of /:   57.0% of 15.45GB   Users logged in:     0
  Memory usage: 0%                 IP address for ens5: 172.31.73.77
  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

53 packages can be updated.
0 updates are security updates.


Last login: Wed Apr  1 20:30:39 2020 from 107.13.161.147
ubuntu@ip-172-31-73-77:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
14:04, 04/06; num of cores:36

Traceback (most recent call last):
  File "EC2.py", line 52, in <module>
    printR("final sd_R trend for", sd_R_range, "\n")
TypeError: printR() takes 1 positional argument but 3 were given
ubuntu@ip-172-31-73-77:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
14:05, 04/06; num of cores:36

final sd_R trend for[10, 15]

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, u_O_u_D], sd_R_range, t_func] = [None,
16, None, None, None, 20, 0.5, 1, [True, False, True, 10], [10, 15], None]


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

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!
Rep 5 DONE with time cost 1.0 mins; Rep 10 DONE with time cost 1.0 mins; 1Rep 15 DONE with time cost 1.0 mins;
----------------------------------------
Value of Behaviour policy:61.025
O_threshold = 100
MC for this TARGET:[67.489, 0.142]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.48, -1.56, -2.05]][[-2.79, -67.49, -6.46]]
std:[[0.56, 0.6, 0.4]][[0.27, 0.0, 0.14]]
MSE:[[1.58, 1.67, 2.09]][[2.8, 67.49, 6.46]]
MSE(-DR):[[0.0, 0.09, 0.51]][[1.22, 65.91, 4.88]]
***
==============
O_threshold = 105
MC for this TARGET:[67.925, 0.139]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.69, -3.77, -4.46]][[-5.67, -67.92, -6.9]]
```

```
std:[[0.68, 0.7, 0.39]][[0.27, 0.0, 0.14]]
MSE:[[3.75, 3.83, 4.48]][[5.68, 67.92, 6.9]]
MSE(-DR):[[0.0, 0.08, 0.73]][[1.93, 64.17, 3.15]]
***
==============
O_threshold = 110
MC for this TARGET:[66.862, 0.137]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.64, -3.7, -4.19]][[-6.26, -66.86, -5.84]]
std:[[0.57, 0.58, 0.45]][[0.28, 0.0, 0.14]]
MSE:[[3.68, 3.75, 4.21]][[6.27, 66.86, 5.84]]
MSE(-DR):[[0.0, 0.07, 0.53]][[2.59, 63.18, 2.16]]
***
==============
O_threshold = 115
MC for this TARGET:[67.538, 0.134]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.7, -6.71, -6.76]][[-11.36, -67.54, -6.51]]
std:[[0.57, 0.57, 0.45]][[0.28, 0.0, 0.14]]
MSE:[[6.72, 6.73, 6.77]][[11.36, 67.54, 6.51]]
MSE(-DR):[[0.0, 0.01, 0.05]][[4.64, 60.82, -0.21]]
***
==============
******************** THIS SETTING IS GOOD ********************
[[ 1.58  1.67  2.09  2.8  67.49  6.46]
 [ 3.75  3.83  4.48  5.68 67.92  6.9 ]
 [ 3.68  3.75  4.21  6.27 66.86  5.84]
 [ 6.72  6.73  6.77 11.36 67.54  6.51]]


time spent until now: 22.8 mins


---------------------------------------
[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!
Rep 5 DONE with time cost 1.0 mins; Rep 10 DONE with time cost 1.0 mins; Rep 15 DONE with time cost 1.0 mins;
---------------------------------------
Value of Behaviour policy:61.018
O_threshold = 100
MC for this TARGET:[67.497, 0.193]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.48, -1.55, -2.12]][[-2.81, -67.5, -6.48]]
std:[[0.71, 0.74, 0.47]][[0.31, 0.0, 0.15]]
MSE:[[1.64, 1.72, 2.17]][[2.83, 67.5, 6.48]]
MSE(-DR):[[0.0, 0.08, 0.53]][[1.19, 65.86, 4.84]]
***
==============
O_threshold = 105
MC for this TARGET:[67.933, 0.19]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.59, -3.69, -4.47]][[-5.7, -67.93, -6.92]]
std:[[0.93, 0.97, 0.48]][[0.3, 0.0, 0.15]]
MSE:[[3.71, 3.82, 4.5]][[5.71, 67.93, 6.92]]
MSE(-DR):[[0.0, 0.11, 0.79]][[2.0, 64.22, 3.21]]
***
==============
O_threshold = 110
MC for this TARGET:[66.87, 0.189]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.61, -3.66, -4.25]][[-6.28, -66.87, -5.85]]
std:[[0.79, 0.79, 0.54]][[0.32, 0.0, 0.15]]
MSE:[[3.7, 3.74, 4.28]][[6.29, 66.87, 5.85]]
MSE(-DR):[[0.0, 0.04, 0.58]][[2.59, 63.17, 2.15]]
***
==============
O_threshold = 115
MC for this TARGET:[67.546, 0.188]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.7, -6.71, -6.75]][[-11.38, -67.55, -6.53]]
std:[[0.79, 0.77, 0.62]][[0.34, 0.0, 0.15]]
MSE:[[6.75, 6.75, 6.78]][[11.39, 67.55, 6.53]]
MSE(-DR):[[0.0, 0.0, 0.03]][[4.64, 60.8, -0.22]]
***
==============
******************** THIS SETTING IS GOOD ********************
```

```
[[ 1.58  1.67  2.09  2.8  67.49  6.46]
 [ 3.75  3.83  4.48  5.68 67.92  6.9 ]
 [ 3.68  3.75  4.21  6.27 66.86  5.84]
 [ 6.72  6.73  6.77 11.36 67.54  6.51]]


[[ 1.64  1.72  2.17  2.83 67.5   6.48]
 [ 3.71  3.82  4.5   5.71 67.93  6.92]
 [ 3.7   3.74  4.28  6.29 66.87  5.85]
 [ 6.75  6.75  6.78 11.39 67.55  6.53]]


time spent until now: 45.3 mins

ubuntu@ip-172-31-73-77:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
14:57, 04/06; num of cores:36

final sd_R trend for[10, 15]

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, u_O_u_D], sd_R_range, t_func] = [None,
16, None, None, None, 20, 0.5, 1, [True, False, True, 10], [10, 15], None]


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

max(u_O) =  145.4
O_threshold = 100
number of reward locations:  18
O_threshold = 105
number of reward locations:  16
O_threshold = 110
number of reward locations:  11
O_threshold = 115
number of reward locations:  10
target 1 in 4 DONE!
^CProcess Process-16:
Process Process-15:
Process Process-14:
Process Process-1:
Traceback (most recent call last):
Process Process-9:
  File "EC2.py", line 82, in <module>
Process Process-4:
    with_MF = with_MF,
  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)]
  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-10:
Process Process-7:
Process Process-12:
Process Process-2:
Process Process-8:
Process Process-11:
Exception ignored in: <function ScopedTFGraph.__del__ at 0x7fd2405d1bf8>
Traceback (most recent call last):
  File "/home/ubuntu/.local/lib/python3.7/site-packages/tensorflow/python/framework/c_api_util.py", line 48, in __del__
    def __del__(self):
KeyboardInterrupt
Process Process-13:
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
ubuntu@ip-172-31-73-77:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
15:09, 04/06; num of cores:36

final sd_R trend for[10, 15]

Basic setting:[T, rep_times, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, u_O_u_D], sd_R_range, t_func] = [None,
16, None, None, None, 20, 0.5, 1, [True, False, True, 10], [10, 15], None]


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

max(u_O) =  145.4
O_threshold = 100
number of reward locations:  18
O_threshold = 105
number of reward locations:  16
O_threshold = 110
number of reward locations:  11
O_threshold = 115
number of reward locations:  10
target 1 in 4 DONE!
target 2 in 4 DONE!
```

```
target 3 in 4 DONE!
target 4 in 4 DONE!
Rep 5 DONE with time cost 1.0 mins; Rep 10 DONE with time cost 1.0 mins; Rep 15 DONE with time cost 1.0 mins;
----------------------------------------
Value of Behaviour policy:69.85
O_threshold = 100
MC for this TARGET:[81.358, 0.148]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.31, 0.13, -1.67]][[2.3, -81.36, -11.51]]
std:[[0.68, 0.69, 0.39]][[0.39, 0.0, 0.18]]
MSE:[[0.75, 0.7, 1.71]][[2.33, 81.36, 11.51]]
MSE(-DR):[[0.0, -0.05, 0.96]][[1.58, 80.61, 10.76]]
***
==============
O_threshold = 105
MC for this TARGET:[81.335, 0.145]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.74, -0.96, -3.1]][[0.53, -81.33, -11.49]]
std:[[0.58, 0.58, 0.43]][[0.43, 0.0, 0.18]]
MSE:[[0.94, 1.12, 3.13]][[0.68, 81.33, 11.49]]
MSE(-DR):[[0.0, 0.18, 2.19]][[-0.26, 80.39, 10.55]]
==============
O_threshold = 110
MC for this TARGET:[76.846, 0.143]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.15, -1.24, -1.62]][[-1.69, -76.85, -7.0]]
std:[[0.57, 0.58, 0.38]][[0.36, 0.0, 0.18]]
MSE:[[1.28, 1.37, 1.66]][[1.73, 76.85, 7.0]]
MSE(-DR):[[0.0, 0.09, 0.38]][[0.45, 75.57, 5.72]]
***
==============
O_threshold = 115
MC for this TARGET:[76.005, 0.143]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.04, -2.09, -2.21]][[-2.57, -76.0, -6.16]]
std:[[0.68, 0.68, 0.41]][[0.4, 0.0, 0.18]]
MSE:[[2.15, 2.2, 2.25]][[2.6, 76.0, 6.16]]
MSE(-DR):[[0.0, 0.05, 0.1]][[0.45, 73.85, 4.01]]
***
==============
[[ 0.75  0.7   1.71  2.33 81.36 11.51]
 [ 0.94  1.12  3.13  0.68 81.33 11.49]
 [ 1.28  1.37  1.66  1.73 76.85  7.  ]
 [ 2.15  2.2   2.25  2.6  76.    6.16]]


time spent until now: 22.6 mins


----------------------------------------
[pattern_seed, day, sd_R] = [0, 7, 15]

max(u_O) =  145.4
O_threshold = 100
number of reward locations:  18
O_threshold = 105
number of reward locations:  16
O_threshold = 110
number of reward locations:  11
O_threshold = 115
number of reward locations:  10
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!
Rep 5 DONE with time cost 1.0 mins; Rep 10 DONE with time cost 1.0 mins; Rep 15 DONE with time cost 1.0 mins;
----------------------------------------
Value of Behaviour policy:69.843
O_threshold = 100
MC for this TARGET:[81.366, 0.198]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.37, 0.19, -1.67]][[2.32, -81.37, -11.52]]
std:[[0.89, 0.88, 0.5]][[0.49, 0.0, 0.21]]
MSE:[[0.96, 0.9, 1.74]][[2.37, 81.37, 11.52]]
MSE(-DR):[[0.0, -0.06, 0.78]][[1.41, 80.41, 10.56]]
***
==============
O_threshold = 105
MC for this TARGET:[81.343, 0.197]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.72, -0.96, -3.12]][[0.55, -81.34, -11.5]]
std:[[0.79, 0.79, 0.54]][[0.52, 0.0, 0.21]]
MSE:[[1.07, 1.24, 3.17]][[0.76, 81.34, 11.5]]
MSE(-DR):[[0.0, 0.17, 2.1]][[-0.31, 80.27, 10.43]]
==============
O_threshold = 110
MC for this TARGET:[76.854, 0.195]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.25, -1.34, -1.67]][[-1.74, -76.85, -7.01]]
```

```
std:[[0.75, 0.76, 0.42]][[0.42, 0.0, 0.21]]
MSE:[[1.46, 1.54, 1.72]][[1.79, 76.85, 7.01]]
MSE(-DR):[[0.0, 0.08, 0.26]][[0.33, 75.39, 5.55]]
***
==============
O_threshold = 115
MC for this TARGET:[76.013, 0.196]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.12, -2.15, -2.28]][[-2.61, -76.01, -6.17]]
std:[[0.9, 0.91, 0.45]][[0.46, 0.0, 0.21]]
MSE:[[2.3, 2.33, 2.32]][[2.65, 76.01, 6.17]]
MSE(-DR):[[0.0, 0.03, 0.02]][[0.35, 73.71, 3.87]]
***
==============
[[ 0.75  0.7   1.71  2.33 81.36 11.51]
 [ 0.94  1.12  3.13  0.68 81.33 11.49]
 [ 1.28  1.37  1.66  1.73 76.85  7.  ]
 [ 2.15  2.2   2.25  2.6  76.    6.16]]


[[ 0.96  0.9   1.74  2.37 81.37 11.52]
 [ 1.07  1.24  3.17  0.76 81.34 11.5 ]
 [ 1.46  1.54  1.72  1.79 76.85  7.01]
 [ 2.3   2.33  2.32  2.65 76.01  6.17]]


time spent until now: 45.2 mins

ubuntu@ip-172-31-73-77:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
16:03, 04/06; num of cores:36

final sd_R trend for[10, 15]

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, 20, 0.5, 1, [True, False, True, 10], [10, 15], None]


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

max(u_O) =  145.4
O_threshold = 90
number of reward locations:  21
O_threshold = 120
number of reward locations:  6
target 1 in 2 DONE!
target 2 in 2 DONE!
Rep 5 DONE with time cost 0.0 mins; Rep 10 DONE with time cost 0.0 mins; Rep 15 DONE with time cost 0.0 mins;
----------------------------------------
Value of Behaviour policy:69.85
O_threshold = 90
MC for this TARGET:[78.197, 0.151]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.79, 1.66, 1.29]][[5.24, -78.2, -8.35]]
std:[[1.13, 1.14, 0.61]][[0.42, 0.0, 0.18]]
MSE:[[2.12, 2.01, 1.43]][[5.26, 78.2, 8.35]]
MSE(-DR):[[0.0, -0.11, -0.69]][[3.14, 76.08, 6.23]]
**
==============
O_threshold = 120
MC for this TARGET:[81.239, 0.132]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-10.24, -10.22, -9.5]][[-13.18, -81.24, -11.39]]
std:[[1.01, 1.04, 0.48]][[0.38, 0.0, 0.18]]
MSE:[[10.29, 10.27, 9.51]][[13.19, 81.24, 11.39]]
MSE(-DR):[[0.0, -0.02, -0.78]][[2.9, 70.95, 1.1]]
**
==============
[[ 2.12  2.01  1.43  5.26 78.2   8.35]
 [10.29 10.27  9.51 13.19 81.24 11.39]]


time spent until now: 11.3 mins


----------------------------------------
[pattern_seed, day, sd_R] = [0, 7, 15]

max(u_O) =  145.4
O_threshold = 90
number of reward locations:  21
O_threshold = 120
number of reward locations:  6
target 1 in 2 DONE!
target 2 in 2 DONE!
Rep 5 DONE with time cost 0.0 mins; Rep 10 DONE with time cost 0.0 mins; Rep 15 DONE with time cost 0.0 mins;
----------------------------------------
Value of Behaviour policy:69.843
O_threshold = 90
```

```
MC for this TARGET:[78.205, 0.201]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[1.78, 1.66, 1.26]][[5.25, -78.2, -8.36]]
std:[[1.41, 1.42, 0.82]][[0.53, 0.0, 0.21]]
MSE:[[2.27, 2.18, 1.5]][[5.28, 78.2, 8.36]]
MSE(-DR):[[0.0, -0.09, -0.77]][[3.01, 75.93, 6.09]]
**
==============
O_threshold = 120
MC for this TARGET:[81.247, 0.184]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-10.38, -10.37, -9.57]][[-13.2, -81.25, -11.4]]
std:[[1.25, 1.26, 0.68]][[0.45, 0.0, 0.21]]
MSE:[[10.45, 10.45, 9.59]][[13.21, 81.25, 11.4]]
MSE(-DR):[[0.0, 0.0, -0.86]][[2.76, 70.8, 0.95]]
**
==============
[[ 2.12  2.01  1.43  5.26 78.2   8.35]
 [10.29 10.27  9.51 13.19 81.24 11.39]]


[[ 2.27  2.18  1.5   5.28 78.2   8.36]
 [10.45 10.45  9.59 13.21 81.25 11.4 ]]


time spent until now: 22.6 mins

ubuntu@ip-172-31-73-77:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
16:31, 04/06; num of cores:36

final sd_R trend for[10, 15]

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, 20, 0.5, 1, [True, False, True, 10], [10, 15], None]


----------------------------------------
[pattern_seed, day, sd_R] = [0, 14, 10]

max(u_O) =  145.4
O_threshold = 90
number of reward locations:  21
O_threshold = 100
number of reward locations:  18
O_threshold = 110
number of reward locations:  11
O_threshold = 120
number of reward locations:  6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!
Rep 5 DONE with time cost 3.0 mins; Rep 10 DONE with time cost 3.0 mins; Rep 15 DONE with time cost 3.0 mins;
----------------------------------------
Value of Behaviour policy:69.847
O_threshold = 90
MC for this TARGET:[78.176, 0.103]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[3.0, 2.77, 1.45]][[5.11, -78.18, -8.33]]
std:[[0.73, 0.72, 0.29]][[0.25, 0.0, 0.11]]
MSE:[[3.09, 2.86, 1.48]][[5.12, 78.18, 8.33]]
MSE(-DR):[[0.0, -0.23, -1.61]][[2.03, 75.09, 5.24]]
**
==============
O_threshold = 100
MC for this TARGET:[81.347, 0.105]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.8, 0.56, -1.67]][[2.12, -81.35, -11.5]]
std:[[0.51, 0.51, 0.28]][[0.24, 0.0, 0.11]]
MSE:[[0.95, 0.76, 1.69]][[2.13, 81.35, 11.5]]
MSE(-DR):[[0.0, -0.19, 0.74]][[1.18, 80.4, 10.55]]
***
==============
O_threshold = 110
MC for this TARGET:[76.83, 0.094]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.92, -1.03, -1.5]][[-1.71, -76.83, -6.98]]
std:[[0.28, 0.28, 0.17]][[0.3, 0.0, 0.11]]
MSE:[[0.96, 1.07, 1.51]][[1.74, 76.83, 6.98]]
MSE(-DR):[[0.0, 0.11, 0.55]][[0.78, 75.87, 6.02]]
***
==============
O_threshold = 120
MC for this TARGET:[81.215, 0.091]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-10.63, -10.64, -9.43]][[-13.17, -81.22, -11.37]]
std:[[0.59, 0.59, 0.38]][[0.26, 0.0, 0.11]]
MSE:[[10.65, 10.66, 9.44]][[13.17, 81.22, 11.37]]
MSE(-DR):[[0.0, 0.01, -1.21]][[2.52, 70.57, 0.72]]
```

**
==========
[[3.09 2.86 1.48 5.12 78.18 8.33]
 [0.95 0.76 1.69 2.13 81.35 11.5]
 [0.96 1.07 1.51 1.74 76.83 6.98]
 [10.65 10.66 9.44 13.17 81.22 11.37]]

time spent until now: 52.5 mins

[pattern_seed, day, sd_R] = [0, 14, 15]

max(u_O) = 145.4
O_threshold = 90
number of reward locations: 21
O_threshold = 100
number of reward locations: 18
O_threshold = 110
number of reward locations: 11
O_threshold = 120
number of reward locations: 6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!
Rep 5 DONE with time cost 3.0 mins;