```
Last login: Mon Apr  6 18:56:02 on ttys001
Run-Mac:~ mac$ cd ~/.ssh
Run-Mac:.ssh mac$ ssh -i "Runzhe_Song_0110.pem" ubuntu@ec2-34-231-20-47.compute-1.amazonaws.com
The authenticity of host 'ec2-34-231-20-47.compute-1.amazonaws.com (34.231.20.47)' can't be established.
ECDSA key fingerprint is SHA256:n7bRJuo4XU64lzI0mkplQDtBWz4L/URgJAKL+9Dwd0w.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-34-231-20-47.compute-1.amazonaws.com,34.231.20.47' (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 22:59:53 UTC 2020

  System load:  1.37               Processes:           229
  Usage of /:   57.0% of 15.45GB   Users logged in:     0
  Memory usage: 1%                 IP address for ens5: 172.31.79.228
  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
export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.pyubuntu@ip-172-31-79-228:~$ export openblas_num_threads=1; export O
MP_NUM_THREADS=1; python EC2.py
19:01, 04/06; num of cores:16

final sd_R trend for[0, 10, 20]

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, 30, 1, 1, [True, False, True, 10], [0, 10, 20], None]


---------------------------------------
[pattern_seed, day, sd_R] = [1, 7, 0]

max(u_O) =  152.3
O_threshold = 80
number of reward locations:  18
O_threshold = 90
number of reward locations:  15
O_threshold = 100
number of reward locations:  12
O_threshold = 110
number of reward locations:  10
O_threshold = 120
number of reward locations:  8
O_threshold = 130
number of reward locations:  5
target 1 in 6 DONE!
target 2 in 6 DONE!
target 3 in 6 DONE!
target 4 in 6 DONE!
target 5 in 6 DONE!
target 6 in 6 DONE!

---------------------------------------
Value of Behaviour policy:62.735
O_threshold = 80
MC for this TARGET:[71.776, 0.065]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-1.27, -1.44, -2.06]][[0.96, -71.78, -9.04]]
std:[[0.38, 0.38, 0.31]][[0.24, 0.0, 0.2]]
MSE:[[1.33, 1.49, 2.08]][[0.99, 71.78, 9.04]]
MSE(-DR):[[0.0, 0.16, 0.75]][[-0.34, 70.45, 7.71]]
==============
O_threshold = 90
MC for this TARGET:[72.101, 0.063]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.83, -3.0, -3.73]][[-0.91, -72.1, -9.37]]
std:[[0.31, 0.3, 0.31]][[0.28, 0.0, 0.2]]
MSE:[[2.85, 3.01, 3.74]][[0.95, 72.1, 9.37]]
MSE(-DR):[[0.0, 0.16, 0.89]][[-1.9, 69.25, 6.52]]
```

```
==============
O_threshold = 100
MC for this TARGET:[74.355, 0.062]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.14, -5.28, -5.34]][[-4.48, -74.36, -11.62]]
std:[[0.52, 0.54, 0.4]][[0.26, 0.0, 0.2]]
MSE:[[5.17, 5.31, 5.35]][[4.49, 74.36, 11.62]]
MSE(-DR):[[0.0, 0.14, 0.18]][[-0.68, 69.19, 6.45]]
==============
O_threshold = 110
MC for this TARGET:[77.337, 0.07]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-7.68, -7.81, -7.41]][[-8.22, -77.34, -14.6]]
std:[[0.61, 0.61, 0.43]][[0.28, 0.0, 0.2]]
MSE:[[7.7, 7.83, 7.42]][[8.22, 77.34, 14.6]]
MSE(-DR):[[0.0, 0.13, -0.28]][[0.52, 69.64, 6.9]]
**
==============
O_threshold = 120
MC for this TARGET:[76.883, 0.065]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.59, -9.66, -8.59]][[-10.42, -76.88, -14.15]]
std:[[0.67, 0.67, 0.41]][[0.25, 0.0, 0.2]]
MSE:[[9.61, 9.68, 8.6]][[10.42, 76.88, 14.15]]
MSE(-DR):[[0.0, 0.07, -1.01]][[0.81, 67.27, 4.54]]
**
==============
O_threshold = 130
MC for this TARGET:[74.581, 0.072]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.42, -9.44, -8.94]][[-11.18, -74.58, -11.85]]
std:[[0.54, 0.54, 0.45]][[0.26, 0.0, 0.2]]
MSE:[[9.44, 9.46, 8.95]][[11.18, 74.58, 11.85]]
MSE(-DR):[[0.0, 0.02, -0.49]][[1.74, 65.14, 2.41]]
**
==============
[[ 1.33  1.49  2.08  0.99 71.78  9.04]
 [ 2.85  3.01  3.74  0.95 72.1   9.37]
 [ 5.17  5.31  5.35  4.49 74.36 11.62]
 [ 7.7   7.83  7.42  8.22 77.34 14.6 ]
 [ 9.61  9.68  8.6  10.42 76.88 14.15]
 [ 9.44  9.46  8.95 11.18 74.58 11.85]]


time spent until now: 59.1 mins


---------------------------------------
[pattern_seed, day, sd_R] = [1, 7, 10]

max(u_O) =  152.3
O_threshold = 80
number of reward locations:  18
O_threshold = 90
number of reward locations:  15
O_threshold = 100
number of reward locations:  12
O_threshold = 110
number of reward locations:  10
O_threshold = 120
number of reward locations:  8
O_threshold = 130
number of reward locations:  5
^[[A^CProcess Process-21:
Process Process-30:
Process Process-22:
Traceback (most recent call last):
  File "EC2.py", line 87, in <module>
Process Process-17:
Process Process-32:
Process Process-18:
Process Process-28:
    with_MF = with_MF,
  File "/home/ubuntu/simu_funs.py", line 62, in simu
Process Process-29:
    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-31:
Process Process-19:
Process Process-23:
Process Process-25:
Process Process-27:
Process Process-26:
    if not self._sem.acquire(block, timeout):
KeyboardInterrupt
```

```
ubuntu@ip-172-31-79-228:~$ export openblas_num_threads=1; export OMP_NUM_THREADS=1; python EC2.py
20:02, 04/06; num of cores:16

final sd_R trend for[0, 10, 20]

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] = [1, 7, 0]

max(u_O) =  152.3
O_threshold = 80
number of reward locations:  18
O_threshold = 90
number of reward locations:  15
O_threshold = 100
number of reward locations:  12
O_threshold = 110
number of reward locations:  10
O_threshold = 120
number of reward locations:  8
O_threshold = 130
number of reward locations:  5
target 1 in 6 DONE!
target 2 in 6 DONE!
target 3 in 6 DONE!
target 4 in 6 DONE!
target 5 in 6 DONE!
target 6 in 6 DONE!

--------------------------------------
Value of Behaviour policy:59.399
O_threshold = 80
MC for this TARGET:[69.796, 0.068]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.68, -0.92, -1.87]][[1.95, -69.8, -10.4]]
std:[[0.42, 0.42, 0.35]][[0.26, 0.0, 0.24]]
MSE:[[0.8, 1.01, 1.9]][[1.97, 69.8, 10.4]]
MSE(-DR):[[0.0, 0.21, 1.1]][[1.17, 69.0, 9.6]]
***
==============
O_threshold = 90
MC for this TARGET:[69.945, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.52, -2.73, -3.71]][[-0.17, -69.94, -10.55]]
std:[[0.35, 0.34, 0.32]][[0.3, 0.0, 0.24]]
MSE:[[2.54, 2.75, 3.72]][[0.34, 69.94, 10.55]]
MSE(-DR):[[0.0, 0.21, 1.18]][[-2.2, 67.4, 8.01]]
==============
O_threshold = 100
MC for this TARGET:[72.012, 0.065]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.56, -4.74, -4.92]][[-3.94, -72.01, -12.61]]
std:[[0.64, 0.66, 0.43]][[0.32, 0.0, 0.24]]
MSE:[[4.6, 4.79, 4.94]][[3.95, 72.01, 12.61]]
MSE(-DR):[[0.0, 0.19, 0.34]][[-0.65, 67.41, 8.01]]
==============
O_threshold = 110
MC for this TARGET:[75.367, 0.069]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-6.83, -7.01, -6.8]][[-8.4, -75.37, -15.97]]
std:[[0.62, 0.63, 0.48]][[0.36, 0.0, 0.24]]
MSE:[[6.86, 7.04, 6.82]][[8.41, 75.37, 15.97]]
MSE(-DR):[[0.0, 0.18, -0.04]][[1.55, 68.51, 9.11]]
**
==============
O_threshold = 120
MC for this TARGET:[75.745, 0.06]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-10.29, -10.38, -9.38]][[-12.29, -75.74, -16.35]]
std:[[0.72, 0.74, 0.42]][[0.33, 0.0, 0.24]]
MSE:[[10.32, 10.41, 9.39]][[12.29, 75.74, 16.35]]
MSE(-DR):[[0.0, 0.09, -0.93]][[1.97, 65.42, 6.03]]
**
==============
O_threshold = 130
MC for this TARGET:[72.691, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-9.63, -9.68, -9.37]][[-13.37, -72.69, -13.29]]
std:[[0.53, 0.53, 0.41]][[0.27, 0.0, 0.24]]
MSE:[[9.64, 9.69, 9.38]][[13.37, 72.69, 13.29]]
MSE(-DR):[[0.0, 0.05, -0.26]][[3.73, 63.05, 3.65]]
**
==============
[[ 0.8   1.01  1.9   1.97 69.8  10.4 ]
 [ 2.54  2.75  3.72  0.34 69.94 10.55]
 [ 4.6   4.79  4.94  3.95 72.01 12.61]
```

```
 [ 6.86  7.04  6.82  8.41 75.37 15.97]
 [10.32 10.41  9.39 12.29 75.74 16.35]
 [ 9.64  9.69  9.38 13.37 72.69 13.29]]


time spent until now: 58.9 mins


--------------------------------------
[pattern_seed, day, sd_R] = [1, 7, 10]

max(u_O) =  152.3
O_threshold = 80
number of reward locations:  18
O_threshold = 90
number of reward locations:  15
O_threshold = 100
number of reward locations:  12
O_threshold = 110
number of reward locations:  10
O_threshold = 120
number of reward locations:  8
O_threshold = 130
number of reward locations:  5
^CTraceback (most recent call last):
Process Process-28:
  File "EC2.py", line 87, in <module>
Process Process-18:
Process Process-27:
Process Process-22:
Process Process-30:
Process Process-32:
Process Process-23:
    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
Process Process-21:
    [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
    if not self._sem.acquire(block, timeout):
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 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)
  File "/home/ubuntu/main.py", line 294, in computeQV
    validation_set = valid_tuples)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 297, in _bootstrap
    self.run()
  File "/home/ubuntu/main.py", line 405, in computeQV_basic
    KQ = SA_GRBF(Z_tilde, gamma_q)
  File "/home/ubuntu/anaconda3/lib/python3.7/multiprocessing/process.py", line 99, in run
    self._target(*self._args, **self._kwargs)
Process Process-24:
  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)
  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]
KeyboardInterrupt
  File "/home/ubuntu/main.py", line 344, in SA_GRBF
```

```
    return np.multiply(K, I_A)
ubuntu@ip-172-31-79-228:~$ 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[0, 10, 20] 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, poisO, u_O_u_D], sd_R_range, t_func] = [None,
16, None, None, None, None, 30, 0.5, 1, [True, False, True, 10], [0, 10, 20], None]


----------------------------------------
[pattern_seed, day, sd_R] = [3, 7, 0]

max(u_O) =  153.7
O_threshold = 80
number of reward locations:  19
O_threshold = 90
number of reward locations:  14
O_threshold = 100
number of reward locations:  10
O_threshold = 110
number of reward locations:  7
O_threshold = 120
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:61.172
O_threshold = 80
MC for this TARGET:[71.03, 0.069]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.51, 0.37, -0.61]][[3.29, -71.03, -9.86]]
std:[[0.55, 0.58, 0.27]][[0.16, 0.0, 0.12]]
MSE:[[0.75, 0.69, 0.67]][[3.29, 71.03, 9.86]]
MSE(-DR):[[0.0, -0.06, -0.08]][[2.54, 70.28, 9.11]]
**
==============
O_threshold = 90
MC for this TARGET:[70.92, 0.064]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.52, -0.69, -2.35]][[0.19, -70.92, -9.75]]
std:[[0.4, 0.43, 0.2]][[0.24, 0.0, 0.12]]
MSE:[[0.66, 0.81, 2.36]][[0.31, 70.92, 9.75]]
MSE(-DR):[[0.0, 0.15, 1.7]][[-0.35, 70.26, 9.09]]
==============
O_threshold = 100
MC for this TARGET:[71.404, 0.067]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.38, -2.56, -3.67]][[-3.58, -71.4, -10.23]]
std:[[0.3, 0.32, 0.19]][[0.25, 0.0, 0.12]]
MSE:[[2.4, 2.58, 3.67]][[3.59, 71.4, 10.23]]
MSE(-DR):[[0.0, 0.18, 1.27]][[1.19, 69.0, 7.83]]
***
==============
O_threshold = 110
MC for this TARGET:[67.789, 0.067]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.67, -2.74, -2.85]][[-4.8, -67.79, -6.62]]
std:[[0.5, 0.53, 0.24]][[0.22, 0.0, 0.12]]
MSE:[[2.72, 2.79, 2.86]][[4.81, 67.79, 6.62]]
MSE(-DR):[[0.0, 0.07, 0.14]][[2.09, 65.07, 3.9]]
***
==============
O_threshold = 120
MC for this TARGET:[65.862, 0.071]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.06, -2.12, -2.07]][[-4.59, -65.86, -4.69]]
std:[[0.45, 0.47, 0.21]][[0.22, 0.0, 0.12]]
MSE:[[2.11, 2.17, 2.08]][[4.6, 65.86, 4.69]]
MSE(-DR):[[0.0, 0.06, -0.03]][[2.49, 63.75, 2.58]]
**
==============
[[ 0.75  0.69  0.67  3.29 71.03  9.86]
 [ 0.66  0.81  2.36  0.31 70.92  9.75]
 [ 2.4   2.58  3.67  3.59 71.4  10.23]
 [ 2.72  2.79  2.86  4.81 67.79  6.62]
 [ 2.11  2.17  2.08  4.6  65.86  4.69]]


time spent until now: 49.1 mins


----------------------------------------
[pattern_seed, day, sd_R] = [3, 7, 10]
```

```
max(u_0) =  153.7
O_threshold = 80
number of reward locations:  19
O_threshold = 90
number of reward locations:  14
O_threshold = 100
number of reward locations:  10
O_threshold = 110
number of reward locations:  7
O_threshold = 120
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:61.157
O_threshold = 80
MC for this TARGET:[71.046, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.62, 0.47, -0.57]][[3.32, -71.05, -9.89]]
std:[[0.75, 0.79, 0.45]][[0.37, 0.0, 0.15]]
MSE:[[0.97, 0.92, 0.73]][[3.34, 71.05, 9.89]]
MSE(-DR):[[0.0, -0.05, -0.24]][[2.37, 70.08, 8.92]]
**
==============
O_threshold = 90
MC for this TARGET:[70.935, 0.13]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.27, -0.46, -2.25]][[0.22, -70.94, -9.78]]
std:[[0.6, 0.6, 0.32]][[0.39, 0.0, 0.15]]
MSE:[[0.66, 0.76, 2.27]][[0.45, 70.94, 9.78]]
MSE(-DR):[[0.0, 0.1, 1.61]][[-0.21, 70.28, 9.12]]
==============
O_threshold = 100
MC for this TARGET:[71.42, 0.136]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.59, -2.76, -3.79]][[-3.56, -71.42, -10.26]]
std:[[0.42, 0.45, 0.35]][[0.37, 0.0, 0.15]]
MSE:[[2.62, 2.8, 3.81]][[3.58, 71.42, 10.26]]
MSE(-DR):[[0.0, 0.18, 1.19]][[0.96, 68.8, 7.64]]
***
==============
O_threshold = 110
MC for this TARGET:[67.804, 0.135]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.64, -2.73, -2.84]][[-4.79, -67.8, -6.65]]
std:[[0.58, 0.62, 0.32]][[0.32, 0.0, 0.15]]
MSE:[[2.7, 2.8, 2.86]][[4.8, 67.8, 6.65]]
MSE(-DR):[[0.0, 0.1, 0.16]][[2.1, 65.1, 3.95]]
***
==============
O_threshold = 120
MC for this TARGET:[65.878, 0.139]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.14, -2.21, -2.04]][[-4.56, -65.88, -4.72]]
std:[[0.63, 0.67, 0.35]][[0.33, 0.0, 0.15]]
MSE:[[2.23, 2.31, 2.07]][[4.57, 65.88, 4.72]]
MSE(-DR):[[0.0, 0.08, -0.16]][[2.34, 63.65, 2.49]]
**
==============
[[ 0.75  0.69  0.67  3.29 71.03  9.86]
 [ 0.66  0.81  2.36  0.31 70.92  9.75]
 [ 2.4   2.58  3.67  3.59 71.4  10.23]
 [ 2.72  2.79  2.86  4.81 67.79  6.62]
 [ 2.11  2.17  2.08  4.6  65.86  4.69]]


[[ 0.97  0.92  0.73  3.34 71.05  9.89]
 [ 0.66  0.76  2.27  0.45 70.94  9.78]
 [ 2.62  2.8   3.81  3.58 71.42 10.26]
 [ 2.7   2.8   2.86  4.8  67.8   6.65]
 [ 2.23  2.31  2.07  4.57 65.88  4.72]]


time spent until now: 98.1 mins


---------------------------------------
[pattern_seed, day, sd_R] = [3, 7, 20]

max(u_0) =  153.7
O_threshold = 80
number of reward locations:  19
O_threshold = 90
number of reward locations:  14
O_threshold = 100
```