```
Basic setting:[T, rep_times, sd_O, sd_D, sd_u_O, w_O, w_A, [M_in_R, mean_reversion, poisO, u_O_u_D], sd_R_range, t_func
] = [None, 32, None, None, 30, 0.5, 1, [True, False, True, 10], [0, 10, 20], None]


--------------------------------------
[pattern_seed, day, sd_R] = [2, 3, 0]

max(u_O) =  168.8
O_threshold = 80
number of reward locations:  15
O_threshold = 90
number of reward locations:  12
O_threshold = 100
number of reward locations:  9
O_threshold = 110
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!

--------------------------------------
Value of Behaviour policy:57.752
O_threshold = 80
MC for this TARGET:[68.365, 0.101]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.76, -0.9, -1.2]][[1.02, -68.36, -10.61]]
std:[[0.75, 0.77, 0.4]][[0.33, 0.0, 0.28]]
MSE:[[1.07, 1.18, 1.26]][[1.07, 68.36, 10.61]]
MSE(-DR):[[0.0, 0.11, 0.19]][[0.0, 67.29, 9.54]]
***
==============
O_threshold = 90
MC for this TARGET:[66.725, 0.104]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.38, -0.52, -0.82]][[-0.57, -66.72, -8.97]]
std:[[0.73, 0.73, 0.39]][[0.36, 0.0, 0.28]]
MSE:[[0.82, 0.9, 0.91]][[0.67, 66.72, 8.97]]
MSE(-DR):[[0.0, 0.08, 0.09]][[-0.15, 65.9, 8.15]]
==============
O_threshold = 100
MC for this TARGET:[66.939, 0.105]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.03, -3.13, -3.2]][[-4.28, -66.94, -9.19]]
std:[[0.63, 0.64, 0.34]][[0.39, 0.0, 0.28]]
MSE:[[3.09, 3.19, 3.22]][[4.3, 66.94, 9.19]]
MSE(-DR):[[0.0, 0.1, 0.13]][[1.21, 63.85, 6.1]]
***
==============
O_threshold = 110
MC for this TARGET:[65.955, 0.122]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.95, -5.01, -5.33]][[-7.43, -65.96, -8.2]]
std:[[0.7, 0.72, 0.43]][[0.36, 0.0, 0.28]]
MSE:[[5.0, 5.06, 5.35]][[7.44, 65.96, 8.2]]
MSE(-DR):[[0.0, 0.06, 0.35]][[2.44, 60.96, 3.2]]
***
==============
```

```
--------------------------------------
Value of Behaviour policy:57.74
O_threshold = 80
MC for this TARGET:[68.384, 0.198]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.83, -0.96, -1.27]][[0.96, -68.38, -10.64]]
std:[[0.84, 0.85, 0.7]][[0.49, 0.0, 0.34]]
MSE:[[1.18, 1.28, 1.45]][[1.08, 68.38, 10.65]]
MSE(-DR):[[0.0, 0.1, 0.27]][[-0.1, 67.2, 9.47]]
==============
O_threshold = 90
MC for this TARGET:[66.744, 0.204]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.43, -0.56, -0.91]][[-0.63, -66.74, -9.0]]
std:[[0.89, 0.88, 0.56]][[0.55, 0.0, 0.34]]
MSE:[[0.99, 1.04, 1.07]][[0.84, 66.74, 9.01]]
MSE(-DR):[[0.0, 0.05, 0.08]][[-0.15, 65.75, 8.02]]
==============
O_threshold = 100
MC for this TARGET:[66.958, 0.196]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.01, -3.11, -3.31]][[-4.41, -66.96, -9.22]]
std:[[0.87, 0.88, 0.62]][[0.6, 0.0, 0.34]]
MSE:[[3.13, 3.23, 3.37]][[4.45, 66.96, 9.23]]
MSE(-DR):[[0.0, 0.1, 0.24]][[1.32, 63.83, 6.1]]
***
==============
O_threshold = 110
MC for this TARGET:[65.974, 0.202]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-4.98, -5.04, -5.5]][[-7.51, -65.97, -8.23]]
std:[[0.98, 0.99, 0.71]][[0.55, 0.0, 0.34]]
MSE:[[5.08, 5.14, 5.55]][[7.53, 65.97, 8.24]]
MSE(-DR):[[0.0, 0.06, 0.47]][[2.45, 60.89, 3.16]]
***
==============
```

```
-------------------------------------
Value of Behaviour policy:57.727
O_threshold = 80
MC for this TARGET:[68.403, 0.361]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.92, -1.06, -1.33]][[0.9, -68.4, -10.68]]
std:[[1.32, 1.31, 1.14]][[0.74, 0.0, 0.47]]
MSE:[[1.61, 1.69, 1.75]][[1.17, 68.4, 10.69]]
MSE(-DR):[[0.0, 0.08, 0.14]][[-0.44, 66.79, 9.08]]
==============
O_threshold = 90
MC for this TARGET:[66.763, 0.367]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.49, -0.64, -0.94]][[-0.71, -66.76, -9.04]]
std:[[1.46, 1.43, 0.94]][[0.82, 0.0, 0.47]]
MSE:[[1.54, 1.57, 1.33]][[1.08, 66.76, 9.05]]
MSE(-DR):[[0.0, 0.03, -0.21]][[-0.46, 65.22, 7.51]]
==============
O_threshold = 100
MC for this TARGET:[66.977, 0.358]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.98, -3.09, -3.38]][[-4.52, -66.98, -9.25]]
std:[[1.32, 1.33, 0.97]][[0.91, 0.0, 0.47]]
MSE:[[3.26, 3.36, 3.52]][[4.61, 66.98, 9.26]]
MSE(-DR):[[0.0, 0.1, 0.26]][[1.35, 63.72, 6.0]]
***
==============
O_threshold = 110
MC for this TARGET:[65.993, 0.36]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.05, -5.12, -5.7]][[-7.6, -65.99, -8.27]]
std:[[1.51, 1.5, 1.16]][[0.82, 0.0, 0.47]]
MSE:[[5.27, 5.34, 5.82]][[7.64, 65.99, 8.28]]
MSE(-DR):[[0.0, 0.07, 0.55]][[2.37, 60.72, 3.01]]
***
==============
```

```
Value of Behaviour policy:57.712
O_threshold = 80
MC for this TARGET:[68.355, 0.062]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.13, -0.35, -1.03]][[1.13, -68.36, -10.64]]
std:[[0.47, 0.49, 0.23]][[0.26, 0.0, 0.21]]
MSE:[[0.49, 0.6, 1.06]][[1.16, 68.36, 10.64]]
MSE(-DR):[[0.0, 0.11, 0.57]][[0.67, 67.87, 10.15]]
***
==============
O_threshold = 90
MC for this TARGET:[66.714, 0.067]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.03, -0.22, -0.75]][[-0.48, -66.71, -9.0]]
std:[[0.36, 0.37, 0.24]][[0.26, 0.0, 0.21]]
MSE:[[0.36, 0.43, 0.79]][[0.55, 66.71, 9.0]]
MSE(-DR):[[0.0, 0.07, 0.43]][[0.19, 66.35, 8.64]]
***
==============
O_threshold = 100
MC for this TARGET:[66.948, 0.071]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.77, -2.91, -3.17]][[-4.24, -66.95, -9.24]]
std:[[0.39, 0.4, 0.25]][[0.25, 0.0, 0.21]]
MSE:[[2.8, 2.94, 3.18]][[4.25, 66.95, 9.24]]
MSE(-DR):[[0.0, 0.14, 0.38]][[1.45, 64.15, 6.44]]
***
==============
O_threshold = 110
MC for this TARGET:[65.968, 0.07]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.06, -5.14, -5.36]][[-7.43, -65.97, -8.26]]
std:[[0.4, 0.42, 0.38]][[0.28, 0.0, 0.21]]
MSE:[[5.08, 5.16, 5.37]][[7.44, 65.97, 8.26]]
MSE(-DR):[[0.0, 0.08, 0.29]][[2.36, 60.89, 3.18]]
***
==============
********** THIS SETTING IS GOOD **********
```

```
--------------------------------------
Value of Behaviour policy:57.715
O_threshold = 80
MC for this TARGET:[68.371, 0.129]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.09, -0.3, -0.98]][[1.15, -68.37, -10.66]]
std:[[0.84, 0.86, 0.39]][[0.42, 0.0, 0.22]]
MSE:[[0.84, 0.91, 1.05]][[1.22, 68.37, 10.66]]
MSE(-DR):[[0.0, 0.07, 0.21]][[0.38, 67.53, 9.82]]
***
==============
O_threshold = 90
MC for this TARGET:[66.73, 0.139]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.06, -0.24, -0.74]][[-0.49, -66.73, -9.02]]
std:[[0.59, 0.58, 0.41]][[0.36, 0.0, 0.22]]
MSE:[[0.59, 0.63, 0.85]][[0.61, 66.73, 9.02]]
MSE(-DR):[[0.0, 0.04, 0.26]][[0.02, 66.14, 8.43]]
***
==============
O_threshold = 100
MC for this TARGET:[66.964, 0.143]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.77, -2.9, -3.16]][[-4.25, -66.96, -9.25]]
std:[[0.52, 0.54, 0.31]][[0.33, 0.0, 0.22]]
MSE:[[2.82, 2.95, 3.18]][[4.26, 66.96, 9.25]]
MSE(-DR):[[0.0, 0.13, 0.36]][[1.44, 64.14, 6.43]]
***
==============
O_threshold = 110
MC for this TARGET:[65.984, 0.143]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.25, -5.32, -5.41]][[-7.45, -65.98, -8.27]]
std:[[0.6, 0.59, 0.49]][[0.37, 0.0, 0.22]]
MSE:[[5.28, 5.35, 5.43]][[7.46, 65.98, 8.27]]
MSE(-DR):[[0.0, 0.07, 0.15]][[2.18, 60.7, 2.99]]
***
==============
********** THIS SETTING IS GOOD **********
```

```
O_threshold = 80
MC for this TARGET:[68.387, 0.236]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.05, -0.27, -0.94]][[1.17, -68.39, -10.67]]
std:[[1.37, 1.38, 0.73]][[0.68, 0.0, 0.28]]
MSE:[[1.37, 1.41, 1.19]][[1.35, 68.39, 10.67]]
MSE(-DR):[[0.0, 0.04, -0.18]][[-0.02, 67.02, 9.3]]
==============
O_threshold = 90
MC for this TARGET:[66.745, 0.246]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.09, -0.27, -0.75]][[-0.49, -66.74, -9.03]]
std:[[1.04, 1.04, 0.76]][[0.65, 0.0, 0.28]]
MSE:[[1.04, 1.07, 1.07]][[0.81, 66.74, 9.03]]
MSE(-DR):[[0.0, 0.03, 0.03]][[-0.23, 65.7, 7.99]]
==============
O_threshold = 100
MC for this TARGET:[66.98, 0.25]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.77, -2.89, -3.16]][[-4.25, -66.98, -9.26]]
std:[[0.9, 0.9, 0.54]][[0.59, 0.0, 0.28]]
MSE:[[2.91, 3.03, 3.21]][[4.29, 66.98, 9.26]]
MSE(-DR):[[0.0, 0.12, 0.3]][[1.38, 64.07, 6.35]]
***
==============
O_threshold = 110
MC for this TARGET:[65.999, 0.25]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.42, -5.49, -5.45]][[-7.48, -66.0, -8.28]]
std:[[1.18, 1.14, 0.76]][[0.6, 0.0, 0.28]]
MSE:[[5.55, 5.61, 5.5]][[7.5, 66.0, 8.28]]
MSE(-DR):[[0.0, 0.06, -0.05]][[1.95, 60.45, 2.73]]
**
==============
```

```
--------------------------------------
Value of Behaviour policy:57.751
O_threshold = 80
MC for this TARGET:[68.367, 0.051]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.6, 0.35, -0.95]][[1.21, -68.37, -10.62]]
std:[[0.28, 0.29, 0.2]][[0.17, 0.0, 0.12]]
MSE:[[0.66, 0.45, 0.97]][[1.22, 68.37, 10.62]]
MSE(-DR):[[0.0, -0.21, 0.31]][[0.56, 67.71, 9.96]]
***
==============
O_threshold = 90
MC for this TARGET:[66.732, 0.052]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[0.21, 0.02, -0.63]][[-0.42, -66.73, -8.98]]
std:[[0.22, 0.23, 0.18]][[0.17, 0.0, 0.12]]
MSE:[[0.3, 0.23, 0.66]][[0.45, 66.73, 8.98]]
MSE(-DR):[[0.0, -0.07, 0.36]][[0.15, 66.43, 8.68]]
***
==============
O_threshold = 100
MC for this TARGET:[66.954, 0.057]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-2.83, -2.97, -3.05]][[-4.14, -66.95, -9.2]]
std:[[0.25, 0.27, 0.18]][[0.17, 0.0, 0.12]]
MSE:[[2.84, 2.98, 3.06]][[4.14, 66.95, 9.2]]
MSE(-DR):[[0.0, 0.14, 0.22]][[1.3, 64.11, 6.36]]
***
==============
O_threshold = 110
MC for this TARGET:[65.975, 0.058]
   [DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.35, -5.43, -5.34]][[-7.33, -65.97, -8.22]]
std:[[0.3, 0.31, 0.19]][[0.16, 0.0, 0.12]]
MSE:[[5.36, 5.44, 5.34]][[7.33, 65.97, 8.22]]
MSE(-DR):[[0.0, 0.08, -0.02]][[1.97, 60.61, 2.86]]
**
==============
[[ 1.07  1.18  1.26  1.07 68.36 10.61]
```

```
==============
[[ 1.07  1.18  1.26  1.07 68.36 10.61]
 [ 0.82  0.9   0.91  0.67 66.72  8.97]
 [ 3.09  3.19  3.22  4.3  66.94  9.19]
 [ 5.    5.06  5.35  7.44 65.96  8.2 ]]


[[ 1.18  1.28  1.45  1.08 68.38 10.65]
 [ 0.99  1.04  1.07  0.84 66.74  9.01]
 [ 3.13  3.23  3.37  4.45 66.96  9.23]
 [ 5.08  5.14  5.55  7.53 65.97  8.24]]


[[ 1.61  1.69  1.75  1.17 68.4  10.69]
 [ 1.54  1.57  1.33  1.08 66.76  9.05]
 [ 3.26  3.36  3.52  4.61 66.98  9.26]
 [ 5.27  5.34  5.82  7.64 65.99  8.28]]


[[ 0.49  0.6   1.06  1.16 68.36 10.64]
 [ 0.36  0.43  0.79  0.55 66.71  9.  ]
 [ 2.8   2.94  3.18  4.25 66.95  9.24]
 [ 5.08  5.16  5.37  7.44 65.97  8.26]]


[[ 0.84  0.91  1.05  1.22 68.37 10.66]
 [ 0.59  0.63  0.85  0.61 66.73  9.02]
 [ 2.82  2.95  3.18  4.26 66.96  9.25]
 [ 5.28  5.35  5.43  7.46 65.98  8.27]]


[[ 1.37  1.41  1.19  1.35 68.39 10.67]
 [ 1.04  1.07  1.07  0.81 66.74  9.03]
 [ 2.91  3.03  3.21  4.29 66.98  9.26]
 [ 5.55  5.61  5.5   7.5  66.    8.28]]


[[ 0.66  0.45  0.97  1.22 68.37 10.62]
 [ 0.3   0.23  0.66  0.45 66.73  8.98]
 [ 2.84  2.98  3.06  4.14 66.95  9.2 ]
 [ 5.36  5.44  5.34  7.33 65.97  8.22]]
```