```
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_distn_infrastructure.py", line 545, in argsreduce
    return [np.extract(cond, arr1 * expand_arr) for arr1 in newargs]
  File "/home/ubuntu/anaconda3/lib/python3.7/site-packages/scipy/stats/_distn_infrastructure.py", line 545, in <listcomp>
    return [np.extract(cond, arr1 * expand_arr) for arr1 in newargs]
KeyboardInterrupt
ubuntu@ip-172-31-9-80:~$ python EC2.py
12:33, 03/30; num of cores:16

Basic setting:[T, sd_O, sd_D, sd_R, sd_u_O, w_O, w_A, lam, simple, M_in_R, u_D_u_O, mean_reversion] = [672, 5, 10, 10, 0.2, 1, 1, 1e-05,
 False, True, 10, False]


---------------------------------------
[pattern_seed, T, sd_R] = [0, 672, 10]

max(u_O) =  156.6
O_threshold = 100
means of Order:

141.6 107.8 121.0 155.7 144.5

81.8 120.3 96.5 97.5 108.0

102.4 133.1 115.8 101.9 108.7

106.3 134.1 95.5 105.9 83.9

59.7 113.4 118.3 85.8 156.6

target policy:

1 1 1 1 1

0 1 0 0 1

1 1 1 1 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  18
O_threshold = 80
target policy:

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

1 1 1 1 1

0 1 1 1 1

number of reward locations:  24
O_threshold = 85
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 1 1

number of reward locations:  22
O_threshold = 90
target policy:

1 1 1 1 1

0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 95
target policy:

1 1 1 1 1
```

```
0 1 1 1 1

1 1 1 1 1

1 1 1 1 0

0 1 1 0 1

number of reward locations:  21
O_threshold = 105
target policy:

1 1 1 1 1

0 1 0 0 1

0 1 1 0 1

1 1 0 1 0

0 1 1 0 1

number of reward locations:  16
O_threshold = 110
target policy:

1 0 1 1 1

0 1 0 0 0

0 1 1 0 0

0 1 0 0 0

0 1 1 0 1

number of reward locations:  11
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:74.886
O_threshold = 100
MC for this TARGET:[85.629, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.13, -1.33, -2.67]][[0.83, 0.4, 0.1]][[-85.63, -85.63, -85.63]][[-2.87, -10.74]]
std:[[0.18, 0.16, 0.19]][[0.1, 0.05, 0.01]][[0.0, 0.0, 0.0]][[0.21, 0.07]]
MSE:[[1.14, 1.34, 2.68]][[0.84, 0.4, 0.1]][[85.63, 85.63, 85.63]][[2.88, 10.74]]
MSE(-DR):[[0.0, 0.2, 1.54]][[-0.3, -0.74, -1.04]][[84.49, 84.49, 84.49]][[1.74, 9.6]]
==============


O_threshold = 80
MC for this TARGET:[83.925, 0.091]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.62, 1.49, 1.1]][[3.32, 2.94, 2.97]][[-83.92, -83.92, -83.92]][[0.96, -9.04]]
std:[[0.2, 0.24, 0.25]][[0.17, 0.14, 0.01]][[0.0, 0.0, 0.0]][[0.21, 0.07]]
MSE:[[1.63, 1.51, 1.13]][[3.32, 2.94, 2.97]][[83.92, 83.92, 83.92]][[0.98, 9.04]]
MSE(-DR):[[0.0, -0.12, -0.5]][[1.69, 1.31, 1.34]][[82.29, 82.29, 82.29]][[-0.65, 7.41]]
better than DR_NO_MARL
MC-based ATE = -1.7
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.76, 2.82, 3.77]][[2.49, 2.54, 2.86]][[1.7, 1.7, 1.7]][3.83]
std:[[0.01, 0.07, 0.06]][[0.07, 0.09, 0.02]][[0.0, 0.0, 0.0]][0.0]
MSE:[[2.76, 2.82, 3.77]][[2.49, 2.54, 2.86]][[1.7, 1.7, 1.7]][3.83]
MSE(-DR):[[0.0, 0.06, 1.01]][[-0.27, -0.22, 0.1]][[-1.06, -1.06, -1.06]][1.07]
==============


O_threshold = 85
MC for this TARGET:[82.783, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.42, 1.29, 0.55]][[3.52, 3.15, 3.11]][[-82.78, -82.78, -82.78]][[0.42, -7.9]]
std:[[0.42, 0.49, 0.1]][[0.15, 0.12, 0.02]][[0.0, 0.0, 0.0]][[0.04, 0.07]]
MSE:[[1.48, 1.38, 0.56]][[3.52, 3.15, 3.11]][[82.78, 82.78, 82.78]][[0.42, 7.9]]
MSE(-DR):[[0.0, -0.1, -0.92]][[2.04, 1.67, 1.63]][[81.3, 81.3, 81.3]][[-1.06, 6.42]]
better than DR_NO_MARL
MC-based ATE = -2.85
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.55, 2.63, 3.22]][[2.7, 2.75, 3.01]][[2.85, 2.85, 2.85]][3.29]
std:[[0.24, 0.32, 0.09]][[0.05, 0.06, 0.02]][[0.0, 0.0, 0.0]][0.17]
MSE:[[2.56, 2.65, 3.22]][[2.7, 2.75, 3.01]][[2.85, 2.85, 2.85]][3.29]
MSE(-DR):[[0.0, 0.09, 0.66]][[0.14, 0.19, 0.45]][[0.29, 0.29, 0.29]][0.73]
******
==============


O_threshold = 90
MC for this TARGET:[82.087, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.76, 1.58, 0.65]][[3.66, 3.28, 3.24]][[-82.09, -82.09, -82.09]][[0.48, -7.2]]
```

```
std:[[0.09, 0.15, 0.3]][[0.08, 0.06, 0.08]][[0.0, 0.0, 0.0]][[0.24, 0.07]]
MSE:[[1.76, 1.59, 0.72]][[3.66, 3.28, 3.24]][[82.09, 82.09, 82.09]][[0.54, 7.2]]
MSE(-DR):[[0.0, -0.17, -1.04]][[1.9, 1.52, 1.48]][[80.33, 80.33, 80.33]][[-1.22, 5.44]]
better than DR_NO_MARL
MC-based ATE = -3.54
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.9, 2.92, 3.33]][[2.83, 2.88, 3.14]][[3.54, 3.54, 3.54]][3.35]
std:[[0.09, 0.01, 0.11]][[0.02, 0.01, 0.08]][[0.0, 0.0, 0.0]][0.03]
MSE:[[2.9, 2.92, 3.33]][[2.83, 2.88, 3.14]][[3.54, 3.54, 3.54]][3.35]
MSE(-DR):[[0.0, 0.02, 0.43]][[-0.07, -0.02, 0.24]][[0.64, 0.64, 0.64]][0.45]
==============


O_threshold = 95
MC for this TARGET:[82.087, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.74, 1.58, 0.66]][[3.65, 3.28, 3.24]][[-82.09, -82.09, -82.09]][[0.5, -7.2]]
std:[[0.09, 0.15, 0.33]][[0.09, 0.06, 0.07]][[0.0, 0.0, 0.0]][[0.27, 0.07]]
MSE:[[1.74, 1.59, 0.74]][[3.65, 3.28, 3.24]][[82.09, 82.09, 82.09]][[0.57, 7.2]]
MSE(-DR):[[0.0, -0.15, -1.0]][[1.91, 1.54, 1.5]][[80.35, 80.35, 80.35]][[-1.17, 5.46]]
better than DR_NO_MARL
MC-based ATE = -3.54
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.88, 2.92, 3.33]][[2.82, 2.88, 3.13]][[3.54, 3.54, 3.54]][3.37]
std:[[0.09, 0.01, 0.14]][[0.01, 0.01, 0.08]][[0.0, 0.0, 0.0]][0.06]
MSE:[[2.88, 2.92, 3.33]][[2.82, 2.88, 3.13]][[3.54, 3.54, 3.54]][3.37]
MSE(-DR):[[0.0, 0.04, 0.45]][[-0.06, 0.0, 0.25]][[0.66, 0.66, 0.66]][0.49]
==============


O_threshold = 105
MC for this TARGET:[85.861, 0.084]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.94, -2.13, -3.75]][[-0.81, -1.23, -1.56]][[-85.86, -85.86, -85.86]][[-3.94, -10.97]]
std:[[0.19, 0.12, 0.21]][[0.18, 0.11, 0.03]][[0.0, 0.0, 0.0]][[0.28, 0.07]]
MSE:[[1.95, 2.13, 3.76]][[0.83, 1.23, 1.56]][[85.86, 85.86, 85.86]][[3.95, 10.97]]
MSE(-DR):[[0.0, 0.18, 1.81]][[-1.12, -0.72, -0.39]][[83.91, 83.91, 83.91]][[2.0, 9.02]]
MC-based ATE = 0.23
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.81, -0.8, -1.07]][[-1.64, -1.63, -1.66]][[-0.23, -0.23, -0.23]][-1.07]
std:[[0.01, 0.04, 0.02]][[0.08, 0.06, 0.02]][[0.0, 0.0, 0.0]][0.06]
MSE:[[0.81, 0.8, 1.07]][[1.64, 1.63, 1.66]][[0.23, 0.23, 0.23]][1.07]
MSE(-DR):[[0.0, -0.01, 0.26]][[0.83, 0.82, 0.85]][[-0.58, -0.58, -0.58]][0.26]
******
==============


O_threshold = 110
MC for this TARGET:[83.145, 0.082]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.37, -2.46, -3.38]][[-2.96, -3.25, -3.66]][[-83.14, -83.14, -83.14]][[-3.47, -8.26]]
std:[[0.45, 0.4, 0.14]][[0.1, 0.03, 0.04]][[0.0, 0.0, 0.0]][[0.18, 0.07]]
MSE:[[2.41, 2.49, 3.38]][[2.96, 3.25, 3.66]][[83.14, 83.14, 83.14]][[3.47, 8.26]]
MSE(-DR):[[0.0, 0.08, 0.97]][[0.55, 0.84, 1.25]][[80.73, 80.73, 80.73]][[1.06, 5.85]]
******
MC-based ATE = -2.48
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.24, -1.13, -0.7]][[-3.78, -3.65, -3.77]][[2.48, 2.48, 2.48]][-0.6]
std:[[0.26, 0.24, 0.05]][[0.0, 0.03, 0.04]][[0.0, 0.0, 0.0]][0.03]
MSE:[[1.27, 1.16, 0.7]][[3.78, 3.65, 3.77]][[2.48, 2.48, 2.48]][0.6]
MSE(-DR):[[0.0, -0.11, -0.57]][[2.51, 2.38, 2.5]][[1.21, 1.21, 1.21]][-0.67]
better than DR_NO_MARL
==============


time spent until now: 5.4 mins


---------------------------------------
[pattern_seed, T, sd_R] = [1, 672, 10]

max(u_O) =  141.0
O_threshold = 100
means of Order:

137.7 88.0 89.5 80.3 118.3

62.8 141.0 85.4 106.0 94.6

133.3 65.9 93.3 92.1 124.8

79.8 96.1 83.5 100.3 111.8

79.8 125.1 119.1 110.0 119.1

target policy:

1 0 0 0 1
```

```
0 1 0 1 0

1 0 0 0 1

0 0 0 1 1

0 1 1 1 1

number of reward locations:  12
O_threshold = 80
target policy:

1 1 1 1 1

0 1 1 1 1

1 0 1 1 1

0 1 1 1 1

0 1 1 1 1

number of reward locations:  21
O_threshold = 85
target policy:

1 1 1 0 1

0 1 1 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

number of reward locations:  19
O_threshold = 90
target policy:

1 0 0 0 1

0 1 0 1 1

1 0 1 1 1

0 1 0 1 1

0 1 1 1 1

number of reward locations:  16
O_threshold = 95
target policy:

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 1 0 1 1

0 1 1 1 1

number of reward locations:  13
O_threshold = 105
target policy:

1 0 0 0 1

0 1 0 1 0

1 0 0 0 1

0 0 0 0 1

0 1 1 1 1

number of reward locations:  11
O_threshold = 110
target policy:

1 0 0 0 1

0 1 0 0 0

1 0 0 0 1

0 0 0 0 1
```

```
0 1 1 1 1

number of reward locations:  10
1 2 3 4 5 6 7 1 2 3 4 5 6 7
--------------------------------------
Value of Behaviour policy:66.664
O_threshold = 100
MC for this TARGET:[77.163, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-4.77, -4.83, -5.37]][[-3.97, -4.31, -4.73]][[-77.16, -77.16, -77.16]][[-5.42, -10.5]]
std:[[0.59, 0.62, 0.42]][[0.04, 0.03, 0.21]][[0.0, 0.0, 0.0]][[0.45, 0.12]]
MSE:[[4.81, 4.87, 5.39]][[3.97, 4.31, 4.73]][[77.16, 77.16, 77.16]][[5.44, 10.5]]
MSE(-DR):[[0.0, 0.06, 0.58]][[-0.84, -0.5, -0.08]][[72.35, 72.35, 72.35]][[0.63, 5.69]]
==============


O_threshold = 80
MC for this TARGET:[73.147, 0.087]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.89, 1.73, 1.27]][[3.82, 3.48, 3.33]][[-73.15, -73.15, -73.15]][[1.11, -6.48]]
std:[[0.48, 0.47, 0.52]][[0.19, 0.18, 0.23]][[0.0, 0.0, 0.0]][[0.51, 0.12]]
MSE:[[1.95, 1.79, 1.37]][[3.82, 3.48, 3.34]][[73.15, 73.15, 73.15]][[1.22, 6.48]]
MSE(-DR):[[0.0, -0.16, -0.58]][[1.87, 1.53, 1.39]][[71.2, 71.2, 71.2]][[-0.73, 4.53]]
better than DR_NO_MARL
MC-based ATE = -4.02
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[6.66, 6.55, 6.65]][[7.79, 7.79, 8.06]][[4.02, 4.02, 4.02]][6.54]
std:[[0.11, 0.15, 0.11]][[0.15, 0.15, 0.02]][[0.0, 0.0, 0.0]][0.06]
MSE:[[6.66, 6.55, 6.65]][[7.79, 7.79, 8.06]][[4.02, 4.02, 4.02]][6.54]
MSE(-DR):[[0.0, -0.11, -0.01]][[1.13, 1.13, 1.4]][[-2.64, -2.64, -2.64]][-0.12]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[73.847, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.65, 0.52, -0.49]][[2.67, 2.31, 2.1]][[-73.85, -73.85, -73.85]][[-0.62, -7.18]]
std:[[0.59, 0.55, 0.6]][[0.15, 0.13, 0.22]][[0.0, 0.0, 0.0]][[0.56, 0.12]]
MSE:[[0.88, 0.76, 0.77]][[2.67, 2.31, 2.11]][[73.85, 73.85, 73.85]][[0.84, 7.18]]
MSE(-DR):[[0.0, -0.12, -0.11]][[1.79, 1.43, 1.23]][[72.97, 72.97, 72.97]][[-0.04, 6.3]]
better than DR_NO_MARL
MC-based ATE = -3.32
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[5.43, 5.34, 4.88]][[6.64, 6.62, 6.83]][[3.32, 3.32, 3.32]][4.8]
std:[[0.0, 0.07, 0.18]][[0.11, 0.1, 0.01]][[0.0, 0.0, 0.0]][0.11]
MSE:[[5.43, 5.34, 4.88]][[6.64, 6.62, 6.83]][[3.32, 3.32, 3.32]][4.8]
MSE(-DR):[[0.0, -0.09, -0.55]][[1.21, 1.19, 1.4]][[-2.11, -2.11, -2.11]][-0.63]
better than DR_NO_MARL
==============


O_threshold = 90
MC for this TARGET:[73.511, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.11, -0.03, -1.11]][[1.25, 0.93, 0.66]][[-73.51, -73.51, -73.51]][[-1.25, -6.85]]
std:[[0.37, 0.36, 0.55]][[0.07, 0.05, 0.19]][[0.0, 0.0, 0.0]][[0.55, 0.12]]
MSE:[[0.39, 0.36, 1.24]][[1.25, 0.93, 0.69]][[73.51, 73.51, 73.51]][[1.37, 6.85]]
MSE(-DR):[[0.0, -0.03, 0.85]][[0.86, 0.54, 0.3]][[73.12, 73.12, 73.12]][[0.98, 6.46]]
******
MC-based ATE = -3.65
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[4.88, 4.8, 4.26]][[5.22, 5.23, 5.39]][[3.65, 3.65, 3.65]][4.17]
std:[[0.22, 0.26, 0.13]][[0.04, 0.03, 0.02]][[0.0, 0.0, 0.0]][0.1]
MSE:[[4.88, 4.81, 4.26]][[5.22, 5.23, 5.39]][[3.65, 3.65, 3.65]][4.17]
MSE(-DR):[[0.0, -0.07, -0.62]][[0.34, 0.35, 0.51]][[-1.23, -1.23, -1.23]][-0.71]
better than DR_NO_MARL
==============


O_threshold = 95
MC for this TARGET:[76.208, 0.085]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-3.8, -3.87, -4.97]][[-2.84, -3.17, -3.55]][[-76.21, -76.21, -76.21]][[-5.04, -9.54]]
std:[[0.57, 0.56, 0.45]][[0.04, 0.03, 0.21]][[0.0, 0.0, 0.0]][[0.44, 0.12]]
MSE:[[3.84, 3.91, 4.99]][[2.84, 3.17, 3.56]][[76.21, 76.21, 76.21]][[5.06, 9.54]]
MSE(-DR):[[0.0, 0.07, 1.15]][[-1.0, -0.67, -0.28]][[72.37, 72.37, 72.37]][[1.22, 5.7]]
MC-based ATE = -0.95
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.97, 0.95, 0.4]][[1.12, 1.13, 1.18]][[0.95, 0.95, 0.95]][0.38]
std:[[0.02, 0.06, 0.03]][[0.0, 0.01, 0.01]][[0.0, 0.0, 0.0]][0.01]
MSE:[[0.97, 0.95, 0.4]][[1.12, 1.13, 1.18]][[0.95, 0.95, 0.95]][0.38]
MSE(-DR):[[0.0, -0.02, -0.57]][[0.15, 0.16, 0.21]][[-0.02, -0.02, -0.02]][-0.59]
better than DR_NO_MARL
==============


O_threshold = 105
```

```
MC for this TARGET:[79.33, 0.084]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-6.58, -6.67, -7.38]][[-6.31, -6.66, -7.19]][[-79.33, -79.33, -79.33]][[-7.47, -12.67]]
std:[[0.58, 0.63, 0.26]][[0.03, 0.03, 0.15]][[0.0, 0.0, 0.0]][[0.31, 0.12]]
MSE:[[6.61, 6.7, 7.38]][[6.31, 6.66, 7.19]][[79.33, 79.33, 79.33]][[7.48, 12.67]]
MSE(-DR):[[0.0, 0.09, 0.77]][[-0.3, 0.05, 0.58]][[72.72, 72.72, 72.72]][[0.87, 6.06]]
MC-based ATE = 2.17
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-1.81, -1.84, -2.01]][[-2.34, -2.35, -2.45]][[-2.17, -2.17, -2.17]][-2.04]
std:[[0.01, 0.01, 0.16]][[0.07, 0.06, 0.06]][[0.0, 0.0, 0.0]][0.14]
MSE:[[1.81, 1.84, 2.02]][[2.34, 2.35, 2.45]][[2.17, 2.17, 2.17]][2.04]
MSE(-DR):[[0.0, 0.03, 0.21]][[0.53, 0.54, 0.64]][[0.36, 0.36, 0.36]][0.23]
******
==============


O_threshold = 110
MC for this TARGET:[80.264, 0.083]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-7.25, -7.29, -7.7]][[-7.65, -8.01, -8.61]][[-80.26, -80.26, -80.26]][[-7.74, -13.6]]
std:[[0.49, 0.56, 0.24]][[0.06, 0.04, 0.21]][[0.0, 0.0, 0.0]][[0.31, 0.12]]
MSE:[[7.27, 7.31, 7.7]][[7.65, 8.01, 8.61]][[80.26, 80.26, 80.26]][[7.75, 13.6]]
MSE(-DR):[[0.0, 0.04, 0.43]][[0.38, 0.74, 1.34]][[72.99, 72.99, 72.99]][[0.48, 6.33]]
******
MC-based ATE = 3.1
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-2.48, -2.46, -2.33]][[-3.69, -3.71, -3.88]][[-3.1, -3.1, -3.1]][-2.31]
std:[[0.09, 0.05, 0.18]][[0.02, 0.02, 0.0]][[0.0, 0.0, 0.0]][0.14]
MSE:[[2.48, 2.46, 2.34]][[3.69, 3.71, 3.88]][[3.1, 3.1, 3.1]][2.31]
MSE(-DR):[[0.0, -0.02, -0.14]][[1.21, 1.23, 1.4]][[0.62, 0.62, 0.62]][-0.17]
better than DR_NO_MARL
==============


time spent until now: 10.8 mins


----------------------------------------
[pattern_seed, T, sd_R] = [2, 672, 10]

max(u_O) =  157.3
O_threshold = 100
means of Order:

91.5 98.4 64.9 138.1 69.5

84.1 110.0 77.6 80.5 82.9

111.1 157.3 100.3 79.6 110.8

88.3 99.1 125.8 85.7 99.7

83.5 96.4 104.7 81.6 93.0

target policy:

0 0 0 1 0

0 1 0 0 0

1 1 1 0 1

0 0 1 0 0

0 0 1 0 0

number of reward locations:  8
O_threshold = 80
target policy:

1 1 0 1 0

1 1 0 1 1

1 1 1 0 1

1 1 1 1 1

1 1 1 1 1

number of reward locations:  21
O_threshold = 85
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1
```

```
1 1 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 90
target policy:

1 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 1

number of reward locations:  14
O_threshold = 95
target policy:

0 1 0 1 0

0 1 0 0 0

1 1 1 0 1

0 1 1 0 1

0 1 1 0 0

number of reward locations:  12
O_threshold = 105
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
O_threshold = 110
target policy:

0 0 0 1 0

0 1 0 0 0

1 1 0 0 1

0 0 1 0 0

0 0 0 0 0

number of reward locations:  6
1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:65.201
O_threshold = 100
MC for this TARGET:[71.401, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.07, -2.11, -3.15]][[-3.97, -4.16, -4.48]][[-71.4, -71.4, -71.4]][[-3.19, -6.2]]
std:[[0.06, 0.06, 0.08]][[0.05, 0.07, 0.05]][[0.0, 0.0, 0.0]][[0.08, 0.0]]
MSE:[[2.07, 2.11, 3.15]][[3.97, 4.16, 4.48]][[71.4, 71.4, 71.4]][[3.19, 6.2]]
MSE(-DR):[[0.0, 0.04, 1.08]][[1.9, 2.09, 2.41]][[69.33, 69.33, 69.33]][[1.12, 4.13]]
******
==============


O_threshold = 80
MC for this TARGET:[73.746, 0.09]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.81, 0.67, -0.62]][[3.09, 2.67, 2.5]][[-73.75, -73.75, -73.75]][[-0.77, -8.55]]
std:[[0.27, 0.21, 0.25]][[0.04, 0.0, 0.04]][[0.0, 0.0, 0.0]][[0.19, 0.0]]
MSE:[[0.85, 0.7, 0.67]][[3.09, 2.67, 2.5]][[73.75, 73.75, 73.75]][[0.79, 8.55]]
MSE(-DR):[[0.0, -0.15, -0.18]][[2.24, 1.82, 1.65]][[72.9, 72.9, 72.9]][[-0.06, 7.7]]
better than DR_NO_MARL
MC-based ATE = 2.34
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.89, 2.78, 2.53]][[7.06, 6.83, 6.98]][[-2.34, -2.34, -2.34]][2.43]
std:[[0.21, 0.16, 0.17]][[0.1, 0.08, 0.01]][[0.0, 0.0, 0.0]][0.11]
MSE:[[2.9, 2.78, 2.54]][[7.06, 6.83, 6.98]][[2.34, 2.34, 2.34]][2.43]
```

```
MSE(-DR):[[0.0, -0.12, -0.36]][[4.16, 3.93, 4.08]][[-0.56, -0.56, -0.56]][-0.47]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[72.145, 0.082]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.32, 0.14, -1.0]][[1.41, 1.08, 0.86]][[-72.14, -72.14, -72.14]][[-1.18, -6.94]]
std:[[0.02, 0.01, 0.07]][[0.0, 0.03, 0.08]][[0.0, 0.0, 0.0]][[0.1, 0.0]]
MSE:[[0.32, 0.14, 1.0]][[1.41, 1.08, 0.86]][[72.14, 72.14, 72.14]][[1.18, 6.94]]
MSE(-DR):[[0.0, -0.18, 0.68]][[1.09, 0.76, 0.54]][[71.82, 71.82, 71.82]][[0.86, 6.62]]
******
MC-based ATE = 0.74
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.4, 2.25, 2.16]][[5.38, 5.23, 5.34]][[-0.74, -0.74, -0.74]][2.01]
std:[[0.08, 0.05, 0.01]][[0.05, 0.04, 0.03]][[0.0, 0.0, 0.0]][0.02]
MSE:[[2.4, 2.25, 2.16]][[5.38, 5.23, 5.34]][[0.74, 0.74, 0.74]][2.01]
MSE(-DR):[[0.0, -0.15, -0.24]][[2.98, 2.83, 2.94]][[-1.66, -1.66, -1.66]][-0.39]
better than DR_NO_MARL
==============


O_threshold = 90
MC for this TARGET:[73.239, 0.084]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.27, -0.41, -2.02]][[-0.56, -0.87, -1.12]][[-73.24, -73.24, -73.24]][[-2.15, -8.04]]
std:[[0.19, 0.19, 0.1]][[0.06, 0.09, 0.1]][[0.0, 0.0, 0.0]][[0.1, 0.0]]
MSE:[[0.33, 0.45, 2.02]][[0.56, 0.87, 1.12]][[73.24, 73.24, 73.24]][[2.15, 8.04]]
MSE(-DR):[[0.0, 0.12, 1.69]][[0.23, 0.54, 0.79]][[72.91, 72.91, 72.91]][[1.82, 7.71]]
******
MC-based ATE = 1.84
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.8, 1.71, 1.14]][[3.41, 3.28, 3.36]][[-1.84, -1.84, -1.84]][1.04]
std:[[0.25, 0.24, 0.18]][[0.01, 0.02, 0.05]][[0.0, 0.0, 0.0]][0.17]
MSE:[[1.82, 1.73, 1.15]][[3.41, 3.28, 3.36]][[1.84, 1.84, 1.84]][1.05]
MSE(-DR):[[0.0, -0.09, -0.67]][[1.59, 1.46, 1.54]][[0.02, 0.02, 0.02]][-0.77]
better than DR_NO_MARL
==============


O_threshold = 95
MC for this TARGET:[72.055, 0.082]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-0.41, -0.46, -1.53]][[-1.07, -1.32, -1.59]][[-72.06, -72.06, -72.06]][[-1.58, -6.85]]
std:[[0.27, 0.24, 0.06]][[0.03, 0.05, 0.04]][[0.0, 0.0, 0.0]][[0.04, 0.0]]
MSE:[[0.49, 0.52, 1.53]][[1.07, 1.32, 1.59]][[72.06, 72.06, 72.06]][[1.58, 6.85]]
MSE(-DR):[[0.0, 0.03, 1.04]][[0.58, 0.83, 1.1]][[71.57, 71.57, 71.57]][[1.09, 6.36]]
******
MC-based ATE = 0.65
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.66, 1.65, 1.62]][[2.9, 2.84, 2.89]][[-0.65, -0.65, -0.65]][1.61]
std:[[0.33, 0.3, 0.14]][[0.02, 0.02, 0.01]][[0.0, 0.0, 0.0]][0.11]
MSE:[[1.69, 1.68, 1.63]][[2.9, 2.84, 2.89]][[0.65, 0.65, 0.65]][1.61]
MSE(-DR):[[0.0, -0.01, -0.06]][[1.21, 1.15, 1.2]][[-1.04, -1.04, -1.04]][-0.08]
better than DR_NO_MARL
==============


O_threshold = 105
MC for this TARGET:[72.499, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-5.09, -5.09, -6.18]][[-6.94, -7.07, -7.43]][[-72.5, -72.5, -72.5]][[-6.17, -7.3]]
std:[[0.02, 0.04, 0.25]][[0.06, 0.05, 0.01]][[0.0, 0.0, 0.0]][[0.23, 0.0]]
MSE:[[5.09, 5.09, 6.19]][[6.94, 7.07, 7.43]][[72.5, 72.5, 72.5]][[6.17, 7.3]]
MSE(-DR):[[0.0, 0.0, 1.1]][[1.85, 1.98, 2.34]][[67.41, 67.41, 67.41]][[1.08, 2.21]]
******
MC-based ATE = 1.1
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.02, -2.97, -3.02]][[-2.96, -2.92, -2.95]][[-1.1, -1.1, -1.1]][-2.98]
std:[[0.08, 0.1, 0.17]][[0.0, 0.02, 0.04]][[0.0, 0.0, 0.0]][0.15]
MSE:[[3.02, 2.97, 3.02]][[2.96, 2.92, 2.95]][[1.1, 1.1, 1.1]][2.98]
MSE(-DR):[[0.0, -0.05, 0.0]][[-0.06, -0.1, -0.07]][[-1.92, -1.92, -1.92]][-0.04]
==============


O_threshold = 110
MC for this TARGET:[72.499, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-5.11, -5.09, -6.16]][[-6.94, -7.07, -7.44]][[-72.5, -72.5, -72.5]][[-6.14, -7.3]]
std:[[0.02, 0.04, 0.26]][[0.05, 0.05, 0.01]][[0.0, 0.0, 0.0]][[0.25, 0.0]]
MSE:[[5.11, 5.09, 6.17]][[6.94, 7.07, 7.44]][[72.5, 72.5, 72.5]][[6.15, 7.3]]
MSE(-DR):[[0.0, -0.02, 1.06]][[1.83, 1.96, 2.33]][[67.39, 67.39, 67.39]][[1.04, 2.19]]
******
MC-based ATE = 1.1
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-3.04, -2.97, -3.01]][[-2.97, -2.92, -2.95]][[-1.1, -1.1, -1.1]][-2.94]
std:[[0.08, 0.1, 0.18]][[0.0, 0.02, 0.04]][[0.0, 0.0, 0.0]][0.17]
MSE:[[3.04, 2.97, 3.02]][[2.97, 2.92, 2.95]][[1.1, 1.1, 1.1]][2.94]
```

MSE(-DR):[[0.0, -0.07, -0.02]][[-0.07, -0.12, -0.09]][[-1.94, -1.94, -1.94]][-0.1]
==========

time spent until now: 16.3 mins

[pattern_seed, T, sd_R] = [3, 672, 10]

max(u_O) = 142.3
O_threshold = 100
means of Order:

142.3 108.6 101.4 68.5 94.1

92.7 97.9 87.8 98.6 90.4

76.5 118.7 118.7 140.0 100.5

91.7 89.2 73.0 121.1 79.8

78.5 95.5 133.9 104.3 81.1

target policy:

1 1 1 0 0

0 0 0 0 0

0 1 1 1 1

0 0 0 1 0

0 0 1 1 0

number of reward locations: 10
O_threshold = 80
target policy:

1 1 1 0 1

1 1 1 1 1

0 1 1 1 1

1 1 0 1 0

0 1 1 1 1

number of reward locations: 20
O_threshold = 85
target policy:

1 1 1 0 1

1 1 1 1 1

0 1 1 1 1

1 1 0 1 0

0 1 1 1 0

number of reward locations: 19
O_threshold = 90
target policy:

1 1 1 0 1

1 1 0 1 1

0 1 1 1 1

1 0 0 1 0

0 1 1 1 0

number of reward locations: 17
O_threshold = 95
target policy:

1 1 1 0 0

0 1 0 1 0

0 1 1 1 1

0 0 0 1 0

```
0 1 1 1 0

number of reward locations:  13
O_threshold = 105
target policy:

1 1 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 0 0

number of reward locations:  7
O_threshold = 110
target policy:

1 0 0 0 0

0 0 0 0 0

0 1 1 1 0

0 0 0 1 0

0 0 1 0 0

number of reward locations:  6
1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:67.173
O_threshold = 100
MC for this TARGET:[75.339, 0.095]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.19, -2.28, -3.88]][[-3.38, -3.61, -4.1]][[-75.34, -75.34, -75.34]][[-3.97, -8.17]]
std:[[0.73, 0.75, 0.11]][[0.11, 0.13, 0.04]][[0.0, 0.0, 0.0]][[0.08, 0.05]]
MSE:[[2.31, 2.4, 3.88]][[3.38, 3.61, 4.1]][[75.34, 75.34, 75.34]][[3.97, 8.17]]
MSE(-DR):[[0.0, 0.09, 1.57]][[1.07, 1.3, 1.79]][[73.03, 73.03, 73.03]][[1.66, 5.86]]
******
==============


O_threshold = 80
MC for this TARGET:[75.421, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.4, 1.32, -0.5]][[2.71, 2.4, 2.06]][[-75.42, -75.42, -75.42]][[-0.57, -8.25]]
std:[[0.26, 0.26, 0.07]][[0.16, 0.16, 0.04]][[0.0, 0.0, 0.0]][[0.07, 0.05]]
MSE:[[1.42, 1.35, 0.5]][[2.71, 2.41, 2.06]][[75.42, 75.42, 75.42]][[0.57, 8.25]]
MSE(-DR):[[0.0, -0.07, -0.92]][[1.29, 0.99, 0.64]][[74.0, 74.0, 74.0]][[-0.85, 6.83]]
better than DR_NO_MARL
MC-based ATE = 0.08
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[3.59, 3.6, 3.39]][[6.09, 6.01, 6.16]][[-0.08, -0.08, -0.08]][3.4]
std:[[0.98, 1.01, 0.18]][[0.05, 0.04, 0.0]][[0.0, 0.0, 0.0]][0.15]
MSE:[[3.72, 3.74, 3.39]][[6.09, 6.01, 6.16]][[0.08, 0.08, 0.08]][3.4]
MSE(-DR):[[0.0, 0.02, -0.33]][[2.37, 2.29, 2.44]][[-3.64, -3.64, -3.64]][-0.32]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[74.869, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.21, 1.13, -0.77]][[2.35, 2.07, 1.75]][[-74.87, -74.87, -74.87]][[-0.85, -7.7]]
std:[[0.3, 0.34, 0.09]][[0.2, 0.2, 0.09]][[0.0, 0.0, 0.0]][[0.05, 0.05]]
MSE:[[1.25, 1.18, 0.78]][[2.36, 2.08, 1.75]][[74.87, 74.87, 74.87]][[0.85, 7.7]]
MSE(-DR):[[0.0, -0.07, -0.47]][[1.11, 0.83, 0.5]][[73.62, 73.62, 73.62]][[-0.4, 6.45]]
better than DR_NO_MARL
MC-based ATE = -0.47
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[3.4, 3.41, 3.11]][[5.73, 5.68, 5.84]][[0.47, 0.47, 0.47]][3.12]
std:[[1.03, 1.09, 0.2]][[0.09, 0.07, 0.05]][[0.0, 0.0, 0.0]][0.14]
MSE:[[3.55, 3.58, 3.12]][[5.73, 5.68, 5.84]][[0.47, 0.47, 0.47]][3.12]
MSE(-DR):[[0.0, 0.03, -0.43]][[2.18, 2.13, 2.29]][[-3.08, -3.08, -3.08]][-0.43]
better than DR_NO_MARL
==============


O_threshold = 90
MC for this TARGET:[76.607, 0.089]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.05, -0.12, -2.32]][[0.65, 0.34, -0.06]][[-76.61, -76.61, -76.61]][[-2.49, -9.43]]
std:[[0.12, 0.12, 0.13]][[0.1, 0.1, 0.0]][[0.0, 0.0, 0.0]][[0.13, 0.05]]
MSE:[[0.13, 0.17, 2.32]][[0.66, 0.35, 0.06]][[76.61, 76.61, 76.61]][[2.49, 9.43]]
MSE(-DR):[[0.0, 0.04, 2.19]][[0.53, 0.22, -0.07]][[76.48, 76.48, 76.48]][[2.36, 9.3]]
```

******
MC-based ATE = 1.27
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[2.23, 2.16, 1.56]][[4.03, 3.95, 4.04]][[-1.27, -1.27, -1.27]][1.48]
std:[[0.85, 0.87, 0.24]][[0.01, 0.03, 0.04]][[0.0, 0.0, 0.0]][0.21]
MSE:[[2.39, 2.33, 1.58]][[4.03, 3.95, 4.04]][[1.27, 1.27, 1.27]][1.49]
MSE(-DR):[[0.0, -0.06, -0.81]][[1.64, 1.56, 1.65]][[-1.12, -1.12, -1.12]][-0.9]
better than DR_NO_MARL
==============

O_threshold = 95
MC for this TARGET:[74.084, 0.094]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-1.18, -1.29, -2.63]][[-0.71, -0.93, -1.28]][[-74.08, -74.08, -74.08]][[-2.74, -6.91]]
std:[[0.13, 0.14, 0.02]][[0.04, 0.07, 0.05]][[0.0, 0.0, 0.0]][[0.04, 0.05]]
MSE:[[1.19, 1.3, 2.63]][[0.71, 0.93, 1.28]][[74.08, 74.08, 74.08]][[2.74, 6.91]]
MSE(-DR):[[0.0, 0.11, 1.44]][[-0.48, -0.26, 0.09]][[72.89, 72.89, 72.89]][[1.55, 5.72]]
MC-based ATE = -1.25
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.01, 0.99, 1.25]][[2.68, 2.67, 2.82]][[1.25, 1.25, 1.25]][1.24]
std:[[0.6, 0.61, 0.13]][[0.06, 0.06, 0.09]][[0.0, 0.0, 0.0]][0.12]
MSE:[[1.17, 1.16, 1.26]][[2.68, 2.67, 2.82]][[1.25, 1.25, 1.25]][1.25]
MSE(-DR):[[0.0, -0.01, 0.09]][[1.51, 1.5, 1.65]][[0.08, 0.08, 0.08]][0.08]
******
==============

O_threshold = 105
MC for this TARGET:[72.771, 0.096]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.98, -2.96, -3.67]][[-4.63, -4.74, -5.12]][[-72.77, -72.77, -72.77]][[-3.65, -5.6]]
std:[[0.44, 0.4, 0.11]][[0.02, 0.01, 0.04]][[0.0, 0.0, 0.0]][[0.15, 0.05]]
MSE:[[3.01, 2.99, 3.67]][[4.63, 4.74, 5.12]][[72.77, 72.77, 72.77]][[3.65, 5.6]]
MSE(-DR):[[0.0, -0.02, 0.66]][[1.62, 1.73, 2.11]][[69.76, 69.76, 69.76]][[0.64, 2.59]]
******
MC-based ATE = -2.57
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.79, -0.69, 0.22]][[-1.25, -1.13, -1.02]][[2.57, 2.57, 2.57]][0.32]
std:[[0.28, 0.35, 0.0]][[0.12, 0.14, 0.08]][[0.0, 0.0, 0.0]][0.07]
MSE:[[0.84, 0.77, 0.22]][[1.26, 1.14, 1.02]][[2.57, 2.57, 2.57]][0.33]
MSE(-DR):[[0.0, -0.07, -0.62]][[0.42, 0.3, 0.18]][[1.73, 1.73, 1.73]][-0.51]
better than DR_NO_MARL
==============

O_threshold = 110
MC for this TARGET:[71.568, 0.097]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-2.69, -2.69, -3.31]][[-4.72, -4.8, -5.14]][[-71.57, -71.57, -71.57]][[-3.3, -4.39]]
std:[[0.41, 0.35, 0.01]][[0.01, 0.01, 0.02]][[0.0, 0.0, 0.0]][[0.08, 0.05]]
MSE:[[2.72, 2.71, 3.31]][[4.72, 4.8, 5.14]][[71.57, 71.57, 71.57]][[3.3, 4.39]]
MSE(-DR):[[0.0, -0.01, 0.59]][[2.0, 2.08, 2.42]][[68.85, 68.85, 68.85]][[0.58, 1.67]]
******
MC-based ATE = -3.77
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.51, -0.41, 0.58]][[-1.34, -1.19, -1.04]][[3.77, 3.77, 3.77]][0.67]
std:[[0.31, 0.4, 0.1]][[0.12, 0.14, 0.06]][[0.0, 0.0, 0.0]][0.01]
MSE:[[0.6, 0.57, 0.59]][[1.35, 1.2, 1.04]][[3.77, 3.77, 3.77]][0.67]
MSE(-DR):[[0.0, -0.03, -0.01]][[0.75, 0.6, 0.44]][[3.17, 3.17, 3.17]][0.07]
better than DR_NO_MARL
==============

time spent until now: 21.7 mins

[pattern_seed, T, sd_R] = [4, 672, 10]

max(u_O) = 155.2
O_threshold = 100
means of Order:

100.5 109.9 81.5 114.3 91.5

72.5 87.4 112.1 106.3 79.1

112.6 97.7 108.3 106.3 78.9

106.7 88.1 135.6 115.0 100.4

81.7 100.6 102.7 78.1 155.2

target policy:

1 1 0 1 0

0 0 1 1 0

```
1 0 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  15
O_threshold = 80
target policy:

1 1 1 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

1 1 1 0 1

number of reward locations:  21
O_threshold = 85
target policy:

1 1 0 1 1

0 1 1 1 0

1 1 1 1 0

1 1 1 1 1

0 1 1 0 1

number of reward locations:  19
O_threshold = 90
target policy:

1 1 0 1 1

0 0 1 1 0

1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  17
O_threshold = 95
target policy:

1 1 0 1 0

0 0 1 1 0

1 1 1 1 0

1 0 1 1 1

0 1 1 0 1

number of reward locations:  16
O_threshold = 105
target policy:

0 1 0 1 0

0 0 1 1 0

1 0 1 1 0

1 0 1 1 0

0 0 0 0 1

number of reward locations:  11
O_threshold = 110
target policy:

0 0 0 1 0

0 0 1 0 0

1 0 0 0 0

0 0 1 1 0
```

```
0 0 0 0 1

number of reward locations:  6
1 2 3 4 5 6 7 1 2 3 4 5 6 7
----------------------------------------
Value of Behaviour policy:68.414
O_threshold = 100
MC for this TARGET:[75.62, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.73, 0.56, -0.98]][[0.94, 0.67, 0.39]][[-75.62, -75.62, -75.62]][[-1.14, -7.21]]
std:[[0.08, 0.09, 0.03]][[0.17, 0.17, 0.06]][[0.0, 0.0, 0.0]][[0.02, 0.01]]
MSE:[[0.73, 0.57, 0.98]][[0.96, 0.69, 0.39]][[75.62, 75.62, 75.62]][[1.14, 7.21]]
MSE(-DR):[[0.0, -0.16, 0.25]][[0.23, -0.04, -0.34]][[74.89, 74.89, 74.89]][[0.41, 6.48]]
******
==============


O_threshold = 80
MC for this TARGET:[75.342, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[2.06, 1.87, 0.54]][[3.71, 3.38, 3.34]][[-75.34, -75.34, -75.34]][[0.35, -6.93]]
std:[[0.68, 0.76, 0.04]][[0.11, 0.14, 0.03]][[0.0, 0.0, 0.0]][[0.05, 0.01]]
MSE:[[2.17, 2.02, 0.54]][[3.71, 3.38, 3.34]][[75.34, 75.34, 75.34]][[0.35, 6.93]]
MSE(-DR):[[0.0, -0.15, -1.63]][[1.54, 1.21, 1.17]][[73.17, 73.17, 73.17]][[-1.82, 4.76]]
better than DR_NO_MARL
MC-based ATE = -0.28
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.33, 1.31, 1.52]][[2.76, 2.71, 2.95]][[0.28, 0.28, 0.28]][1.5]
std:[[0.76, 0.85, 0.07]][[0.07, 0.03, 0.03]][[0.0, 0.0, 0.0]][0.02]
MSE:[[1.53, 1.56, 1.52]][[2.76, 2.71, 2.95]][[0.28, 0.28, 0.28]][1.5]
MSE(-DR):[[0.0, 0.03, -0.01]][[1.23, 1.18, 1.42]][[-1.25, -1.25, -1.25]][-0.03]
better than DR_NO_MARL
==============


O_threshold = 85
MC for this TARGET:[74.187, 0.087]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[1.89, 1.76, 0.3]][[3.22, 2.93, 2.93]][[-74.19, -74.19, -74.19]][[0.18, -5.77]]
std:[[0.4, 0.46, 0.04]][[0.09, 0.11, 0.03]][[0.0, 0.0, 0.0]][[0.02, 0.01]]
MSE:[[1.93, 1.82, 0.3]][[3.22, 2.93, 2.93]][[74.19, 74.19, 74.19]][[0.18, 5.77]]
MSE(-DR):[[0.0, -0.11, -1.63]][[1.29, 1.0, 1.0]][[72.26, 72.26, 72.26]][[-1.75, 3.84]]
better than DR_NO_MARL
MC-based ATE = -1.43
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[1.17, 1.2, 1.29]][[2.28, 2.27, 2.54]][[1.43, 1.43, 1.43]][1.32]
std:[[0.48, 0.55, 0.07]][[0.08, 0.06, 0.03]][[0.0, 0.0, 0.0]][0.0]
MSE:[[1.26, 1.32, 1.29]][[2.28, 2.27, 2.54]][[1.43, 1.43, 1.43]][1.32]
MSE(-DR):[[0.0, 0.06, 0.03]][[1.02, 1.01, 1.28]][[0.17, 0.17, 0.17]][0.06]
******
==============


O_threshold = 90
MC for this TARGET:[75.814, 0.087]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.79, 0.65, -0.95]][[1.99, 1.67, 1.43]][[-75.81, -75.81, -75.81]][[-1.09, -7.4]]
std:[[0.26, 0.25, 0.02]][[0.12, 0.13, 0.03]][[0.0, 0.0, 0.0]][[0.03, 0.01]]
MSE:[[0.83, 0.7, 0.95]][[1.99, 1.68, 1.43]][[75.81, 75.81, 75.81]][[1.09, 7.4]]
MSE(-DR):[[0.0, -0.13, 0.12]][[1.16, 0.85, 0.6]][[74.98, 74.98, 74.98]][[0.26, 6.57]]
******
MC-based ATE = 0.19
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[0.07, 0.09, 0.03]][[1.04, 1.01, 1.04]][[-0.19, -0.19, -0.19]][0.06]
std:[[0.34, 0.34, 0.05]][[0.05, 0.04, 0.03]][[0.0, 0.0, 0.0]][0.05]
MSE:[[0.35, 0.35, 0.06]][[1.04, 1.01, 1.04]][[0.19, 0.19, 0.19]][0.08]
MSE(-DR):[[0.0, 0.0, -0.29]][[0.69, 0.66, 0.69]][[-0.16, -0.16, -0.16]][-0.27]
better than DR_NO_MARL
==============


O_threshold = 95
MC for this TARGET:[75.114, 0.088]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.65, 0.51, -0.85]][[1.76, 1.49, 1.24]][[-75.11, -75.11, -75.11]][[-0.99, -6.7]]
std:[[0.3, 0.27, 0.1]][[0.19, 0.19, 0.08]][[0.0, 0.0, 0.0]][[0.12, 0.01]]
MSE:[[0.72, 0.58, 0.86]][[1.77, 1.5, 1.24]][[75.11, 75.11, 75.11]][[1.0, 6.7]]
MSE(-DR):[[0.0, -0.14, 0.14]][[1.05, 0.78, 0.52]][[74.39, 74.39, 74.39]][[0.28, 5.98]]
******
MC-based ATE = -0.51
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.08, -0.06, 0.13]][[0.82, 0.83, 0.85]][[0.51, 0.51, 0.51]][0.15]
std:[[0.38, 0.36, 0.13]][[0.01, 0.02, 0.02]][[0.0, 0.0, 0.0]][0.14]
MSE:[[0.39, 0.36, 0.18]][[0.82, 0.83, 0.85]][[0.51, 0.51, 0.51]][0.21]
MSE(-DR):[[0.0, -0.03, -0.21]][[0.43, 0.44, 0.46]][[0.12, 0.12, 0.12]][-0.18]
better than DR_NO_MARL
==============
```

```
O_threshold = 105
MC for this TARGET:[72.071, 0.086]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[0.37, 0.34, -0.35]][[0.08, -0.06, -0.32]][[-72.07, -72.07, -72.07]][[-0.37, -3.66]]
std:[[0.46, 0.47, 0.23]][[0.3, 0.28, 0.15]][[0.0, 0.0, 0.0]][[0.23, 0.01]]
MSE:[[0.59, 0.58, 0.42]][[0.31, 0.29, 0.35]][[72.07, 72.07, 72.07]][[0.44, 3.66]]
MSE(-DR):[[0.0, -0.01, -0.17]][[-0.28, -0.3, -0.24]][[71.48, 71.48, 71.48]][[-0.15, 3.07]]
MC-based ATE = -3.55
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-0.36, -0.22, 0.64]][[-0.86, -0.72, -0.71]][[3.55, 3.55, 3.55]][0.77]
std:[[0.39, 0.38, 0.26]][[0.12, 0.11, 0.09]][[0.0, 0.0, 0.0]][0.26]
MSE:[[0.53, 0.44, 0.69]][[0.87, 0.73, 0.72]][[3.55, 3.55, 3.55]][0.81]
MSE(-DR):[[0.0, -0.09, 0.16]][[0.34, 0.2, 0.19]][[3.02, 3.02, 3.02]][0.28]
******
==============


O_threshold = 110
MC for this TARGET:[74.977, 0.079]
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR/QV/IS]_NO_MF; [DR2, V_behav]
bias:[[-5.96, -5.92, -6.66]][[-7.26, -7.33, -7.74]][[-74.98, -74.98, -74.98]][[-6.62, -6.56]]
std:[[0.01, 0.01, 0.35]][[0.14, 0.14, 0.08]][[0.0, 0.0, 0.0]][[0.36, 0.01]]
MSE:[[5.96, 5.92, 6.67]][[7.26, 7.33, 7.74]][[74.98, 74.98, 74.98]][[6.63, 6.56]]
MSE(-DR):[[0.0, -0.04, 0.71]][[1.3, 1.37, 1.78]][[69.02, 69.02, 69.02]][[0.67, 0.6]]
******
MC-based ATE = -0.64
   [DR/QV/IS]; [DR/QV/IS]_NO_MARL; [DR2]
bias:[[-6.69, -6.48, -5.67]][[-8.21, -8.0, -8.13]][[0.64, 0.64, 0.64]][-5.47]
std:[[0.07, 0.08, 0.38]][[0.04, 0.04, 0.02]][[0.0, 0.0, 0.0]][0.38]
MSE:[[6.69, 6.48, 5.68]][[8.21, 8.0, 8.13]][[0.64, 0.64, 0.64]][5.48]
MSE(-DR):[[0.0, -0.21, -1.01]][[1.52, 1.31, 1.44]][[-6.05, -6.05, -6.05]][-1.21]
better than DR_NO_MARL
==============


time spent until now: 27.2 mins

ubuntu@ip-172-31-9-80:~$
```