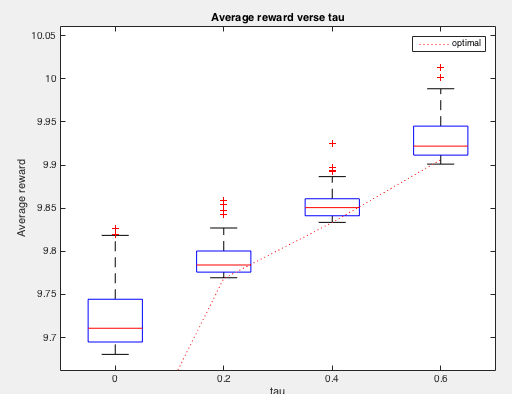
Project: Test bandit algorithm performance taking into consider the burden effect (when context are not iid)

The assumption of bandit algorithm is that context is iid, and cannot be affected by action taken earlier. So we want to see the performance of the algorithm when action can affect context.

Last time, we simulated the experiment with **tau varying from 0 to 0.6** and see the performance of the algorithm. It works well when tau is not equal to 0. The population average cost points are at the bottom of boxplots. But for tau = 0, the optimal point does not touch the boxplot. **So we suspect that the algorithm does not work well (or there might be a bug) when tau = 0. So we fix tau = 0 and set coefficients of the action interaction terms as variables, increasing from 0 to 1.**



On the other hand, this experiment can be seen as a test for **learning effect** in the model. **Tau = 0, there is no burden effect**

The S2 can be seen as a learning factor, so we see the performance of the algorithm under the learning effect. S2 has interaction with action. So when the coefficient of S2= …+ coefficient\*A increase, we expect to see the average cost to reduce.

As the coefficient increases, the average cost as expected decreases. Also the optimal average cost line touch the bottom of the boxplots, which proves the performance of the algorithm when tau = 0 is good.

Compare population level theta and theta we got from the simulation experiment:

When changing x coefficient from 0, 0.2, … to 1

**Optimal theta** is always [0.64 0.60 0.96 0]. (Why these are all same?)

Mean of theta got from experiments

exp\_mean\_theta =

 x\_coeff 0  0.6428    0.5413    0.9273   -0.0157

 x\_coeff 0.2  0.6826    0.5146    0.8653   -0.0541

 x\_coeff 0.4   0.6793    0.4875    0.8575   -0.0084

 x\_coeff 0.6   0.7419    0.4492    0.8031   -0.0140

 x\_coeff 0.8   0.8262    0.4251    0.7275   -0.0178

 x\_coeff 1   0.7194    0.3930    0.6190   -0.0050

bias =

 x\_coeff 0   0.0039   -0.0609   -0.0325   -0.0164

 x\_coeff 0.2   0.0437   -0.0875   -0.0944   -0.0548

 x\_coeff 0.4   0.0404   -0.1146   -0.1022   -0.0091

 x\_coeff 0.6   0.1031   -0.1529   -0.1567   -0.0146

 x\_coeff 0.8   0.1874   -0.1770   -0.2322   -0.0185

 x\_coeff 1   0.0805   -0.2092   -0.3407   -0.0056

Since we the population theta are all positive numbers, I calculated the percentage of positive numbers of the experiment theta

percentage =

x\_coeff 0    0.9500    0.9200    0.9900    0.5100

x\_coeff 0.2    0.9500    0.9600    1.0000    0.4500

x\_coeff 0.4    0.9000    0.9500    1.0000    0.5300

x\_coeff 0.6    0.9100    0.9200    0.9900    0.4600

x\_coeff 0.8    0.9700    0.9400    0.9900    0.4800

x\_coeff 1    0.9600    0.9600    0.9800    0.5400

Result for when tau = 0 and context are simulated from:

S(t,1) = S(t-1,1)\*0.4 + e(t,1)

S(t,2) = S(t-1,2)\*0.25 + 8\*A(t-1)\***x\_co** + e(t,2)

S(t,3) = S(t-1,3)\*0.5 + 0.05\*S(t-1,3)\*A(t-1)\***x\_co** + 0.5\*A(t-1)\***x\_co** + e(t,3)

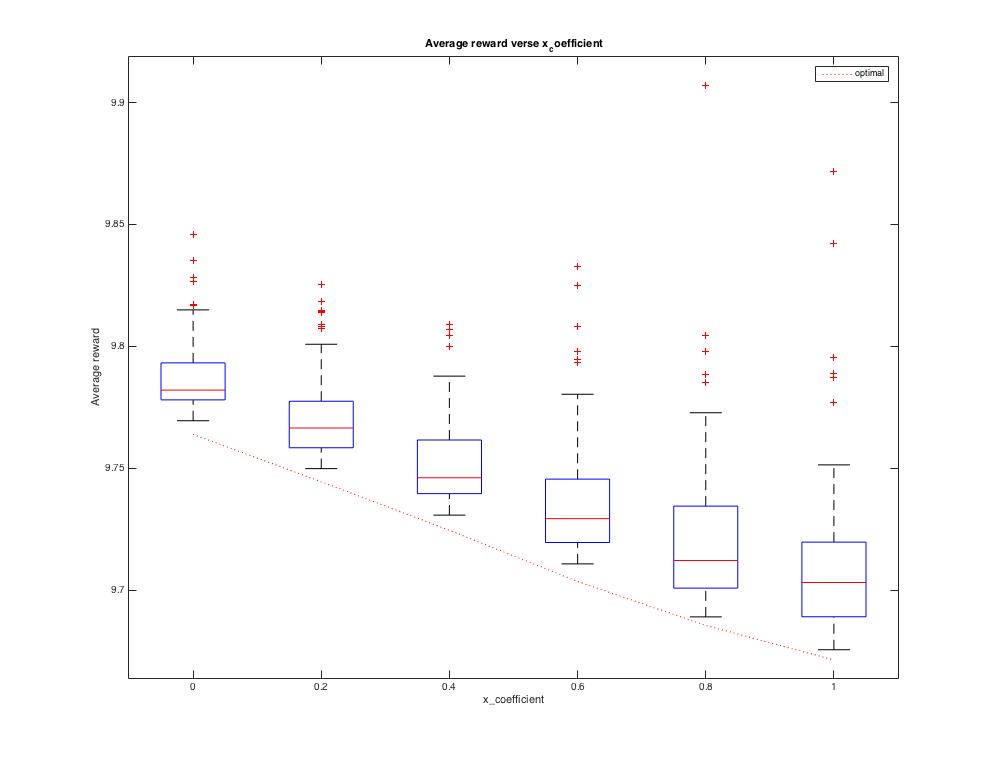
Where x\_co is coefficients we set, varying from 0, 0.2, .. to 1

Cost is generated by

Mu = [10 -1 0 0(tau) -0.25 -0.4 0]

R(t+1) = 10 – S(t,1) + A(t) x [-0.25S(t,1) – 0.4S(t,2)] + e(t)

Then the average cost according to different coefficients is shown in the boxplot

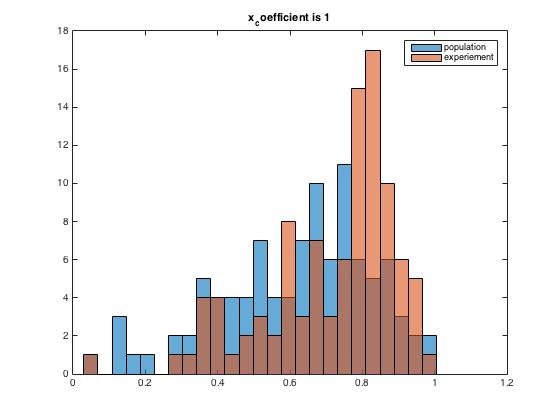
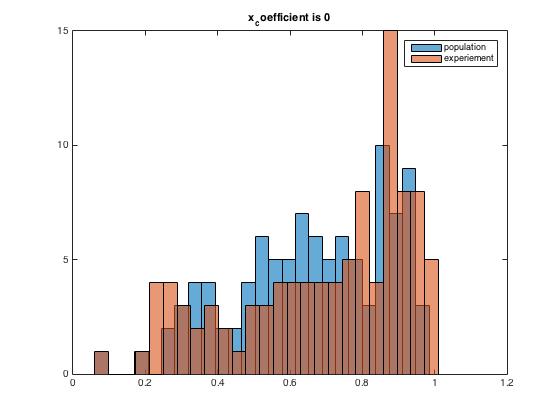


And when x = 0 and when x = 1, to compare the value of the probability

P(a = 1 |S) =

under population level and simulated experiments, I simulated 100 groups of contexts S: then use theta from optimal as well as theta from experiment.

When coefficient is 0, the probability values are almost same for population and experiment. But when coefficient is 1, the probability from experiment is overall higher than it from population.



Left is when coefficient is 0

Right is when coefficient is 1

I think this also meets our expectation. When tau = 0, there is no burden effect, but we have S2, which is a learning factor. When we increase the coefficient, S2 and Action are more positively related. So we expect to see a higher probability value.

Experiment results:

Result for when tau = 0 and context are iid:

This meets our expectation. (That is bandit algorithm work on iid context)

