Experiment 1

February 14, 2018

Simulation Details

Considered $K = 3$, $T = 5005$. Report statistics at $t = 1000, 3000, 5000$
The Bandit priors that were considered:

- Uniform: Draw the mean rewards for the arms from [0.25, 0.75]
- "HeavyTail": We took the mean rewards to be randomly drawn from Beta($\alpha = 0.6, \beta = 0.6$). With this distribution it was likely to have arms that were at the extremes (close to 1 and close to 0) but also some of the arms with intermediate value means.
- Needle-in-haystack
  1. Medium - 9 arms with mean 0.50, 1 arm with mean 0.55 (+ 0.05)
  2. High - 9 arms with mean 0.50, 1 arm with mean 0.70 (+ 0.20)

Algorithms considered:

1. ThompsonSampling with priors of $Beta(1, 1)$ for every arm.
2. DynamicGreedy with priors of $Beta(1, 1)$ for every arm
3. Bayesian Dynamic ϵ-greedy with priors of $Beta(1, 1)$ for every arm and $\epsilon = 0.05$

Agent Algorithms considered:

1. HardMax
2. HardMaxWithRandom
3. SoftMax

Memory Sizes

1. 10
2. 25
3. 100

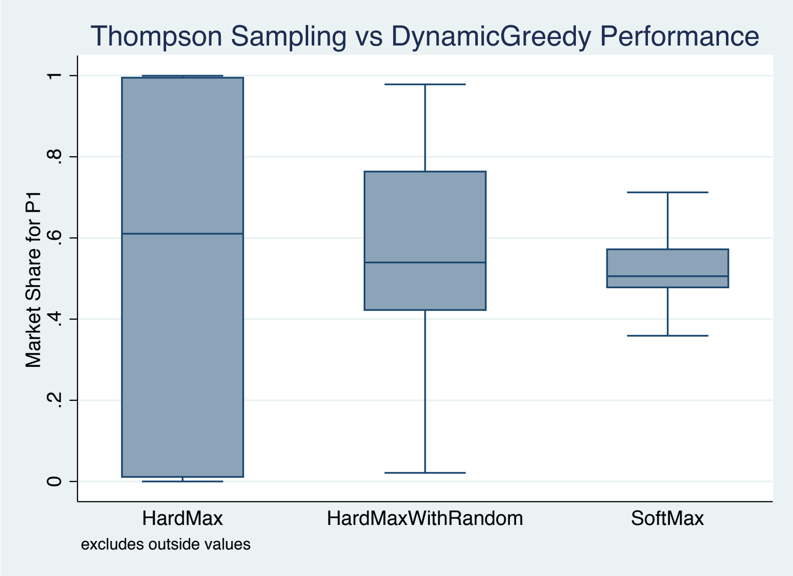
Simulation Procedure

```
for Each prior $p$ do
    for Each agent algorithm $agent_alg$ do
        for Each principal algorithm pair $principal_alg1$, $principal_alg2$ do
            for Each simulation $i$ do
                Generate true distribution from $p$ (except for needle-in- haystack, just use $p$ itself)
                Run simulation for T periods
            end for
        end for
    end for
end for
```

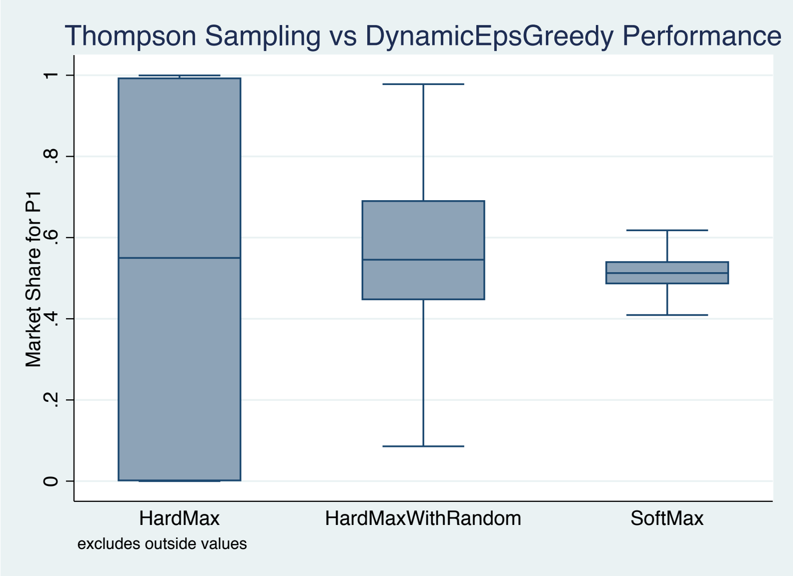
Results

One thing which is ambiguous to define is the regret value to use when a principal never gets chosen in a given simulation. When calculating any of the aggregate regret statistics we drop these simulations, but we do record how many rounds have an undefined regret.

First, we'll restrict focus to $t = 5000$ and look at the performance of ThompsonSampling. Note that the y axis here represents the market share that the ThompsonSampling principal gets. Performance of ThompsonSampling vs DynamicGreedy and DynamicEpsilonGreedy across all agent models, memory sizes, and priors:



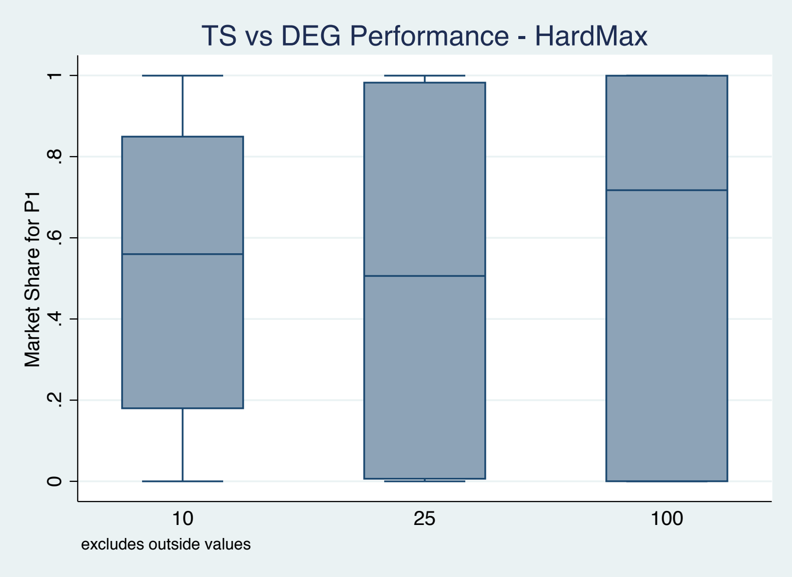
Thompson Sampling vs DynamicGreedy Performance

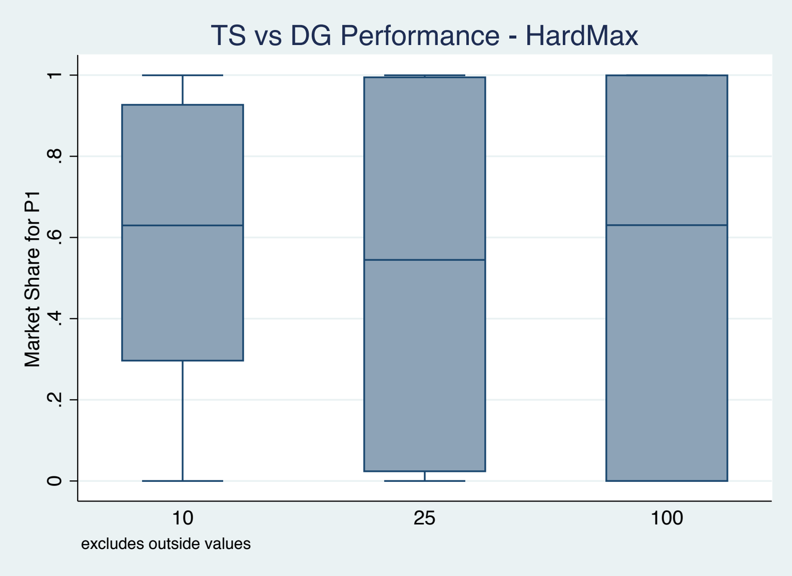


Thompson Sampling vs DynamicEpsGreedy Performance

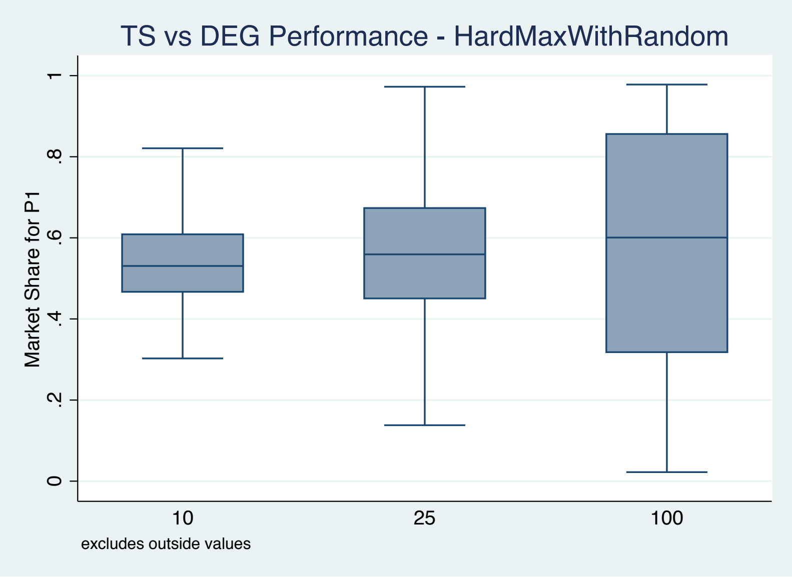
Now, looking across different memory sizes for each agent model (still using each prior):

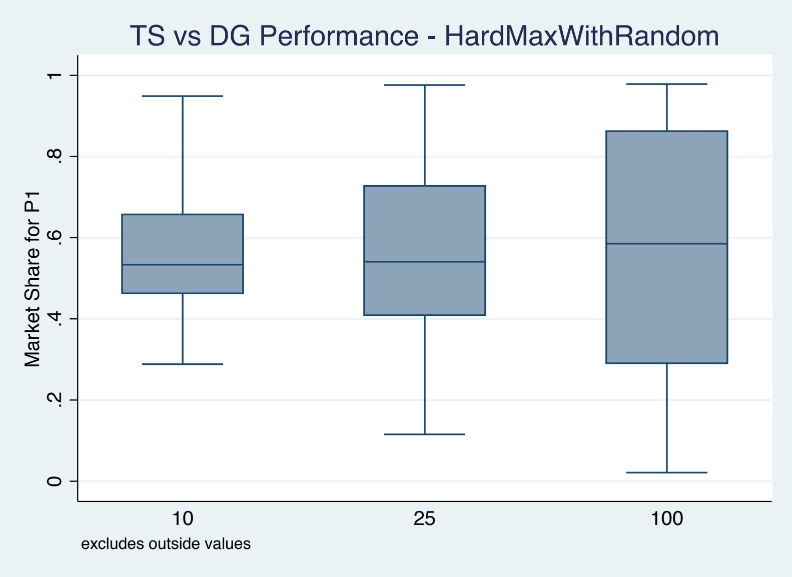
HardMax



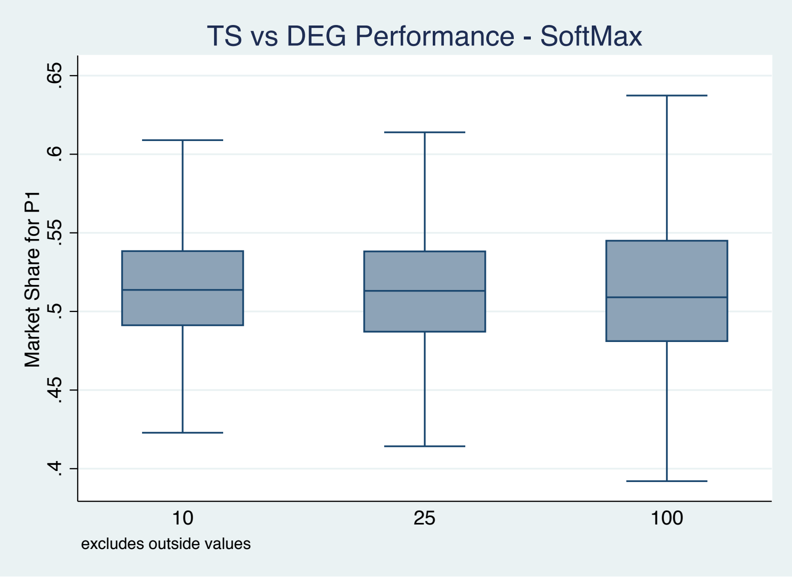


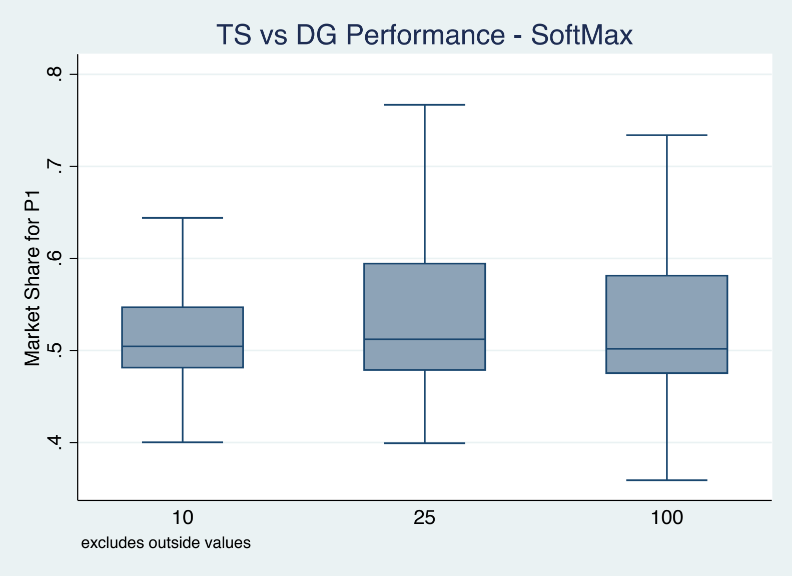
HardMaxWithRandom





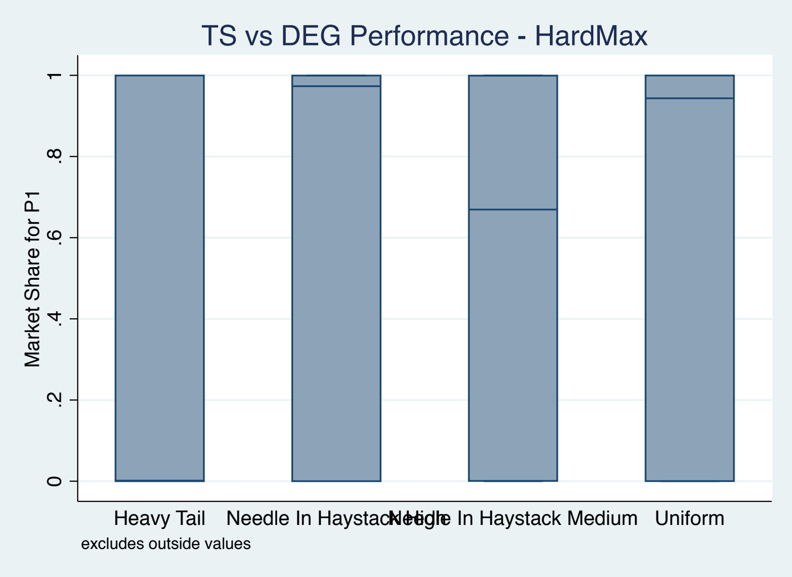
SoftMax

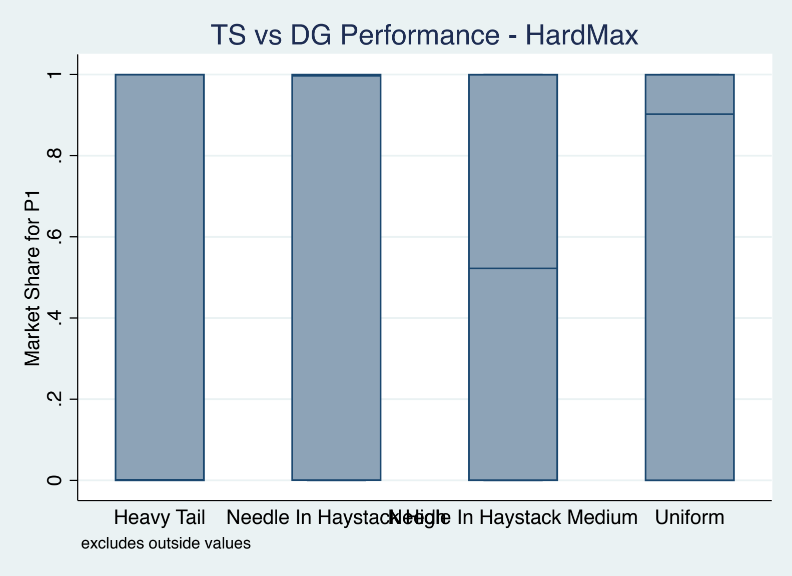

TS vs DEG Performance - SoftMax
Market Share for P1
.4
.45
.5
.55
.6
.65
10
25
100
excludes outside values


TS vs DG Performance - SoftMax
Market Share for P1
.4
.5
.6
.7
.8
10
25
100
excludes outside values

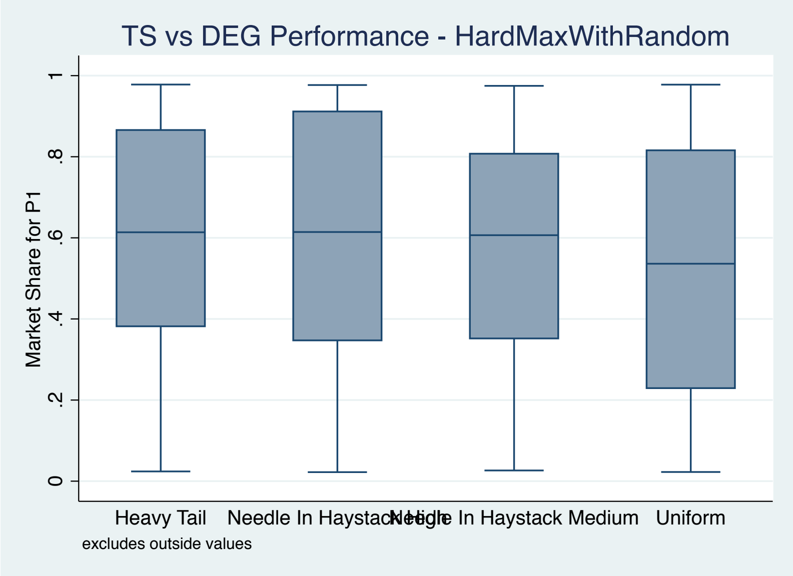
We'll now look fix memory size to be 100 and look at the performance across priors.

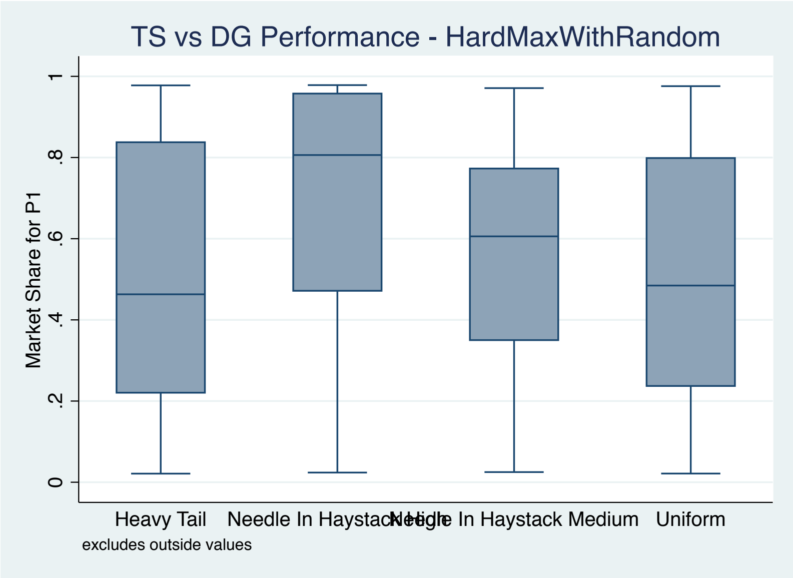
HardMax



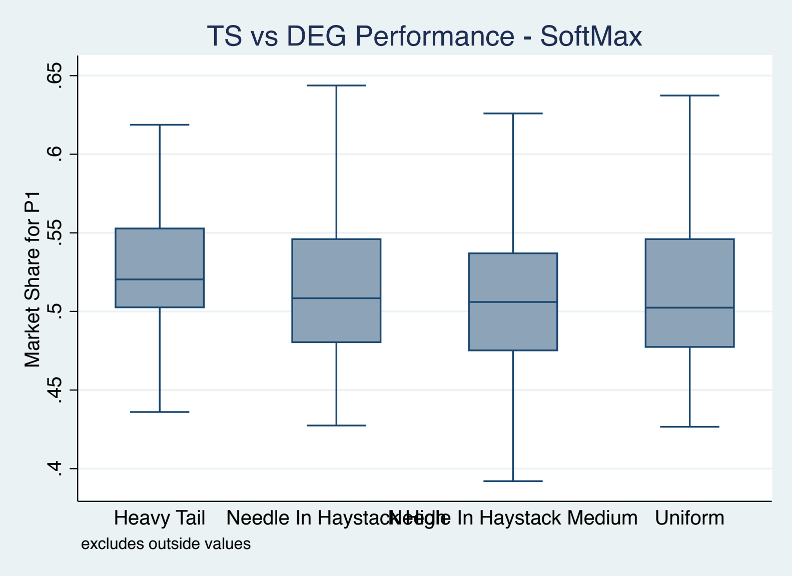


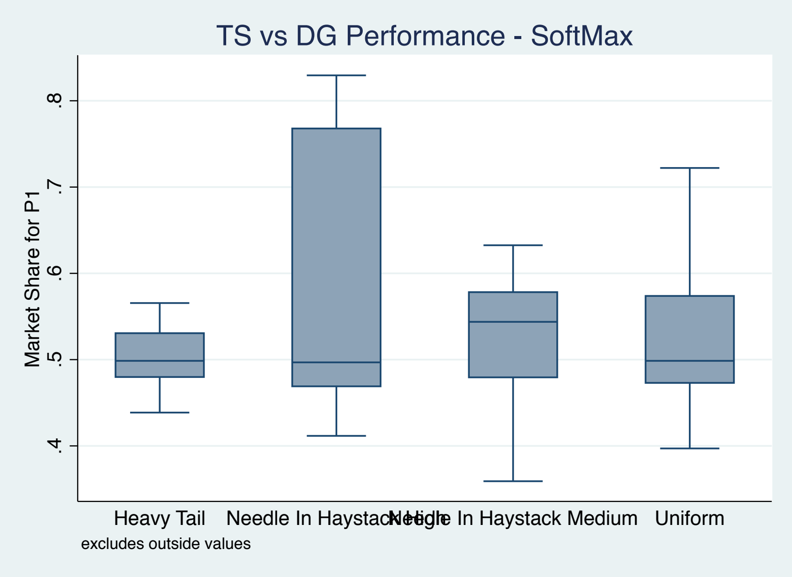
HardMaxWithRandom





SoftMax


TS vs DEG Performance - SoftMax
Market Share for P1
.4
.45
.5
.55
.6
.65
Heavy Tail
Needle In Haystack High
Needle In Haystack Medium
Uniform
excludes outside values


TS vs DG Performance - SoftMax
Market Share for P1
.4
.5
.6
.7
.8
Heavy Tail
Needle In Haystack High
Needle In Haystack Medium
Uniform
excludes outside values