Competing Bandits - Summary

July 2, 2018

Overview of Results

Priors:

1. Needle In a Haystack (x, K) = (K-1) arms with a mean of x and 1 arm with mean x + 0.2
2. Heavy Tail - means of arms randomly drawn from beta(0.6, 0.6)
3. Uniform - means of arms randomly drawn from [0.25, 0.75]
4. .5/.7 Prior - means of arms randomly drawn from $\{0.5, 0.7\}$

Question: What algorithms win when both firms start at the same time (simultaneous entry)? How does this vary across agent response functions?

1. Can we use any metrics of the bandit instances run in isolation that help us predict who wins in the competion game? A metric that seems to work well is the relative reputation plot that simply computes the proportion of instances when an algorithm a has at least as high of a reputation than algorithm b at fixed time step t and on the same bandit instance (with the same realizations).
   (a) For HM - fixing a warm start value t (number of free observations at the start), whichever algorithm has > 0.5 proportion of "victories" at time t according to the relative reputation plot should, on average, win more in competition
   (b) For HMR and SM, whichever algorithm has > 0.5 proportion of "victories" for sufficiently large t (i.e. when learning is more or less done) should win in competition. This is the case since the reputation curve of a given algorithm eventually converges and we simply need to ask which reputation curve is relatively higher at the convergence value.
   (c) **Question**: Does this always line up with looking at the mean reputation plots? In one case, Heavy Tail with $K = 3$, we see that the mean plots show $DEG > DG$, but our simulation results show that $DG > DEG$ in competition and $DG > DEG$ in the relative reputation plots. Relation to Dueling Algorithms paper (?). One intuition to this which plays out in the data is that DG either:
      i. identifies the best arm and thus beats or equals reputation of DEG since DEG explores randomly with ϵ probability.
      ii. does not identify the best arm and does substantially worse than DEG.

As a result, it will have a worse average reputation due to the cases where it identifies the wrong best arm but there may be enough cases where it does so that when only comparing relative performance in a given simulation it does better.

2. What algorithms win when fixing the HM response model?
   (a) We see that there is a low effective end of game[1], which is skewed with many instances having a median of 0, meaning that many of the games end simply from the choices made in the warm start
   (b) Many extreme shares - one principal takes most of the market
   (c) What algorithm wins depends on the warm start parameter (number of initial free observations).
   (d) For low warm start of 5 we have the following results
      i. $TS > DEG, DG$ for Needle In Haystack (0.5, 10), (0.7, 10)
      ii. 50/50 for Needle In Haystack (0.5, 3), (0.1, 10), (0.3, 10), Uniform (K = 3, 10), .5/.7 prior
      iii. $DEG, DG > TS$ for Heavy Tail (K = 3, 10)
      iv. $DG > DEG$ for Heavy Tail (K = 3)
   (e) For moderate warm start (20, 50)
      i. $DEG, DG > TS$ for Uniform (K = 10), Heavy Tail (K = 10)
      ii. $TS > DEG, DG$ for Needle In Haystack (0.5, 10), (0.7, 10)
   (f) For large warm start
      i. $TS > DEG, DG$ for everything except .5/.7 prior (why not for .5/.7 prior? Will run more simulations to see if it narrowly wins)
   (g) **Question**: Why do results vary by warm start value?
      i. On "harder" learning instances, learning takes longer and pure exploration has early reputation costs. Warm start needs to be sufficiently high to have recovered the reputation costs of exploration.
      ii. On "harder" learning instances, there is a benefit to exploration in the sense that it allows the principal to differentiate itself in reputation against an algorithm like dynamic greedy in the long run.
      iii. On "easier" learning instances, little reputation cost but also little benefit to deploying a smarter learning algorithm?
      iv. How to define "harder?" Use $H = \mathbb{E}[\sum\limits_{\mu_i < \mu^*} \frac{1}{\mu^* - \mu_i}]$, where μ^* is the mean of the best arm to define instance difficulty. On the "harder" instances we seem to see larger warm start values required in order for the better algorithm to win but not sure this explains everything.

3. For HMR:

[1] Effective End of Game (EEOG) is defined as the last round when the agent "switched" the firm she chose

    (a) "Better algorithm" wins with sufficiently large time horizon, where "better algorithm" is defined as the algorithm that has a larger proportion of wins in the relative reputation plot
    (b) One exception is the .5/.7 prior (where means of the arms are randomly drawn from 0.5, 0.7) where even this is 50/50
    (c) Moderate variance in the market shares

4. For SM

    (a) Qualitatively similar to HMR where better algorithm wins for sufficiently large time horizon
    (b) Results closer to 50/50 than HMR
    (c) Very low variance in the shares

Question: What happens if there is asymmetric entry so that one principal has an incumbency advantage?

1. If the incumbency advantage is sufficiently long, then TS is a dominant strategy across priors and agent models that we have tested out.
2. If the incumbency advantage is not sufficiently long, TS is no longer a dominant strategy
3. For HM, TS leads to one firm (the incumbent) dominating the market.
4. For HMR, TS leads to one firm dominating the market but not as dramatically compared to HM.
5. For SM, TS still wins but closer to 50/50. than HMR/HM.
6. For the entrant, it is ambiguous what is the best strategy

Question: Incumbency gives both an informational and a reputational advantage. If we artificially erase one upon entry of the entrant, does information or reputation play a bigger role as barriers to entry?

1. For HM erasing reputation hurts incumbent more (reputational advantage is more important than data advantage) (Why?)
2. For HMR/SM - erasing either hurts a bit, but erasing both hurts a lot implying that reputational and data advantage substitute for one another

Story

- Competition in learning environments - compete for users based on quality alone but quality of different actions is unknown and need to learn the quality of actions (modeled as an iid MAB). Learn by committing to a learning algorithm at the start of the world. Want the adoption of traditionally "better" algorithms since it is socially beneficial but when does this happen?

- We introduce a simple metric of reputation - agents want to maximize their reward and use a reputation score as a proxy for that. Reputation score is a sliding window average of the reward experienced by past agents that had selected this firm. Agents are myopic and non-strategic so they do not consider other factors such as the algorithm employed by the firms.

- **Problem**: Competition imposes incentive compatibility constraints on myopic consumers (i.e. firm needs to choose actions so that they actually get selected by users in competition). Start with consumers using the HardMax decision rule.

- Firms face a tradeoff between picking algorithms that maximize short-term reputation and those engage in pure exploration (which assists in maximizing long-term reputation). Consumers don't care about information gain from exploration but do care about reputation. Thus, algorithms that engage in pure exploration suffer short-term reputational costs and lose in the competition game to algorithms that maximize short-term reputation. Another aspect of this is that it depends on the hardness of the learning instance - the "harder" learning instances have larger (short-term) reputational costs but potentially larger (long-run) reputational advantage.

- How to incentivize firms better algorithms? Sufficient number of consumers whose incentive compatibility constraints are effectively removed. A few ways this can happen

  1. Sufficient number of periods as the monopolist in the market (incumbent experiment results)
  2. "Random agents" or probabilistic decision rules (HMR or SM response)
  3. "Warm start" or free agents at the start of the game

- Focusing on the monopoly case and HM decision rule, the firm accumulates both a reputational advantage and a data advantage. Which is more important in the competition game for serving as barriers to entry. We see that it is reputation (I have no good intuition for this yet).

- **Takeaway**: competition can stifle "innovation" if learning has reputational effects due to incentive constraints on consumers. Counterintuitively, monopoly can have benefit for innovation since it allows for exploration by not having to worry about incentive constraint of consumers. However, sometimes even a traditionally "better" algorithm according to mean regret cannot win in competition since relative reputation matters more.