Competing Bandits

Guy Aridor, Kevin Liu

October 28, 2017

In this project we hope to explore the use of bandit models in explaining product selection and innovation in firms. To begin our project, we will focus on a simulation project based on Mansour, Slivkins, and Wu (2017) and then proceed to investigating alternative specifications of the model. Thus, our project is will start as a simulation project and, depending on the time and success of the simulation, will consider exploring original research in this topic. Our intention is to proceed as follows:

1. We hope to reproduce the results of the aforementioned paper via simulation. The method by which we intend to proceed for this is to take the theoretical results regarding the optimal algorithm chosen conditional on the decision-rule of the agents and the assumptions on the prior as given and then to write code for the optimal bandit algorithms for the principals and the decision-rules of the agents. We hope to reproduce the inverted-U relationship between competition and innovation.

2. We will continue utilizing the same model, but investigate the effects of different parameterizations of the model. Namely, we will use the simulation code in the first part of the project to examine the robustness of the results for different distributions for the arms and priors. In doing so we hope to understand what the effects of distributional assumptions on the tradeoff between competition and innovation in this model.

3. After this we will attempt to both modify and extend the model to be focused on a particular setting where we interpret the principals as firms and the agents as consumers. We do this to investigate the motivating question of utilizing a bandit model to explain product selection and research and development between competing firms. This part of the project will be a mix of simulations and theoretical exercises:

   (a) The sequence of decisions is taken to be the same as the original model. However, we consider an alternative modelling assumption that breaks up the consumer and firm payoffs. Here we interpret the different arms as being different products (bundles of characteristics) offered by the firm and the firms' payoffs (reward) are interpreted as being the profitability of the arm. Where we deviate from the original model is to model consumers as having stochastic heterogenous utility for different products and not (exactly) linked to the stochastic reward from the arms. This modelling assumption comes from the fact that different arms will have different costs associated with them for the firm (the original model effectively assumes no cost) and thus the optimal arm for the consumer is not necessarily the optimal arm for the firm.

   (b) The sequence of events in the original model are a good model for online web services where a service determines in realtime which arm to pull. However, we will consider a model where a firm decides which product to produce for a fixed time period and then observes its reward over that time period (the number of agents it attracts) and then decides which product to produce (arm to pull) in the next time period. This kind of modelling decision is more in line with firms that are not online web services.

   (c) We will likely not have the time to undertake this potential extension of the model as it is will likely be quite difficult to tackle in a month, but we note it here since, if time permits, at least thinking about the intuition and rudimentary simulations for it would be useful. We consider an extension of the model where instead of committing to a MAB algorithm in the first period, the firms update their beliefs not only from the rewards they acrue from consumers who buy from them but also from their observations of the actions of the other firms. The intuition as to why this might be a reasonable extension is that the actions of firms reveal private information about their beliefs about the reward distributions and thus if the actions of the firm are observable then this is also information the other firm can use to update its belief. Thus, this kind of model is more in line with strategic experimentation models.