Topics in Game Theory: Learning, Experimentation, and Information

Outline for the lectures of Part 2
Pauli Murto
Aalto University and Helsinki GSE

March 23, 2022

1 Outline for the lectures of Part 2

In these three lectures, we will focus on the two-armed bandit model in continuous time. Two versions of that model, a Poisson bandit model and a Brownian bandit model have both been in extensive use in economics. First, we will analyze the experimentation problem of a single decision maker in both the Poisson and Brownian versions of that model. Next, we will intrdoduce strategic elements into that model in order to focus on informational externalities. Finally, we will look at some selected extensions/applications. The goal of these lectures is to give a basis for reading applied papers based on this framework. A very non-exhaustive list of references to such papers is given below.

1.1 Lecture 1 (24.3.)

In the first lecture, we will introduce the classical two-arm bandit problem in continuous time. To see clearly its connection to the models analyzed so far, I state here the discrete-time version of the model:

- Time periods $t = 0, 1, ...$
- Discounting per period $\delta = e^{-r\Delta}$, where Δ is the period length
- A decision maker chooses in each period an action $a_t \in \{\text{safe } (S), \text{risky } (R)\}$
- The risky arm has an unknown permanent state $\theta \in \{\text{good } (G), \text{ bad } (B)\}$ with prior $p_0 = \Pr(\theta = G) \in (0, 1)$.
- The decision maker maximizes the expected discounted stream of rewards:

$$\mathbb{E} \sum_{t=0,1,...} e^{-rt\Delta} \pi(a_t, \theta),$$

  where expectation is over θ and the chosen sequence $\{a_t\}$.

- The reward using the safe arm is constant $\pi\,(S,\theta_t) = s\cdot\Delta$ irrespective of state θ
- The reward for the risky arm is random, and we will focus on two alternative formulations:
- Model 1:
$$\pi\,(\,R,\theta) = \begin{cases} h \text{ with probability } \lambda_\theta\cdot\Delta, \\ 0 \text{ with probability } 1-\lambda_\theta\cdot\Delta, \end{cases}$$
where s, h, λ_G, λ_B are positive parameters such that $\lambda_B h < s < \lambda_G h$.
- Model 2:
$$\pi\,(\,R,\theta) = \mu_\theta\Delta + \sigma\sqrt{\Delta}Z,$$
where $\mu_B < s < \mu_G$, $\sigma > 0$, and Z is a standard normal random variable.
- In the continuous time limit $\Delta \to 0$, Model 1 is the "Poisson bandit" and Model 2 is the "Brownian bandit" model.
- We will do the analysis in the lecture (see references below for material)

1.1.1 Key references for lecture 1:

- For the Poisson bandit problem, see e.g. the cooperative solution in Keller and Rady (2010). A more mathematical treatment is in Presman (1990).
- For the Brownian problem, see e.g. the cooperative solution in Bolton and Harris (1999). A more general and mathematically rigorous treatment is in Cohen and Solan (2013)

1.2 Lecture 2 (29.3.)

In the second lecture, we will analyze a multi-player version of the two-armed bandit problem. In such a model of "strategic experimentation", N players face a choice problem analyzed in lecture one, and each player observes in real time the action choices and realized payoffs of each other. Hence, in addition to their own experimentation, the players also learn from the other players' experimentation, which creates an informational externality. We will discuss how this distorts the incentives to experiment and what this means in applications.

1.2.1 Key references for lecture 2:

The first paper that analyzed a model of strategic experimentation was Bolton and Harris (1999), who utilized the Brownian bandit model. Keller, Rady, and Cripps (2005) analyzed a "perfectly revealing news" version (i.e. one where $\lambda_B = 0$) of the Poisson bandit model and Keller and Rady (2010) extended that analysis to the more general Poisson bandit model. All these papers restrict the solution concept to be Markov Perfect Equilibrium. Hörner, Klein, and Rady (forthcoming) extend beyond that solution concept and show that better outcomes are achievalbe in "strongly symmetric equilibrium". Their analysis also extends to a more general Levy-bandit model that encompasses both Brownian and Poisson models as special cases.

1.3 Lecture 3 (31.3.)

In the third lecture, we will discuss some modifications/applications of the model of strategic experimentation. In the model analyzed in Lecture 2, the agents observe both the action choices and the action outcomes (rewards) of the other agents. One may ask what happens if agents' observations are more limited:

- What if agents only observe the action choices of each other, not rewards? Rosenberg, Solan, and Vieille (2007) and Murto and Välimäki (2011) analyze such models.

- What if agents only observe action outcomes but not action choices? Bonatti and Hörner (2011) is such a model.

1.4 Possible papers for student presentations

The two-armed bandit framework has been extensively utilized as an incredient in applications. Here are some examples of such papers. We don't have time to discuss these in the lecture, but they make suitable topics for student presentations:

- Papers utilizing Poisson bandit framework: Bergemann and Hege (2005), Bonatti and Hörner (2011), Che and Hörner (2018), Halac, Kartik and Liu (2017), Hörner and Samuelson (2013), Malueg and Tsutsui (1997), Strulovici (2010).

- Papers utilizing Brownian bandit framework: Bergemann and Välimäki (1997), Bergemann and Välimäki (2000), Bonatti (2011), Gul and Pesendorfer (2012), Laiho, Murto and Salmi (2022).

2 References

- Bergemann and Hege (2005), "The Financing of Innovation: Learning and Stopping", RAND Journal of Economics.

- Bergemann and Välimäki (1997), "Market diffusion with two-sided learning," RAND Journal of Economics.

- Bergemann and Välimäki (2000), "Experimentation in markets," Review of Economic Studies.

- Bolton and Harris (1999), "Strategic Experimentation," Econometrica, 67, 349–374.

- Bonatti (2011), “Menu pricing and learning,” American Economic Journal: Microeconomics.
- Bonatti and Hörner (2011), "Collaborating", American Economic Review.
- Che and Hörner (2018), "Recommender Systems as Mechanisms for Social Learning", Quarterly Journal of Economics.
- Cohen and Solan (2013), "Bandit Problems with Lévy Processes", Mathematics of Operations Research 38(1):92-107.
- Gul and Pesendorfer (2012), "The War of Information", Review of Economic Studies.
- Halac, Kartik and Liu (2017), "Contests for Experimentation", Journal of Political Economy.
- Hörner, Klein and Rady (forthcoming), "Overcoming Free-Riding in Bandit Games", Review of Economic Studies.
- Hörner and Samuelson (2013), "Incentives for experimenting agents", RAND Journal of Economics.
- Keller and Rady (2010), “Strategic Experimentation with Poisson Bandits,” Theoretical Economics, 5, 275–311.
- Keller, Rady and Cripps (2005), “Strategic Experimentation with Exponential Bandits,” Econometrica, 73, 39–68.
- Laiho, Murto and Salmi (2022), "Gradual Learning from Incremental Actions", working paper.
- Malueg and Tsutsui (1997), "Dynamic R&D Competition with Learning", RAND Journal of Economics.
- Murto and Välimäki (2011), "Learning and Information Aggregation in an Exit Game", Review of Economic Studies.
- Presman (1990), “Poisson version of the two-armed bandit problem with discounting”, Theory of Probability and Its Applications, 35, 307–317.
- Rosenberg, Solan and Vieille (2007), “Social Learning in One-Arm Bandit Problems,” Econometrica, 75, 1591—1611.
- Strulovici (2010), “Learning while voting: Determinants of collective experimentation.” Econometrica, 78, 933–971.