Strategic Experimentation

Lecture 1: Poisson Bandits
A Model of Rational Learning

People are uncertain about their environment

They learn from experience

Bayesian learning

- Agents possess a prior belief about some unknown parameter(s) of the environment
- They know the distribution of possible outcomes conditional on the parameter(s)
- They use Bayes’ rule to update their belief on the basis of what they observe
A Single-Agent Example

Optimal learning typically involves experimentation
– Sacrifice current rewards for better information
  which will generate better rewards in the future
– Exploitation versus exploration

Examples:
– Seller of a new good
– Farmer choosing between a traditional crop
  and a gene-modified one
– Researcher pursuing a new research agenda
A Single-Agent Example

Multi-period monopolist learning about the demand curve

No physical links between periods

- Rothschild (1974)
- Aghion, Bolton, Harris & Julien (1991)
- Keller & Rady (1999)

Optimal learning may be incomplete

Identical sellers may charge different prices in the long run because of different sequences of observations
Strategic Experimentation

Agents learn from the experiments of others, as well as from their own

Examples:

– Oligopolists in a new market
– Farmers with neighboring fields
– Researchers pursuing a common research agenda

With publicly observable actions and outcomes, there is an informational externality
Strategic Experimentation

  - Differentiated-goods duopoly

  - Sellers and buyers learning about a new product

  - Pure informational externality
  - Two-armed bandits in continuous time
Multi-Armed Bandit Problem

Stylized model of the exploitation-exploration trade-off
  – Agent decides repeatedly which slot machine to play
  – Each machine produces a random stream of payoffs
  – Uncertainty about the distribution of payoff streams

Sequential design of experiments

- Robbins (1952)
- Bellman (1956), Bradt, Johnson & Karlin (1956)
- Gittins and Jones (1974)
- Karatzas (1984), Presman (1990)
- Bergemann & Välimäki (2006)
Bolton & Harris (1999):
- Two-armed bandits in continuous time
- One arm is “safe”
- The other arm is either “good” or “bad” (Brownian motion with high or low drift)
- Type of risky arm identical across players
- Publicly observable actions and outcomes
- Pure informational externality
- Unique symmetric Markov perfect equilibrium
- Free-riding versus encouragement
Strategic Experimentation with Bandits

Keller, Rady & Cripps (2005), Keller & Rady (2010):

- Poisson process with high or low arrival rate
- Unique symmetric Markov perfect equilibrium
- Multiplicity of asymmetric Markov perfect equilibria
- No equilibria in cutoff strategies
- Inefficiency
- Encouragement effect
The Model

Introduction

The Model
- Setup
- Beliefs
- Strategies

The Single-Agent Problem

The Cooperative Problem

The Strategic Problem

Conclusion
Poisson Bandits

Two-armed bandit in continuous time

One arm is **safe** \((S)\)

- generates a known flow payoff \(s\)

Other arm is **risky** \((R)\)

- yields i.i.d. *lump-sums* of known mean \(h\) which arrive according to a Poisson process

  - if **good** \((\theta = 1)\), Poisson intensity is \(\lambda_1 (\equiv \text{flow payoff } \lambda_1 h)\)
  - if **bad** \((\theta = 0)\), Poisson intensity is \(\lambda_0 (\equiv \text{flow payoff } \lambda_0 h)\)

\[ s > 0 \text{ and } \lambda_1 > \lambda_0 \geq 0 \text{ known to players} \]

True value of \(\theta\) initially unknown to players

Assumption: \(\lambda_1 h > s > \lambda_0 h\)
Exponential Bandits

Keller, Rady & Cripps (2005):

- \( \lambda_0 = 0 \)
- First lump-sum resolves all uncertainty about \( \theta \)
- Arrives at an exponentially distributed random time
  ( \( \rightarrow \) “exponential bandit”)

Much simpler than the case \( \lambda_0 > 0 \) in Keller & Rady (2010):

- No encouragement effect
- Backward induction from single-agent cut-off
Actions and Payoffs

One unit of a perfectly divisible resource per unit of time

If fraction \( k_t \in [0, 1] \) is allocated to \( R \) on \([t, t + dt]\)

\[ S \text{ yields } (1 - k_t) s \, dt \]

\[ R \text{ pays a lump-sum with probability } k_t \lambda_\theta \, dt \]

Expected payoff increment conditional on \( \theta \):

\[ [(1 - k_t)s + k_t \lambda_\theta h] \, dt \]
Payoffs and Beliefs

\( p_t = \) posterior probability that \( \theta \) equals 1

\[
E_t[\lambda_\theta] = p_t \lambda_1 + (1 - p_t) \lambda_0 = \lambda(p_t)
\]

Expected payoff increment conditional on observations up to time \( t \):

\[
[(1 - k_t)s + k_t \lambda(p_t)h] \, dt
\]

Total payoff in per-period units:

\[
E \left[ \int_0^\infty r \, e^{-rt} \left[ (1 - k_t)s + k_t \lambda(p_t)h \right] \, dt \right]
\]
Common Posterior Beliefs

Each player has a replica two-armed bandit
  – same $\theta$
  – i.i.d. arrival times

Common prior

Observable actions and outcomes

Hence common posterior
Applying Bayes’ Rule

Let player \( n = 1, \ldots, N \) allocate \( k_{n,t} \) to \( R \) on \([t, t + dt]\)

Define

\[
K_t = \sum_{n=1}^{N} k_{n,t}
\]

Conditional on \( \theta \), the probability of none of the players receiving a lump-sum in \([t, t + dt]\) is

\[
\prod_{n=1}^{N} e^{-k_{n,t} \lambda_{\theta} dt} = \prod_{n=1}^{N} (1 - k_{n,t} \lambda_{\theta} dt) = 1 - K_t \lambda_{\theta} dt
\]
No News is Bad News

Given $p_t$ at time $t$ and no lump-sum in $[t, t + dt[$:

$$p_t + dp_t = \frac{p_t (1 - K_t \lambda_1 dt)}{(1 - p_t)(1 - K_t \lambda_0 dt) + p_t (1 - K_t \lambda_1 dt)}$$

or

$$dp_t = -K_t \Delta \lambda p_t (1 - p_t) dt$$

with $\Delta \lambda = \lambda_1 - \lambda_0$

When there is a lump-sum on any of the risky arms, the posterior belief jumps up from $p_{t-}$ to

$$p_t = j(p_{t-}) = \frac{\lambda_1 p_{t-}}{\lambda(p_{t-})}$$
Stationary Markov Strategies

With $p_t$ as the natural state variable:

$$k_t = k(p_t)$$

where $k : [0, 1] \rightarrow [0, 1]$ is left-continuous and piecewise Lipschitz-continuous

Call the strategy simple if $k(p) \in \{0, 1\}$ for all $p$
For some $\hat{p} \in [0, 1]$:

If $p > \hat{p}$, then $k(p) = 1$ (play $R$ exclusively)
If $p \leq \hat{p}$, then $k(p) = 0$ (play $S$ exclusively)

Example:

**Myopic** agent ignores information content of own actions (as if $r = \infty$)

Payoff $= \max \{ s, \lambda(p)h \}$

Cut-off:

$$p^m = \frac{s - \lambda_0 h}{\lambda_1 h - \lambda_0 h}$$
The Single-Agent Problem

- Principle of Optimality
- Bellman Equation
- Solution

Conclusion
 Principle of Optimality

\[ u(p) = \max_{k \in [0,1]} \left\{ r \left[ (1 - k)s + k \lambda(p) h \right] dt \right. \]

\[ \left. + e^{-rt} \mathbb{E} \left[ u(\tilde{p}) \mid p, k \right] \right\} \]

With probability \( k \lambda(p) dt \):

Lump-sum,

\[ u(\tilde{p}) = u(j(p)) \]

With probability \( 1 - k \lambda(p) dt \):

No lump-sum,

\[ u(\tilde{p}) = u(p) - k \Delta \lambda p(1 - p) u'(p) dt \]

Insert and simplify to get Bellman equation!
Bellman Equation

\[ u(p) = s + \max_{k \in [0,1]} k \{ b(p, u) - c(p) \} \]

with the expected short-term opportunity cost

\[ c(p) = s - \lambda(p) h \]

and the expected long-term learning benefit

\[ b(p, u) = \left[ \lambda(p) (u(j(p)) - u(p)) - \Delta \lambda p(1 - p)u'(p) \right] / r \]

(left-hand derivative!)

\[ \Rightarrow k = 1 \text{ or } k = 0 \text{ is optimal} \]
When $k = 1$ is optimal, $u$ satisfies a differential-difference equation (DDE) that admits an explicit solution.

Optimal cut-off $p_1^* < p^m$ determined by

- value matching: $u(p_1^*) = s$
- smooth pasting: $u'(p_1^*) = 0$
The Cooperative Problem
Bellman equation for $N$ players jointly maximising their \textit{average} payoff:

$$u(p) = s + \max_{K \in [0,N]} K \left\{ b(p, u) - c(p) / N \right\}$$

$\Rightarrow K = N$ or $K = 0$ is optimal

$\Rightarrow$ Optimal cut-off $p^*_N < p^*_1$, decreasing in $N$
The Strategic Problem

Introduction

The Model

The Single-Agent Problem

The Cooperative Problem

The Strategic Problem

- Best Responses
- General Results
- Symmetric MPE
- Asymmetric MPE
- Simple MPE
- More Asymmetry

Conclusion
Bellman equation for player $n$, given $K_{-n} = K - k_n$:

$$u_n(p) = s + K_{-n} b(p, u_n) + \max_{k_n \in [0,1]} k_n \{ b(p, u_n) - c(p) \}$$

Indifference diagonal:

$$D_{K_{-n}} = \{ (p, u) \in [0,1] \times \mathbb{R}_+: u = s + K_{-n} c(p) \}$$
General Results for Markov Perfect Equilibria

(1) In any MPE, each player obtains at least the single-agent optimum payoff

(2) All MPE are inefficient (free-riding)

(3) There is no MPE where all players use cut-off strategies

(4) In any MPE, at least one player experiments at some beliefs below $p_1^*$ (encouragement) if and only if $\lambda_0 > 0$
No MPE in Cut-Off Strategies

Player 1 must change action at $p_2$ as well as at $p_1$
Encouragement Effect for $\lambda_0 > 0$

Suppose all players play $S$ at all beliefs $p \leq p_1^*$

$$u_n(p_1^*) = V_1^*(p_1^*) = s, \quad u'_n(p_1^*) = (V_1^*)'(p_1^*) = 0$$

$$b(p_1^*, u_n) \leq c(p_1^*) = b(p_1^*, V_1^*)$$

$$\implies u_n(j(p_1^*)) \leq V_1^*(j(p_1^*))$$

$$\implies u_n(j(p_1^*)) = V_1^*(j(p_1^*))$$

$$u_n - V_1^* \text{ minimal at } j(p_1^*)$$

$$\implies u'_n(j(p_1^*)) \leq (V_1^*)'(j(p_1^*))$$

$$u_n(j^2(p_1^*)) \geq V_1^*(j^2(p_1^*))$$

$$\implies b(j(p_1^*), u_n) \geq b(j(p_1^*), V_1^*) > c(j(p_1^*))$$

So all players must use $R$ at the belief $j(p_1^*)$
Encouragement Effect for $\lambda_0 > 0$

Each player’s Bellman equation now yields

$$u_n(j(p_1^*)) = s + N b(j(p_1^*), u_n) - c(j(p_1^*))$$

$$\geq s + N b(j(p_1^*), V_1^*) - c(j(p_1^*))$$

$$> s + b(j(p_1^*), V_1^*) - c(j(p_1^*))$$

$$= V_1^*(j(p_1^*))$$

— a contradiction with $u_n(j(p_1^*)) = V_1^*(j(p_1^*))$
Encouragement effect rests on two conditions

- Experimentation by a pioneer must increase the likelihood that others will return to the risky arm
- These future experiments must be valuable to the pioneer

The second condition is not met for \( \lambda_0 = 0 \)

So for \( \lambda_0 = 0 \), experimentation in any MPE stops at \( p_1^* \)
Symmetric Equilibrium

No symmetric MPE in pure strategies

Indifference between $R$ and $S$ requires

$$b(p, u) = c(p)$$

— DDE with closed-form solution

Bellman equation then implies

$$k(p) = \frac{u(p) - s}{(N - 1)c(p)}$$
Symmetric MPE

Construction and uniqueness by elementary methods

Smooth pasting at $\tilde{p}_N$

Comparative statics as $N \uparrow$: common payoffs $\uparrow$, $\tilde{p}_N \downarrow$, $p_N^\uparrow$
Symmetric MPE

Interior allocation between cut-offs $\tilde{p}_N < p_N^\dagger$

Inefficiency: $\tilde{p}_N > p_N^*$ (not reached in finite time)

Encouragement effect: $\tilde{p}_N < p_1^*$ if $\lambda_0 > 0$
Asymmetric Equilibria

No asymmetric MPE demonstrated in Bolton & Harris (1999)

Keller, Rady & Cripps (2005):

- $\lambda_0 = 0$ (no encouragement effect)
- variety of asymmetric MPE in simple strategies
- complete classification for $N = 2$
- most inequitable MPE for arbitrary $N$
- higher average payoff than symmetric MPE

Keller, Rady & Cripps (2010):

- $\lambda_0 > 0$ (encouragement effect)
- construction of a particular asymmetric MPE
- Pareto improvement over symmetric MPE
The Most Inequitable MPE for $\lambda_0 = 0$ and $N = 2$

Average payoff higher than in symmetric MPE

Average payoff increases as burden of experimentation is shared more equitably
Constructing Asymmetric MPE for $\lambda_0 > 0$

Two ideas:

- Give all players identical continuation values after a success on a risky arm
- Let players alternate between the roles of experimenter and free-rider before all experimentation stops

Construction below for $N = 2$ generalises to arbitrary $N$
Asymmetric MPE for $\lambda_0 > 0$ and $N = 2$

- On $[p^b, 1]$ both players play symmetrically.
- On $[p^b, p^\#]$ players take turns playing $R$.
- On $[0, p^b]$ both players play $S$.

Strictly higher average payoff on $[p^b, 1]$ than in the symmetric equilibrium.

Pareto improvement for sufficiently frequent turns on $[p^b, p^\#]$.
Why is Taking Turns More Efficient?

Under symmetry:

- A player who deviates to the safe action slows down the gradual slide of beliefs towards more pessimism
- This makes the other players experiment more than they would on the equilibrium path

In the alternation phase of an asymmetric MPE:

- A deviation from the risky to the safe action freezes the belief in its current state
- This delays the time at which another player takes over
Keller, Rady & Cripps (2005):

- $\lambda_0 = 0$
- Symmetric MPE yields uniformly lowest average payoffs
- For $N = 2$, the limits as $\lambda_0 \downarrow 0$ of the MPE just constructed yield uniformly highest average payoffs
- For $N = 2$, the most inequitable MPE yield uniformly extremal individual payoffs; these are in simple strategies

Do such MPE exist for $\lambda_0 > 0$?

Consider simple MPE for $N = 2$ and small $\lambda_0 > 0$
Large Jumps

For $\lambda_0 > 0$ small enough:

- $j(p_2^*) \geq p^m$
- Everybody plays $R$ after a success
- We can again give the players common continuation values after a success on a risky arm

Precise condition:

$$\lambda_0 \left( \frac{\lambda_0}{\lambda_1} \right)^{\frac{\lambda_0}{\Delta \lambda}} \leq \frac{r}{2}$$

Sufficient condition:

$$\lambda_0 \leq \frac{r}{2}$$
A Simple MPE for $N = 2$

- on $[\hat{p}, 1]$ both players play $R$
- on $[p_s, \hat{p}]$ player 1 plays $S$, player 2 plays $R$
- on $[\bar{p}, p_s]$ player 1 plays $R$, player 2 plays $S$
- on $[0, \bar{p}]$ both players play $S$

Lower average payoff than in the previous MPE

- Monotonicity? ‘Anticipation’
Rewarding the Last Experimenter

- Maintain assumption of ‘large jumps’
- Relax payoff symmetry above $D_1$
  - Give player 1 higher payoff at $j(\bar{p})$
  - Two effects:
    - Lower intensity just to the right of $\hat{p}$
    - Player 1 is willing to play $R$ left of $\bar{p}$
  - Average payoff
    - decreases at high beliefs
    - increases around $\bar{p}$
Consequences:

- Can construct equilibria with progressively increasing payoff asymmetry, approaching the most inequitable MPE for $\lambda_0 = 0$
- There exist no uniformly best or worst MPE when $\lambda_0 > 0$
Rewarding the Last Experimenter

- Fix \((\hat{p}_2, u_1)\) above \(D_1\) find \(\hat{p}_1\) and \(\bar{p}\) and \(p_s\)
- Fill in \(u_2\) above \(D_1\) and just below
- Fill in \(u_2\) above \(\bar{p}\) ... does it join up?
- Adjust \((\hat{p}_2, u_1)\) if necessary

Introduction
The Model
The Single-Agent Problem
The Cooperative Problem
The Strategic Problem
  - Best Responses
  - General Results
  - Symmetric MPE
  - Asymmetric MPE
  - Simple MPE
  - More Asymmetry
Conclusion
Conclusion
Concluding Remarks

- **Poisson bandits**
  - Alternative to Brownian setup
  - News comes in ‘lumps’

- **Asymmetric equilibria with encouragement**
  - Construction by elementary methods
  - Taking turns is superior to ‘mixing’
  - No uniformly ‘best’ or ‘worst’ MPE unless $\lambda_0 = 0$
Concluding Remarks

- **What else?**
  - Ongoing work on ‘bad news’ scenario
  - Smooth pasting does not hold
  - Characterization of symmetric MPE
  - Viscosity solutions of the Bellman equation

- **What next?**
  - Introduce negative correlation of type of risky arm across players
  - Consider non-Markovian equilibria