Strongly Symmetric Equilibria in Bandit Games

Johannes Hörner (Yale University)
Nicolas Klein (Université de Montréal)
Sven Rady (University of Bonn)

Workshop on Stochastic Games, Singapore, Nov. 30–Dec. 4, 2015

Strategic Experimentation

Strategic Experimentation

Agents are uncertain about their environment.

They learn from experience in a Bayesian fashion.

Strategic Experimentation

Agents are uncertain about their environment.

They learn from experience in a Bayesian fashion.

Optimal learning typically involves experimentation (Sacrifice of current rewards for better information).

Strategic Experimentation

Agents are uncertain about their environment.

They learn from experience in a Bayesian fashion.

Optimal learning typically involves experimentation (Sacrifice of current rewards for better information).

Strategic Experimentation: Agents learn from the experiments *of others*, as well as from their own.

Strategic Experimentation

Agents are uncertain about their environment.

They learn from experience in a Bayesian fashion.

Optimal learning typically involves experimentation (Sacrifice of current rewards for better information).

Strategic Experimentation: Agents learn from the experiments *of others*, as well as from their own.

Literature thus far (Bolton & Harris, 1999; Keller, Rady, Cripps, 2005; Keller & Rady, 2010):

– **Markov** perfect equilibria;

– inefficiently low levels of experimentation because of free-riding (positive informational externality).

This Paper

This Paper

Explore non-Markovian behaviour:

This Paper

Explore non-Markovian behaviour:

– Freeze actions for a small length of time ($\rightarrow$ stochastic game in discrete time)

This Paper

Explore non-Markovian behaviour:

- Freeze actions for a small length of time ($\rightarrow$ stochastic game in discrete time)
- Construct strongly symmetric perfect Bayesian equilibria

This Paper

Explore non-Markovian behaviour:

– Freeze actions for a small length of time
($\rightarrow$ stochastic game in discrete time)

– Construct strongly symmetric perfect Bayesian equilibria

– Use a recursive approach
(Abreu 1986, 1988, Cronshaw & Luenberger 1994)

This Paper

Explore non-Markovian behaviour:

- Freeze actions for a small length of time
  ($\rightarrow$ stochastic game in discrete time)
- Construct strongly symmetric perfect Bayesian equilibria
- Use a recursive approach
  (Abreu 1986, 1988, Cronshaw & Luenberger 1994)
- Consider the limit of vanishing "inertia"

This Paper

Explore non-Markovian behaviour:

- Freeze actions for a small length of time ($\rightarrow$ stochastic game in discrete time)
- Construct strongly symmetric perfect Bayesian equilibria
- Use a recursive approach (Abreu 1986, 1988, Cronshaw & Luenberger 1994)
- Consider the limit of vanishing “inertia”
- Show that the best (worst) PBE is strongly symmetric

This Paper

Explore non-Markovian behaviour:

- Freeze actions for a small length of time ($\rightarrow$ stochastic game in discrete time)
- Construct strongly symmetric perfect Bayesian equilibria
- Use a recursive approach (Abreu 1986, 1988, Cronshaw & Luenberger 1994)
- Consider the limit of vanishing "inertia"
- Show that the best (worst) PBE is strongly symmetric

How close to efficiency can we get in the continuous-time limit?

Model

Poisson Bandits

Poisson Bandits

N players; two-armed bandits in continuous time.

Poisson Bandits

N players; two-armed bandits in continuous time.

One arm is **safe** (S),
generates a known flow payoff s.

Poisson Bandits

N players; two-armed bandits 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 λ_1 ($\equiv$ *flow* payoff $\lambda_1 h$);
if **bad** ($\theta = 0$), Poisson intensity is λ_0 ($\equiv$ *flow* payoff $\lambda_0 h$).

Poisson Bandits

N players; two-armed bandits 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 λ_1 ($\equiv$ *flow* payoff $\lambda_1 h$);
if **bad** ($\theta = 0$), Poisson intensity is λ_0 ($\equiv$ *flow* payoff $\lambda_0 h$).

$s > 0$ and $\lambda_1 > \lambda_0 \geq 0$ known to players.

True value of θ initially unknown to players.

Poisson Bandits

N players; two-armed bandits 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 λ_1 ($\equiv$ *flow* payoff $\lambda_1 h$);
if **bad** ($\theta = 0$), Poisson intensity is λ_0 ($\equiv$ *flow* payoff $\lambda_0 h$).

$s > 0$ and $\lambda_1 > \lambda_0 \geq 0$ known to players.

True value of θ initially unknown to players.

Assumption: $\lambda_1 h > s > \lambda_0 h$.

Beliefs

Players can adjust their actions at $t = 0, \Delta, 2\Delta, 3\Delta, \ldots$.

Beliefs

Players can adjust their actions at $t = 0, \Delta, 2\Delta, 3\Delta, \ldots$.

Each player has a replica two-armed bandit:

– same θ;

– independent Poisson processes.

Beliefs

Players can adjust their actions at $t = 0, \Delta, 2\Delta, 3\Delta, \ldots$.

Each player has a replica two-armed bandit:

– same θ;

– independent Poisson processes.

Common prior p_0

Observable actions and outcomes

Hence common posterior p_t (from Bayes' Rule)

Strategies

Strategies

For $t = 0, \Delta, 2\Delta, \ldots$, let H_t be the set of all **histories**

$$\left((k_{n,0})_{n=1}^N, (j_{n,\Delta})_{n=1}^N, \ldots, (k_{n,t-\Delta})_{n=1}^N, (j_{n,t})_{n=1}^N \right)$$

such that $k_{n,\tau} = 0 \Rightarrow j_{n,\tau+\Delta} = 0$.

Strategies

For $t = 0, \Delta, 2\Delta, \ldots$, let H_t be the set of all **histories**

$$\left((k_{n,0})_{n=1}^N, (j_{n,\Delta})_{n=1}^N, \ldots, (k_{n,t-\Delta})_{n=1}^N, (j_{n,t})_{n=1}^N \right)$$

such that $k_{n,\tau} = 0 \Rightarrow j_{n,\tau+\Delta} = 0$.

Each history h_t generates a unique sequence of beliefs $(p_0, p_\Delta, \ldots, p_{t-\Delta}, p_t)$.

Strategies

For $t = 0, \Delta, 2\Delta, \ldots$, let H_t be the set of all **histories**

$$\left((k_{n,0})_{n=1}^{N}, (j_{n,\Delta})_{n=1}^{N}, \ldots, (k_{n,t-\Delta})_{n=1}^{N}, (j_{n,t})_{n=1}^{N}\right)$$

such that $k_{n,\tau} = 0 \Rightarrow j_{n,\tau+\Delta} = 0$.

Each history h_t generates a unique sequence of beliefs $(p_0, p_\Delta, \ldots, p_{t-\Delta}, p_t)$.

A **strategy** is a sequence $\{k_t\}_{t=0,\Delta,2\Delta,\ldots}$ of measurable mappings

$$k_t : H_t \to \{0, 1\}$$

specifying an action $k_t(h_t)$ for each history $h_t \in H_t$.

Markov Strategies

Markov Strategies

A player's strategy $\{k_t\}_{t=\Delta,2\Delta,\ldots}$ is a **Markov strategy** if for all t

$$k_t(h_t) = \kappa(p_t)$$

where

- $\kappa\colon [0,1] \to \{0,1\}$ is measurable
- p_t is the posterior belief at the end of history h_t

Equilibria

Definition

Definition

A **strongly symmetric equilibrium (SSE)** is a perfect Bayesian equilibrium where

$$k_{1,t}(h_t) = k_{2,t}(h_t) = \ldots = k_{N,t}(h_t)$$

for all $t = 0, \Delta, 2\Delta, \ldots$ and all histories $h_t \in H_t$.

Payoff Functions

Payoff Functions

With each SSE we can associate a measurable **equilibrium payoff function**

$$w : [0, 1] \to [s, \lambda_1 h].$$

Payoff Functions

With each SSE we can associate a measurable **equilibrium payoff function**

$$w : [0,1] \to [s, \lambda_1 h].$$

For given $\Delta > 0$, the set of equilibrium payoff functions has

- a pointwise supremum

$$\overline{W}^{\Delta} : [0,1] \to [s, \lambda_1 h],$$

- a pointwise infimum

$$\underline{W}^{\Delta} : [0,1] \to [s, \lambda_1 h].$$

A Lower Bound on Equilibrium Payoffs

A Lower Bound on Equilibrium Payoffs

Note that

$$\underline{W}^{\Delta} \geq W_1^{\Delta},$$

– the single-agent value function.

A Lower Bound on Equilibrium Payoffs

Note that

$$\underline{W}^{\Delta} \geq W_1^{\Delta},$$

– the single-agent value function.

For $\Delta \to 0$, we have uniform convergence

$$W_1^{\Delta} \to V_1^*,$$

with an explicit representation for the limit function.

An Upper Bound on Equilibrium Payoffs

An Upper Bound on Equilibrium Payoffs

Let

$$\tilde{p}^{\Delta} = \inf\left\{p\colon \overline{W}^{\Delta}(p) > s\right\}$$

and

$$\tilde{p} = \liminf_{\Delta\to 0} \tilde{p}^{\Delta}$$

$\tilde{p} \geq p_N^*$ = efficient cut-off in continuous time

An Upper Bound on Equilibrium Payoffs

An Upper Bound on Equilibrium Payoffs

For any fixed Δ, consider the problem of maximizing the players' average payoff subject to

- symmetry of actions after all histories
- no use of R at beliefs $p < \tilde{p}$

Write $\widetilde{W}^{\Delta}$ for the corresponding value function

An Upper Bound on Equilibrium Payoffs

For any fixed Δ, consider the problem of maximizing the players' average payoff subject to

- symmetry of actions after all histories
- no use of R at beliefs $p < \tilde{p}$

Write $\widetilde{W}^{\Delta}$ for the corresponding value function

Then, there exists a $\bar{\Delta} > 0$ s.t. for $\Delta < \bar{\Delta}$:

$$\overline{W}^{\Delta} \leq \widetilde{W}^{\Delta}$$

An Upper Bound on Equilibrium Payoffs

For any fixed Δ, consider the problem of maximizing the players' average payoff subject to

- symmetry of actions after all histories
- no use of R at beliefs $p < \tilde{p}$

Write $\widetilde{W}^{\Delta}$ for the corresponding value function

Then, there exists a $\bar{\Delta} > 0$ s.t. for $\Delta < \bar{\Delta}$:

$$\overline{W}^{\Delta} \leq \widetilde{W}^{\Delta}$$

For $\Delta \to 0$, we have uniform convergence

$$\widetilde{W}^{\Delta} \to V_N(\cdot; \tilde{p})$$

again with an explicit representation for the limit function.

A Bound on Experimentation

A Bound on Experimentation

For a sequence of Δ's converging to 0 with $\tilde{p}^\Delta \to \tilde{p}$, choose $p^\Delta > \tilde{p}^\Delta$ with the following property:

> If the players start at the belief p^Δ, and $N-1$ of them use R for Δ units of time without success, then the posterior belief ends up below $\tilde{p}^\Delta$.

A Bound on Experimentation

Playing R at p^Δ yields at most

$$(1-\delta)\lambda(p^\Delta)h + \delta \mathsf{E}^\Delta\left[\widetilde{W}^\Delta | N, p^\Delta\right]$$

A Bound on Experimentation

Playing R at p^Δ yields at most

$$(1-\delta)\lambda(p^\Delta)h + \delta \mathsf{E}^\Delta\left[\widetilde{W}^\Delta | N, p^\Delta\right]$$

$$= \quad r\Delta\lambda(p^\Delta)h + (1-r\Delta)\Big\{(1-N\lambda(p^\Delta)\Delta)s + N\lambda(p^\Delta)\Delta\,\widetilde{W}^\Delta\Big(\frac{p^\Delta\lambda_1 e^{-\lambda_1\Delta K}}{p^\Delta\lambda_1 e^{-\lambda_1\Delta K}+(1-p^\Delta)\lambda_0 e^{-\lambda_0\Delta K}}\Big)\Big\} + o(\Delta)$$

A Bound on Experimentation

Playing R at p^Δ yields at most

$$(1-\delta)\lambda(p^\Delta)h + \delta \mathsf{E}^\Delta\left[\widetilde{W}^\Delta | N, p^\Delta\right]$$

$$= \quad r\Delta\lambda(p^\Delta)h + (1-r\Delta)\Big\{(1-N\lambda(p^\Delta)\Delta)s$$

$$+ N\lambda(p^\Delta)\Delta\,\widetilde{W}^\Delta\Big(\frac{p^\Delta\lambda_1 e^{-\lambda_1\Delta K}}{p^\Delta\lambda_1 e^{-\lambda_1\Delta K}+(1-p^\Delta)\lambda_0 e^{-\lambda_0\Delta K}}\Big)\Big\} + o(\Delta)$$

$$= \quad s + \Big\{r[\lambda(\tilde{p})h - s] + N\lambda(\tilde{p})[V_N(j(\tilde{p});\tilde{p}) - s]\Big\}\Delta + o(\Delta)$$

A Bound on Experimentation

Playing R at p^Δ yields at most

$$(1-\delta)\lambda(p^\Delta)h + \delta \mathsf{E}^\Delta\left[\widetilde{W}^\Delta | N, p^\Delta\right]$$

$$= \quad r\Delta\lambda(p^\Delta)h + (1-r\Delta)\left\{(1-N\lambda(p^\Delta)\Delta)s + N\lambda(p^\Delta)\Delta\,\widetilde{W}^\Delta\left(\frac{p^\Delta\lambda_1 e^{-\lambda_1\Delta K}}{p^\Delta\lambda_1 e^{-\lambda_1\Delta K}+(1-p^\Delta)\lambda_0 e^{-\lambda_0\Delta K}}\right)\right\} + o(\Delta)$$

$$= \quad s + \left\{r[\lambda(\tilde{p})h - s] + N\lambda(\tilde{p})[V_N(j(\tilde{p});\tilde{p}) - s]\right\}\Delta + o(\Delta)$$

Playing S yields at least

$$(1-\delta)s + \delta \mathsf{E}^\Delta\left[W_1^\Delta | N-1, p^\Delta\right]$$

A Bound on Experimentation

Playing R at p^Δ yields at most

$$(1-\delta)\lambda(p^\Delta)h + \delta \mathsf{E}^\Delta\left[\widetilde{W}^\Delta | N, p^\Delta\right]$$

$$= \quad r\Delta\lambda(p^\Delta)h + (1-r\Delta)\Big\{(1-N\lambda(p^\Delta)\Delta)s$$

$$+ N\lambda(p^\Delta)\Delta\,\widetilde{W}^\Delta\left(\frac{p^\Delta\lambda_1 e^{-\lambda_1\Delta K}}{p^\Delta\lambda_1 e^{-\lambda_1\Delta K}+(1-p^\Delta)\lambda_0 e^{-\lambda_0\Delta K}}\right)\Big\} + o(\Delta)$$

$$= \quad s + \Big\{r[\lambda(\tilde{p})h - s] + N\lambda(\tilde{p})[V_N(j(\tilde{p});\tilde{p}) - s]\Big\}\Delta + o(\Delta)$$

Playing S yields at least

$$(1-\delta)s + \delta \mathsf{E}^\Delta\left[W_1^\Delta | N-1, p^\Delta\right]$$

$$= \quad s + \Big\{(N-1)\lambda(\tilde{p})[V_1^*(j(\tilde{p})) - s]\Big\}\Delta + o(\Delta)$$

Trade-Off at p^Δ

Incentive compatibility at p^Δ requires

$$r(s-\lambda(\tilde{p})h) \leq \lambda(\tilde{p})\left[NV_{N,\tilde{p}}(j(\tilde{p})) - (N-1)V_1^*(j(\tilde{p})) - s\right],$$

Trade-Off at p^{Δ}

Incentive compatibility at p^{Δ} requires

$$r(s-\lambda(\tilde{p})h) \leq \lambda(\tilde{p})\left[NV_{N,\tilde{p}}(j(\tilde{p})) - (N-1)V_1^*(j(\tilde{p})) - s\right],$$

i.e.

$$\tilde{p} \geq \hat{p},$$

where $\hat{p}$ is the unique belief in $[p_N^*, p_1^*]$ making this condition bind.

Trade-Off at p^Δ

Incentive compatibility at p^Δ requires

$$r(s-\lambda(\tilde{p})h) \leq \lambda(\tilde{p})\left[NV_{N,\tilde{p}}(j(\tilde{p})) - (N-1)V_1^*(j(\tilde{p})) - s\right],$$

i.e.

$$\tilde{p} \geq \hat{p},$$

where $\hat{p}$ is the unique belief in $[p_N^*, p_1^*]$ making this condition bind.

$\hat{p} = p_N^*$ if and only if $j(p_N^*) \leq p_1^*$ (i.e., λ_0 close to λ_1);
$\hat{p} = p_1^*$ if and only if $\lambda_0 = 0$.

Construction of Equilibria

Assume $\lambda_0 > 0$ from now on so that $\hat{p} < p_1^*$

Want to establish that $\tilde{p} = \hat{p}$

Construct equilibria for small Δ that achieve payoffs arbitrarily close to $V_N(\cdot;\hat{p})$ as $\Delta \to 0$

Construction of Equilibria

Construction of Equilibria

Two-state automaton with public randomization

Construction of Equilibria

Two-state automaton with public randomization

Normal state:

- Common action $\overline{\kappa}(p)$ (independent of Δ)
- Go to punishment state after unilateral deviations
- Otherwise remain in normal state

Construction of Equilibria

Two-state automaton with public randomization

Normal state:

- Common action $\overline{\kappa}(p)$ (independent of Δ)
- Go to punishment state after unilateral deviations
- Otherwise remain in normal state

Punishment state:

- Common action $\underline{\kappa}(p)$ (independent of Δ)
- Remain in this state after unilateral deviations
- Otherwise go to normal state with probability $\gamma^{\Delta}(p)$

Actions

Actions

Take cut-off beliefs $\underline{p} < \bar{p}$ such that

$$\hat{p} < \underline{p} < \hat{p} + \epsilon \quad \text{and} \quad 1 - \epsilon < \bar{p} < 1$$

Actions

Take cut-off beliefs $\underline{p} < \bar{p}$ such that

$$\hat{p} < \underline{p} < \hat{p} + \epsilon \quad \text{and} \quad 1 - \epsilon < \bar{p} < 1$$

Set

$$\overline{\kappa}(p) = \begin{cases} 1 & \text{for} \quad p > \underline{p} \\ 0 & \text{for} \quad p \leq \underline{p} \end{cases}$$

and

$$\underline{\kappa}(p) = \begin{cases} 1 & \text{for} \quad p > \bar{p} \\ 0 & \text{for} \quad p \leq \bar{p} \end{cases}$$

Verification of Incentive Compatibility

Verification of Incentive Compatibility

No incentives needed at beliefs $p > \bar{p}$ or $p < \underline{p}$.

Verification of Incentive Compatibility

No incentives needed at beliefs $p > \bar{p}$ or $p < \underline{p}$.

Away from $\underline{p}$, have that $\overline{w}^{\Delta} - \underline{w}^{\Delta} > \nu > 0$, while benefit from deviation is of order Δ.

Verification of Incentive Compatibility

No incentives needed at beliefs $p > \bar{p}$ or $p < \underline{p}$.

Away from $\underline{p}$, have that $\overline{w}^{\Delta} - \underline{w}^{\Delta} > \nu > 0$, while benefit from deviation is of order Δ.

"Close to $\underline{p}$," $\overline{w}^{\Delta}$ gets close to $\underline{w}^{\Delta}$, but terms of order Δ go the right way (as $\underline{p} > \hat{p}$).

An Upper Bound on Equilibrium Payoffs

$\breve{p}^{\Delta}$: infimum of set of beliefs at which there is some PBE giving a payoff $> s$ to at least one player, and

$$\breve{p} = \liminf_{\Delta \to 0} \breve{p}^{\Delta}$$

An Upper Bound on Equilibrium Payoffs

$\breve{p}^{\Delta}$: infimum of set of beliefs at which there is some PBE giving a payoff $> s$ to at least one player, and

$$\breve{p} = \liminf_{\Delta \to 0} \breve{p}^{\Delta}$$

By construction, $\hat{p} \geq \breve{p} \geq p_N^*$.

An Upper Bound on Equilibrium Payoffs

$\breve{p}^{\Delta}$: infimum of set of beliefs at which there is some PBE giving a payoff $> s$ to at least one player, and

$$\breve{p} = \liminf_{\Delta \to 0} \breve{p}^{\Delta}$$

By construction, $\hat{p} \geq \breve{p} \geq p_N^*$.

* Can show that players' average payoff is bounded above by a function which converges to the **same** function $V_{N,\breve{p}}$.

An Upper Bound on Equilibrium Payoffs

$\breve{p}^\Delta$: infimum of set of beliefs at which there is some PBE giving a payoff $> s$ to at least one player, and

$$\breve{p} = \liminf_{\Delta \to 0} \breve{p}^\Delta$$

By construction, $\hat{p} \geq \breve{p} \geq p_N^*$.

* Can show that players' average payoff is bounded above by a function which converges to the **same** function $V_{N,\breve{p}}$.

* If L players play risky with positive probability, they can get at most $N\breve{W}^\Delta - (N-L)W_1^\Delta$ after any history.

An Upper Bound on Equilibrium Payoffs

$\breve{p}^{\Delta}$: infimum of set of beliefs at which there is some PBE giving a payoff $> s$ to at least one player, and

$$\breve{p} = \liminf_{\Delta \to 0} \breve{p}^{\Delta}$$

By construction, $\hat{p} \geq \breve{p} \geq p_N^*$.

* Can show that players' average payoff is bounded above by a function which converges to the **same** function $V_{N,\breve{p}}$.

* If L players play risky with positive probability, they can get at most $N\breve{W}^{\Delta} - (N - L)W_1^{\Delta}$ after any history.

Using these two facts, one shows that $\breve{p} = \tilde{p} = \hat{p}$.

An Upper Bound on Equilibrium Payoffs

$\breve{p}^{\Delta}$: infimum of set of beliefs at which there is some PBE giving a payoff $> s$ to at least one player, and

$$\breve{p} = \liminf_{\Delta \to 0} \breve{p}^{\Delta}$$

By construction, $\hat{p} \geq \breve{p} \geq p_N^*$.

* Can show that players' average payoff is bounded above by a function which converges to the **same** function $V_{N,\breve{p}}$.

* If L players play risky with positive probability, they can get at most $N\breve{W}^{\Delta} - (N - L)W_1^{\Delta}$ after any history.

Using these two facts, one shows that $\breve{p} = \tilde{p} = \hat{p}$. Thus:

Proposition: The set of PBE average payoffs coincides with the set of SSE average payoffs.

Conclusion

Concluding Remarks

- **Poisson bandits**

Concluding Remarks

- **Poisson bandits**
  - Continuous time

Concluding Remarks

- **Poisson bandits**
  - Continuous time
  - News comes in ‘lumps’

Concluding Remarks

- **Poisson bandits**
  - Continuous time
  - News comes in ‘lumps’
- **Strongly symmetric equilibria**

Concluding Remarks

- **Poisson bandits**
  - Continuous time
  - News comes in ‘lumps’
- **Strongly symmetric equilibria**
  - Freeze actions for Δ units of time

Concluding Remarks

- **Poisson bandits**
  - Continuous time
  - News comes in ‘lumps’
- **Strongly symmetric equilibria**
  - Freeze actions for Δ units of time
  - Characterize range of experimentation and supremum payoffs as Δ vanishes

Concluding Remarks

- **Poisson bandits**
  - Continuous time
  - News comes in 'lumps'
- **Strongly symmetric equilibria**
  - Freeze actions for Δ units of time
  - Characterize range of experimentation and supremum payoffs as Δ vanishes
- **Results**

Concluding Remarks

- **Poisson bandits**
  - Continuous time
  - News comes in ‘lumps’
- **Strongly symmetric equilibria**
  - Freeze actions for Δ units of time
  - Characterize range of experimentation and supremum payoffs as Δ vanishes
- **Results**
  - Efficiency is possible in the limit of SSE if and only if news events are ‘small’ (λ_0 close to λ_1)

Concluding Remarks

- **Poisson bandits**
  - Continuous time
  - News comes in ‘lumps’
- **Strongly symmetric equilibria**
  - Freeze actions for Δ units of time
  - Characterize range of experimentation and supremum payoffs as Δ vanishes
- **Results**
  - Efficiency is possible in the limit of SSE if and only if news events are ‘small’ (λ_0 close to λ_1)
  - Asymmetric PBE do not increase the range of experimentation beyond $[\hat{p}, 1]$

Concluding Remarks

- **Poisson bandits**
  - Continuous time
  - News comes in ‘lumps’
- **Strongly symmetric equilibria**
  - Freeze actions for Δ units of time
  - Characterize range of experimentation and supremum payoffs as Δ vanishes
- **Results**
  - Efficiency is possible in the limit of SSE if and only if news events are ‘small’ (λ_0 close to λ_1)
  - Asymmetric PBE do not increase the range of experimentation beyond $[\hat{p}, 1]$

Appendix

A Pair of Coupled Functional Equations

A Pair of Coupled Functional Equations

Generalizing Cronshaw & Luenberger (1994):

$$\overline{W}^{\Delta}(p) = \max_{k \in \mathcal{K}(p;\overline{W}^{\Delta},\underline{W}^{\Delta})} \Big\{(1-\delta)[(1-k)s + k\lambda(p)h] + \delta \mathsf{E}^{\Delta}\left[\overline{W}^{\Delta} | Nk, p\right]\Big\}$$

A Pair of Coupled Functional Equations

Generalizing Cronshaw & Luenberger (1994):

$$\overline{W}^{\Delta}(p) = \max_{k \in \mathcal{K}(p;\overline{W}^{\Delta},\underline{W}^{\Delta})} \Big\{ (1-\delta)\big[(1-k)s + k\lambda(p)h\big] + \delta \mathsf{E}^{\Delta}\left[\overline{W}^{\Delta} | Nk, p\right] \Big\}$$

$$\underline{W}^{\Delta}(p) = \min_{k \in \mathcal{K}(p;\overline{W}^{\Delta},\underline{W}^{\Delta})} \max_{k' \in \{0,1\}} \Big\{ (1-\delta)\big[(1-k')s + k'\lambda(p)h\big] + \delta \mathsf{E}^{\Delta}\left[\underline{W}^{\Delta} | (N-1)k + k', p\right] \Big\}$$

A Pair of Coupled Functional Equations

Generalizing Cronshaw & Luenberger (1994):

$$\overline{W}^{\Delta}(p) = \max_{k \in \mathcal{K}(p;\overline{W}^{\Delta},\underline{W}^{\Delta})} \Big\{(1-\delta)[(1-k)s + k\lambda(p)h] + \delta \mathsf{E}^{\Delta}\Big[\overline{W}^{\Delta}|Nk,p\Big]\Big\}$$

$$\underline{W}^{\Delta}(p) = \min_{k \in \mathcal{K}(p;\overline{W}^{\Delta},\underline{W}^{\Delta})} \max_{k' \in \{0,1\}} \Big\{(1-\delta)[(1-k')s + k'\lambda(p)h] + \delta \mathsf{E}^{\Delta}\Big[\underline{W}^{\Delta}|(N-1)k+k',p\Big]\Big\}$$

with $\mathcal{K}(p;\overline{W}^{\Delta},\underline{W}^{\Delta}) \subseteq \{0,1\}$ denoting the set of actions satisfying

$$(1-\delta)[(1-k)s + k\lambda(p)h] + \delta \mathsf{E}^{\Delta}\Big[\overline{W}^{\Delta}|Nk,p\Big]$$
$$\geq (1-\delta)[ks + (1-k)\lambda(p)h] + \delta \mathsf{E}^{\Delta}\Big[\underline{W}^{\Delta}|(N-1)k+1-k,p\Big]$$

Payoff Functions and State Transitions

Payoff Functions and State Transitions

$$\overline{w}^{\Delta}(p) = (1-\delta)\big[(1-\overline{\kappa}(p))s + \overline{\kappa}(p)\lambda(p)h\big] + \delta \mathsf{E}^{\Delta}\left[\overline{w}^{\Delta} | N\overline{\kappa}(p), p\right]$$

Payoff Functions and State Transitions

$$\overline{w}^{\Delta}(p) = (1-\delta)\left[(1-\overline{\kappa}(p))s + \overline{\kappa}(p)\lambda(p)h\right] + \delta \mathsf{E}^{\Delta}\left[\overline{w}^{\Delta} \middle| N\overline{\kappa}(p), p\right]$$

$$\underline{w}^{\Delta}(p) = \max_{k\in\{0,1\}} \left\{(1-\delta)\left[(1-k)s + k\lambda(p)h\right] + \delta \mathsf{E}^{\Delta}\left[\underline{w}^{\Delta} \middle| (N-1)\underline{\kappa}(p) + k, p\right]\right\}$$

Payoff Functions and State Transitions

$$\overline{w}^{\Delta}(p) = (1-\delta)\left[(1-\overline{\kappa}(p))s + \overline{\kappa}(p)\lambda(p)h\right] + \delta \mathsf{E}^{\Delta}\left[\overline{w}^{\Delta} | N\overline{\kappa}(p), p\right]$$

$$\underline{w}^{\Delta}(p) = \max_{k\in\{0,1\}} \left\{(1-\delta)\left[(1-k)s + k\lambda(p)h\right] + \delta \mathsf{E}^{\Delta}\left[\underline{w}^{\Delta} | (N-1)\underline{\kappa}(p) + k, p\right]\right\}$$

$$= (1-\delta)\left[(1-\underline{\kappa}(p))s + \underline{\kappa}(p)\lambda(p)h\right] + \delta \mathsf{E}^{\Delta}\left[\gamma^{\Delta}(p)\overline{w}^{\Delta} + (1-\gamma^{\Delta}(p))\underline{w}^{\Delta} | N\underline{\kappa}(p), p\right]$$

Incentive Compatibility

Incentive Compatibility

The common action k can be sustained with continuation payoffs $\overline{w}^{\Delta}$ and $\underline{w}^{\Delta}$ if and only if

$$(1-\delta)[(1-k)s + k\lambda(p)h] + \delta \mathsf{E}^{\Delta}\left[\overline{w}^{\Delta} | Nk, p\right]$$

$$\geq (1-\delta)[ks + (1-k)\lambda(p)h] + \delta \mathsf{E}^{\Delta}\left[\underline{w}^{\Delta} | (N-1)k + 1 - k, p\right]$$

Incentive Compatibility

The common action k can be sustained with continuation payoffs $\overline{w}^{\Delta}$ and $\underline{w}^{\Delta}$ if and only if

$$(1-\delta)\big[(1-k)s + k\lambda(p)h\big] + \delta \mathsf{E}^{\Delta}\left[\overline{w}^{\Delta}|Nk,p\right]$$

$$\geq (1-\delta)\big[ks + (1-k)\lambda(p)h\big] + \delta \mathsf{E}^{\Delta}\left[\underline{w}^{\Delta}|(N-1)k+1-k,p\right]$$

$\gamma^{\Delta}(p) = 0$ if and only if $k = \underline{\kappa}(p)$ can be sustained with continuation payoff $\underline{w}^{\Delta}$

The Case $\lambda_0 = 0$

– Problem: $\hat{p} = p_1^*$, i.e. we can't squeeze $\underline{p}$ into $(\hat{p}, p_1^*)$ any more
$\Rightarrow$ Analyze the discrete-time game in some detail

The Case $\lambda_0 = 0$

– Problem: $\hat{p} = p_1^*$, i.e. we can't squeeze $\underline{p}$ into $(\hat{p}, p_1^*)$ any more
$\Rightarrow$ Analyze the discrete-time game in some detail

– Check for symmetric MPE with individual randomization first.

The Case $\lambda_0 = 0$

– Problem: $\hat{p} = p_1^*$, i.e. we can't squeeze $\underline{p}$ into $(\hat{p}, p_1^*)$ any more
$\Rightarrow$ Analyze the discrete-time game in some detail

– Check for symmetric MPE with individual randomization first.

– There exist several symmetric MPE on an open interval of beliefs!

The Case $\lambda_0 = 0$

– Problem: $\hat{p} = p_1^*$, i.e. we can't squeeze $\underline{p}$ into $(\hat{p}, p_1^*)$ any more
$\Rightarrow$ Analyze the discrete-time game in some detail

– Check for symmetric MPE with individual randomization first.

– There exist several symmetric MPE on an open interval of beliefs!

– Use MPE as continuation equilibrium to show that $\bar{\kappa} = 1$ can be sustained arbitrarily close to p_1^* as $\Delta \to 0$.

The Case $\lambda_0 = 0$

- Problem: $\hat{p} = p_1^*$, i.e. we can't squeeze $\underline{p}$ into $(\hat{p}, p_1^*)$ any more

  $\Rightarrow$ Analyze the discrete-time game in some detail
- Check for symmetric MPE with individual randomization first.
- There exist several symmetric MPE on an open interval of beliefs!
- Use MPE as continuation equilibrium to show that $\bar{\kappa} = 1$ can be sustained arbitrarily close to p_1^* as $\Delta \to 0$.
- Use this good SSE to show that $\underline{\kappa} = 0$ can be enforced as well.