Overcoming Free-Riding in Bandit Games

Johannes Hörner (Yale University)
Nicolas Klein (University of Montreal)
Sven Rady (University of Bonn)

International Conference on Game Theory
Stony Brook

24 July 2020

Optimal Experimentation

Optimal experimentation (active learning) involves the well-known trade-off between exploitation and exploration

Optimal Experimentation

Optimal experimentation (active learning) involves the well-known trade-off between exploitation and exploration

Multi-armed bandit is a stylized model of this 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

Optimal Experimentation

Optimal experimentation (active learning) involves the well-known trade-off between exploitation and exploration

Multi-armed bandit is a stylized model of this 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

Surveys:
- Bergemann & Välimäki (2008)
- Hörner & Skrzypacz (2016)

Optimal Experimentation

Optimal experimentation (active learning) involves the well-known trade-off between exploitation and exploration

Multi-armed bandit is a stylized model of this 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

Surveys:
- – Bergemann & Välimäki (2008)
- – Hörner & Skrzypacz (2016)

Examples:
- – Farmer choosing between a traditional crop and a gene-modified one
- – Researcher pursuing a new research agenda

Strategic Experimentation

Agents learn from the exploration efforts of others, as well as from their own

Examples:

- Farmers with neighboring fields
- Researchers pursuing a common research agenda

Strategic Experimentation with Two-Armed Bandits

Simplest possible framework:

- One arm is ‘safe’
- The ‘risky’ arm is either ‘good’ or ‘bad’
- Type of risky arm identical across players, unknown
- Publicly observable actions and outcomes
- Pure informational externality

Strategic Experimentation with Two-Armed Bandits

Bolton & Harris (1999):

- Brownian motion with high or low drift
- Unique symmetric Markov perfect equilibrium
- Free-riding versus encouragement

Strategic Experimentation with Two-Armed Bandits

Bolton & Harris (1999):

- Brownian motion with high or low drift
- Unique symmetric Markov perfect equilibrium
- Free-riding versus encouragement

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
- Higher welfare by taking turns

This Paper

Good-news Lévy bandits (Cohen and Solan 2013)

This Paper

Good-news Lévy bandits (Cohen and Solan 2013)

Explore non-Markovian behaviour

– Freeze actions for a small length of time Δ

– Consider the limit $\Delta \to 0$
(Abreu, Milgrom & Pearce 1991; Hellwig & Schmidt 2002; Biais, Mariotti, Plantin & Rochet 2007; Sadzik & Stacchetti 2015)

– Characterize limit PBE payoffs

This Paper

Good-news Lévy bandits (Cohen and Solan 2013)

Explore non-Markovian behaviour

- Freeze actions for a small length of time Δ
- Consider the limit $\Delta \to 0$
  (Abreu, Milgrom & Pearce 1991; Hellwig & Schmidt 2002; Biais, Mariotti, Plantin & Rochet 2007; Sadzik & Stacchetti 2015)
- Characterize limit PBE payoffs
- Construct SSE that attain these payoffs
- Use recursive approach / two-state automaton
  (Abreu 1988; Cronshaw & Luenberger 1994)

This Paper

Good-news Lévy bandits (Cohen and Solan 2013)

Explore non-Markovian behaviour
- Freeze actions for a small length of time Δ
- Consider the limit $\Delta \to 0$ (Abreu, Milgrom & Pearce 1991; Hellwig & Schmidt 2002; Biais, Mariotti, Plantin & Rochet 2007; Sadzik & Stacchetti 2015)
- Characterize limit PBE payoffs
- Construct SSE that attain these payoffs
- Use recursive approach / two-state automaton (Abreu 1988; Cronshaw & Luenberger 1994)
- Provide a necessary and sufficient condition for efficiency in the limit

Preview

When the risky payoff has an informative Brownian component:

– Posterior beliefs have unbounded variation
– Benefit of experimentation vanishes like Δ^ρ with $\rho < 1$
– Opportunity cost of experimentation vanishes like Δ
– Efficiency in the limit

Preview

When the risky payoff has an informative Brownian component:

– Posterior beliefs have unbounded variation

– Benefit of experimentation vanishes like Δ^ρ with $\rho < 1$

– Opportunity cost of experimentation vanishes like Δ

– Efficiency in the limit

When all the learning is from Poisson jumps:

– Posterior beliefs are piecewise deterministic

– Benefit of experimentation vanishes like Δ

– Asymptotic efficiency depends on the ability to punish a deviator after good news

– Efficiency if and only if news are 'small'

Preview

When the risky payoff has an informative Brownian component:

- Posterior beliefs have unbounded variation
- Benefit of experimentation vanishes like Δ^ρ with $\rho < 1$
- Opportunity cost of experimentation vanishes like Δ
- Efficiency in the limit

When all the learning is from Poisson jumps:

- Posterior beliefs are piecewise deterministic
- Benefit of experimentation vanishes like Δ
- Asymptotic efficiency depends on the ability to punish a deviator after good news
- Efficiency if and only if news are ‘small’

Choice of equilibrium concept and specification of payoff-generating processes matter!

Model

A Two-Armed Bandit in Continuous Time

One arm is **safe** (S)

generates a known flow payoff $s > 0$

A Two-Armed Bandit in Continuous Time

One arm is **safe** (S)

generates a known flow payoff $s > 0$

Other arm is **risky** (R)

yields stochastic payoff increments

$$dY_t = \alpha_\theta \, dt + \sigma \, dZ_t + h \, dN_t^\theta$$

where

$\theta \in \{0, 1\}$ unknown state of the world

Z Wiener process

N^θ Poisson process with intensity λ_θ

$\alpha_0, \alpha_1, \sigma, h, \lambda_0, \lambda_1$ known scalar parameters

Assumptions

Expected infinitesimal payoff increment from R in state θ:

$$m_\theta = \alpha_\theta + \lambda_\theta h$$

Assume:

(i) $m_0 < s < m_1$

(ii) $\sigma > 0$ and $h > 0$

(iii) $\lambda_1 \geq \lambda_0 \geq 0$

Special cases:

$\lambda_1 = \lambda_0 \longrightarrow$ Bolton & Harris (1999)

$\alpha_1 = \alpha_0$ and $\lambda_0 > 0 \longrightarrow$ Keller & Rady (2010)

$\alpha_1 = \alpha_0$ and $\lambda_0 = 0 \longrightarrow$ Keller, Rady & Cripps (2005)

Beliefs

Each player $n = 1, \ldots, N$ has a replica two-armed bandit

– same θ

– independent payoff processes

Common prior

Observable actions and outcomes

Hence common posterior

Explicit characterization of the belief process
(Cohen & Solan 2013)

Beliefs and Payoffs

p_t = posterior probability at time t that θ equals 1

Expected flow payoff from R conditional on observations up to time t:

$$m(p_t) = p_t m_1 + (1 - p_t) m_0$$

– ‘myopic payoff’

Total payoff in per-period units:

$$\mathbb{E}\left[\int_0^\infty r\, e^{-r\,t} \,\{(1-k_t)s + k_t m(p_t)\}\; dt\right]$$

where $r > 0$ and $k_t \in \{0, 1\}$

Continuous-Time Analysis

Single-Agent Problem

HJB equation:

$$u(p) = s + \max_{k \in \{0,1\}} k \left\{ b(p,u) - c(p) \right\}$$

with

$$c(p) \quad = \quad s - m(p)$$

Single-Agent Problem

HJB equation:

$$u(p) = s + \max_{k \in \{0,1\}} k \left\{ b(p,u) - c(p) \right\}$$

with

$$\begin{aligned} c(p) &= s - m(p) \\ b(p,u) &= \frac{1}{r} \left\{ \frac{1}{2} \left(\frac{\alpha_1 - \alpha_0}{\sigma} \right)^2 p^2 (1-p)^2 u''(p) \right. \end{aligned}$$

Single-Agent Problem

HJB equation:

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

with

$$\begin{aligned} c(p) &= s - m(p) \\ b(p,u) &= \frac{1}{r}\left\{\frac{1}{2}\left(\frac{\alpha_1 - \alpha_0}{\sigma}\right)^2 p^2(1-p)^2 u''(p) \right. \\ &\qquad \left. - (\lambda_1 - \lambda_0)\, p(1-p)u'(p)\right\} \end{aligned}$$

Single-Agent Problem

HJB equation:

$$u(p) = s + \max_{k\in\{0,1\}} k\left\{b(p,u) - c(p)\right\}$$

with

$$\begin{aligned}
c(p) &= s - m(p) \\
b(p,u) &= \frac{1}{r}\left\{\frac{1}{2}\left(\frac{\alpha_1-\alpha_0}{\sigma}\right)^2 p^2(1-p)^2u''(p)\right. \\
&\quad - (\lambda_1-\lambda_0)\,p(1-p)u'(p) \\
&\quad \left. + \lambda(p)\left[u(j(p)) - u(p)\right]\right\} \\
\lambda(p) &= p\lambda_1 + (1-p)\lambda_0 \\
j(p) &= \frac{p\lambda_1}{\lambda(p)}
\end{aligned}$$

Single-Agent Problem

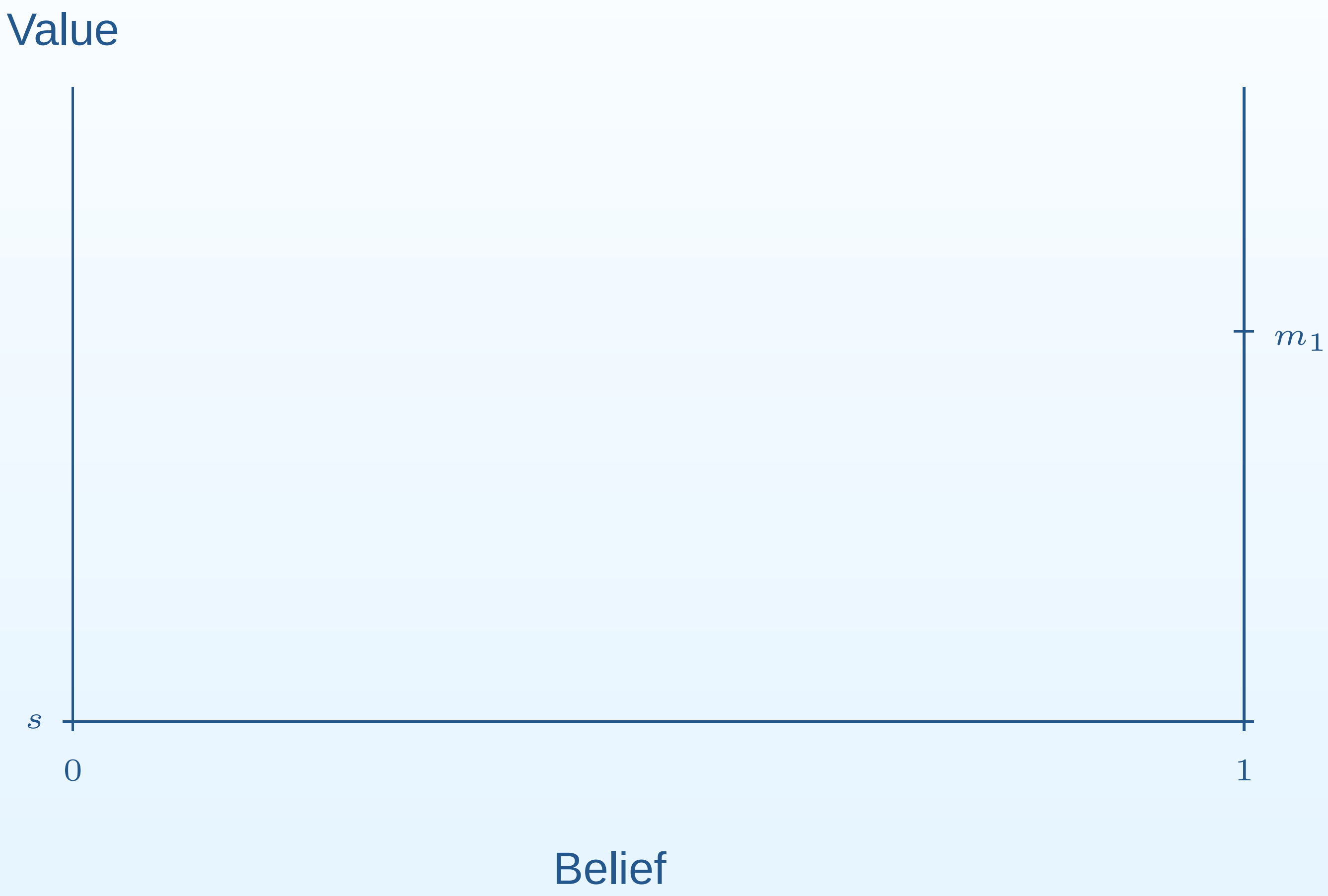
HJB equation:

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

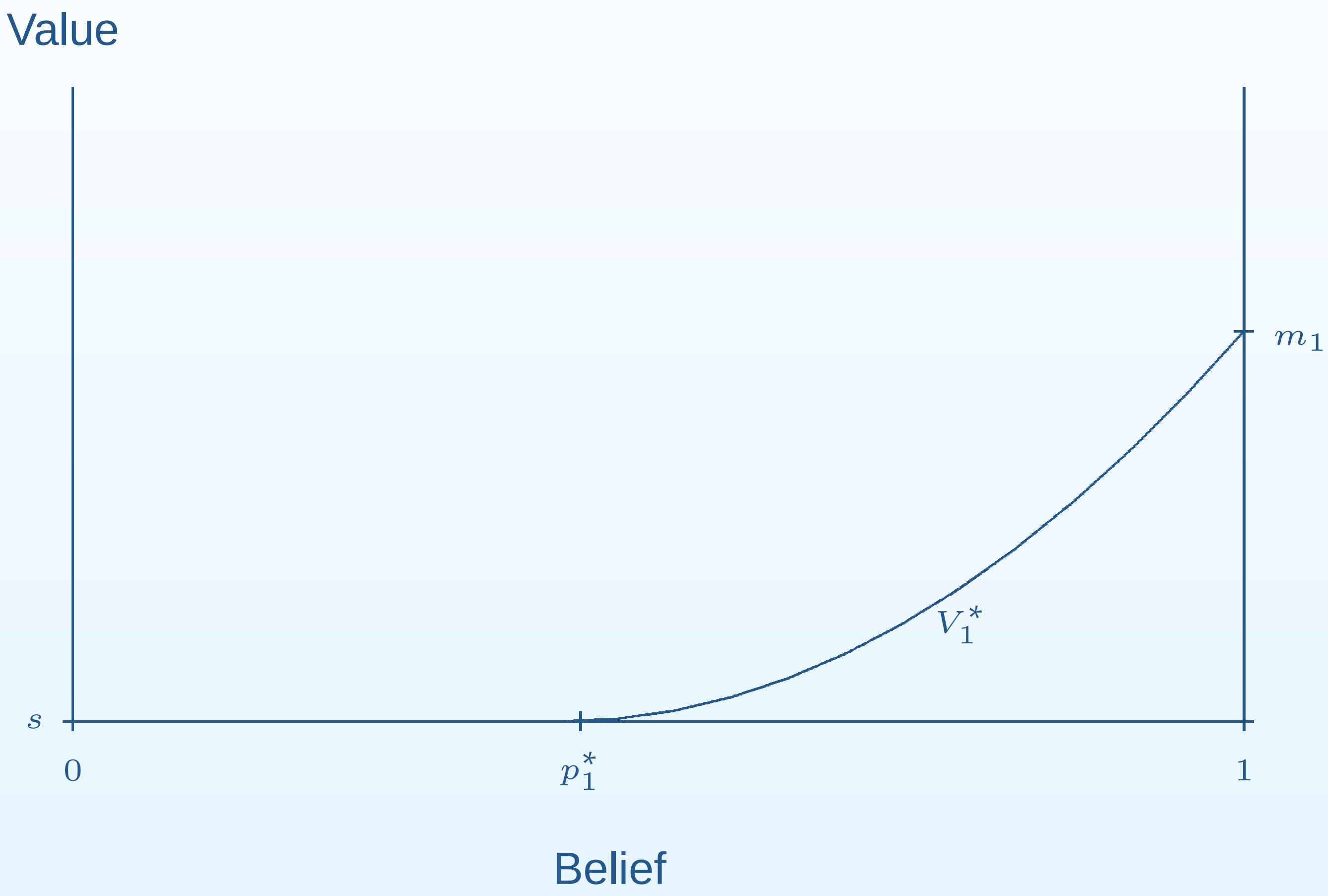
Optimal cut-off p_1^*

Value function V_1^* in closed form

Single-Agent Problem



Single-Agent Problem



Efficient Experimentation

Bellman equation for N players jointly maximising their *average* payoff:

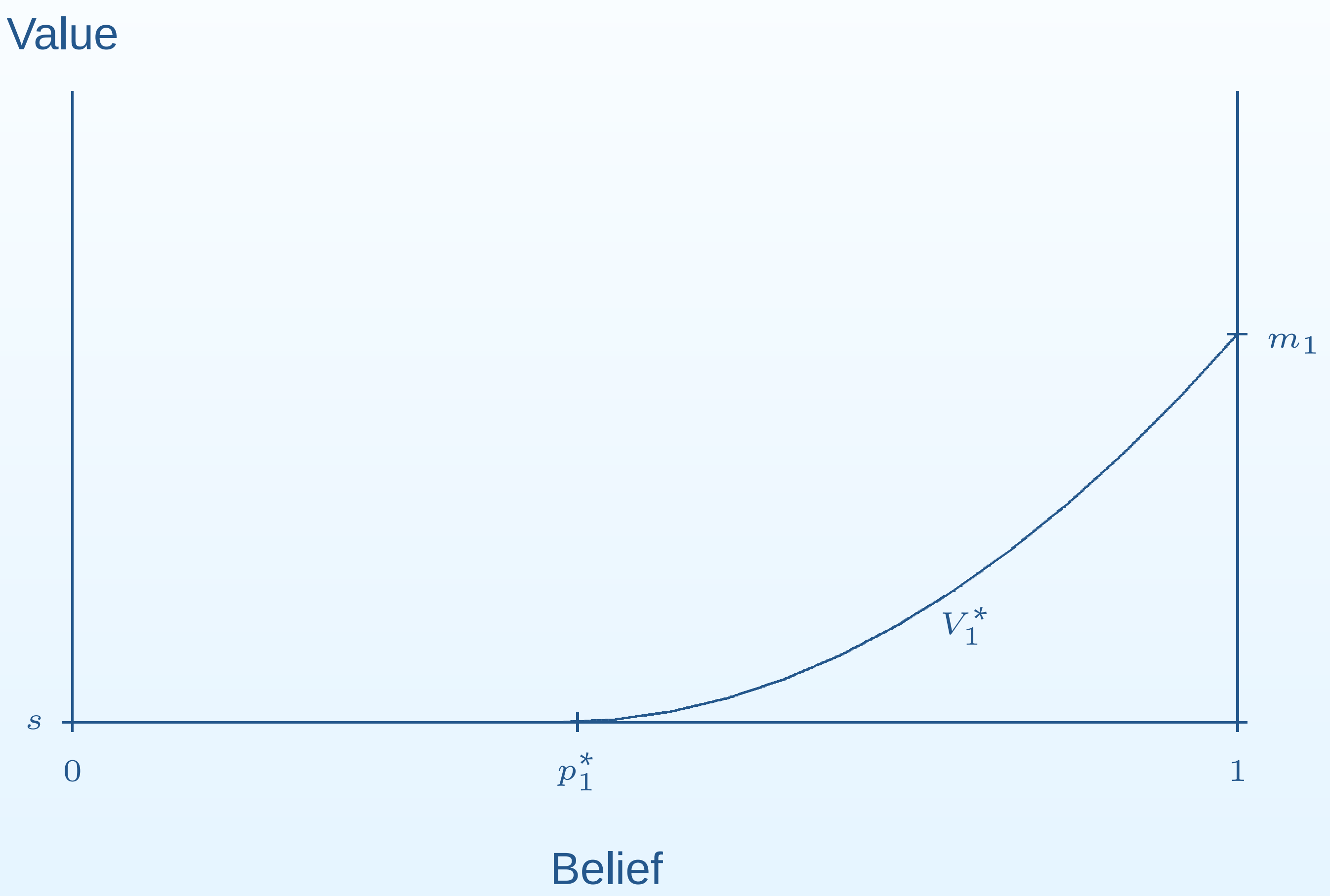
$$u(p) = s + \max_{K \in \{0,1,\ldots,N\}} K \left\{ b(p,u) - c(p)/N \right\}$$

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

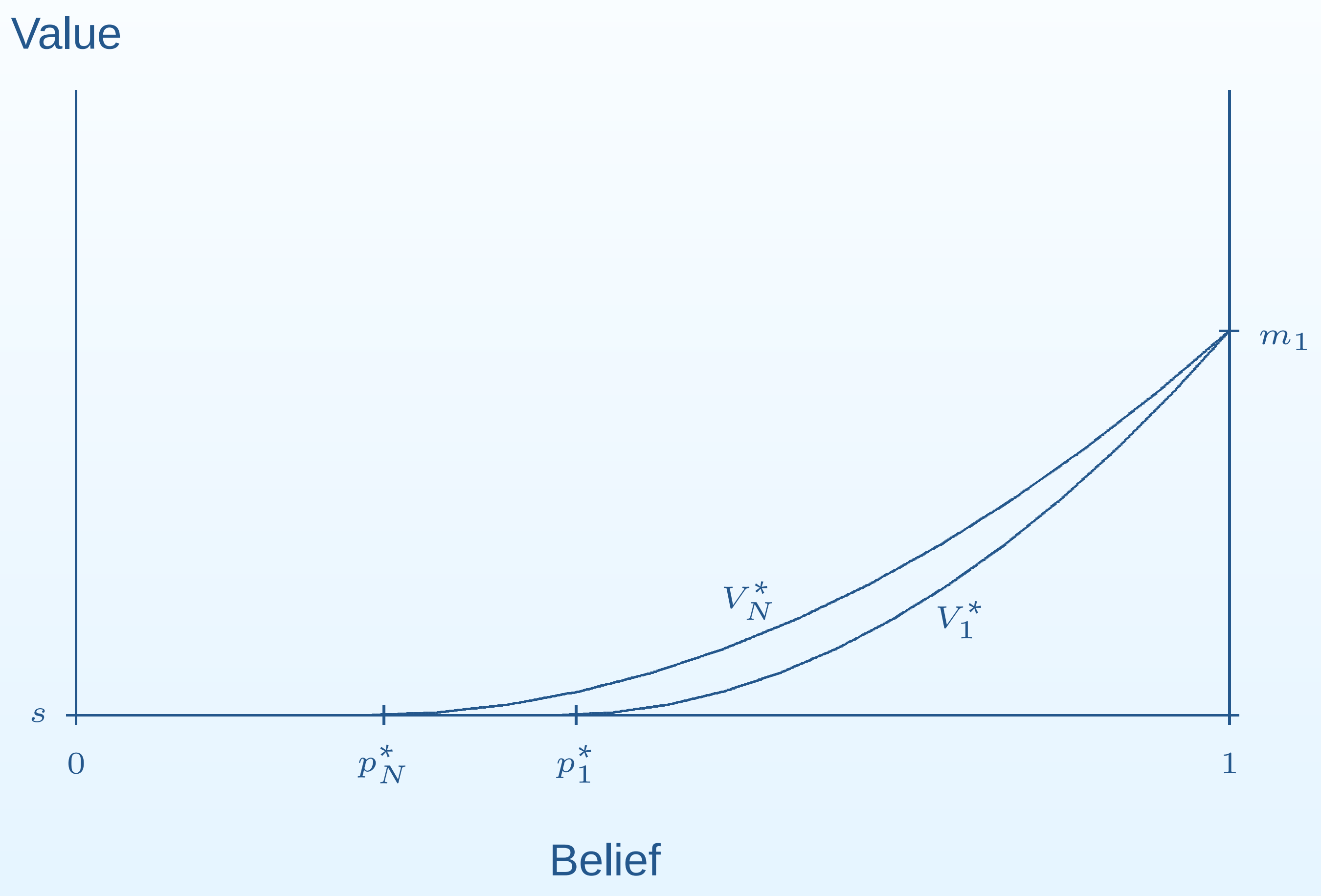
Optimal cut-off $p_N^* < p_1^*$, decreasing in N

Value function V_N^* in closed form

Efficient Experimentation



Efficient Experimentation



Markov Perfect Equilibria

Suppose all players $\ell \neq n$ use Markov strategies:

$$k_{\ell,t} = k_\ell(p_t)$$

Bellman equation for player n:

$$u_n(p) = s + \sum_{\ell \neq n} k_\ell(p) b(p, u_n) + \max_{k_n \in \{0,1\}} k_n \left\{ b(p, u_n) - c(p) \right\}$$

Markov Perfect Equilibria

Suppose all players $\ell \neq n$ use Markov strategies:

$$k_{\ell,t} = k_\ell(p_t)$$

Bellman equation for player n:

$$u_n(p) = s + \sum_{\ell \neq n} k_\ell(p) b(p, u_n) + \max_{k_n \in \{0,1\}} k_n \left\{ b(p, u_n) - c(p) \right\}$$

Efficient behaviour is not a Markov perfect equilibrium:

$$b - c/N \quad \text{versus} \quad b - c$$

– free-rider problem

Discretizing the Game

Strategies and Payoffs

Players can adjust their actions at $t = 0, \Delta, 2\Delta, 3\Delta, \ldots$ for some fixed $\Delta > 0$

Payoff increment from playing S:

$$\int_0^\Delta r\, e^{-r\,t} s\, dt = (1 - \delta) s \quad \text{with} \quad \delta = e^{-r\,\Delta}$$

Expected payoff increment from R conditional on observations up to time $t \in \{0, \Delta, 2\Delta, \ldots\}$:

$$(1 - \delta) m(p_t)$$

Equilibrium Concepts

Perfect Bayesian equilibrium (PBE) is defined as usual

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

Equilibrium Payoffs

For given $\Delta > 0$ and initial belief p, the sets of average (per player) PBE and SSE payoffs have

- suprema $\overline{W}^{\Delta}_{\mathrm{PBE}}(p)$ and $\overline{W}^{\Delta}_{\mathrm{SSE}}(p)$
- infima $\underline{W}^{\Delta}_{\mathrm{PBE}}(p)$ and $\underline{W}^{\Delta}_{\mathrm{SSE}}(p)$

in the interval $[s, m_1]$

Equilibrium Payoffs

For given $\Delta > 0$ and initial belief p, the sets of average (per player) PBE and SSE payoffs have

- suprema $\overline{W}^{\Delta}_{\mathrm{PBE}}(p)$ and $\overline{W}^{\Delta}_{\mathrm{SSE}}(p)$
- infima $\underline{W}^{\Delta}_{\mathrm{PBE}}(p)$ and $\underline{W}^{\Delta}_{\mathrm{SSE}}(p)$

in the interval $[s, m_1]$

With W^{Δ}_1 denoting the single-agent value function:

$$W^{\Delta}_1 \leq \underline{W}^{\Delta}_{\mathrm{PBE}} \leq \underline{W}^{\Delta}_{\mathrm{SSE}} \leq \overline{W}^{\Delta}_{\mathrm{SSE}} \leq \overline{W}^{\Delta}_{\mathrm{PBE}} \leq V^{*}_{N}$$

Equilibrium Payoffs

For given $\Delta > 0$ and initial belief p, the sets of average (per player) PBE and SSE payoffs have

- suprema $\overline{W}^{\Delta}_{\mathrm{PBE}}(p)$ and $\overline{W}^{\Delta}_{\mathrm{SSE}}(p)$
- infima $\underline{W}^{\Delta}_{\mathrm{PBE}}(p)$ and $\underline{W}^{\Delta}_{\mathrm{SSE}}(p)$

in the interval $[s, m_1]$

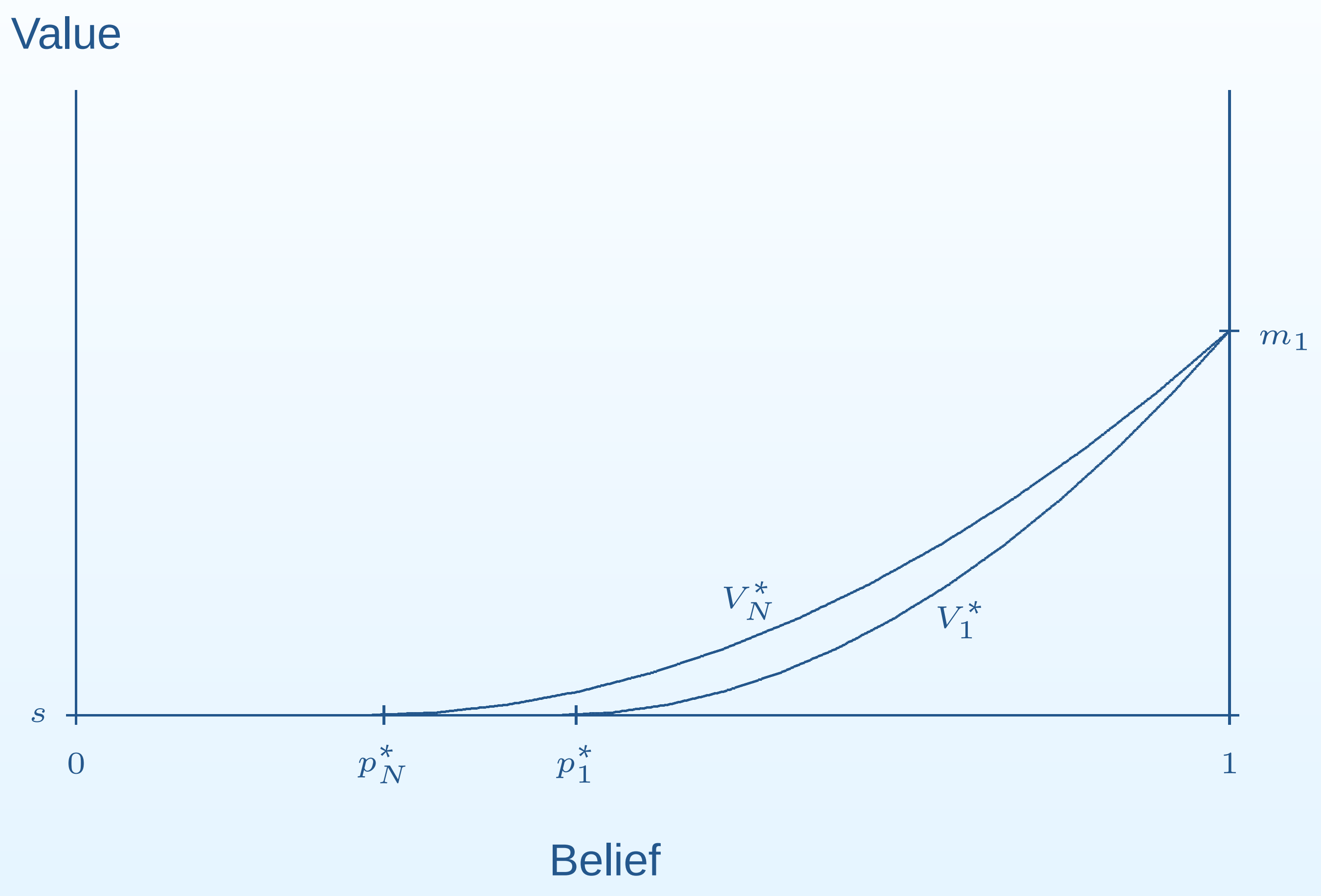
With W^{Δ}_1 denoting the single-agent value function:

$$W^{\Delta}_1 \leq \underline{W}^{\Delta}_{\mathrm{PBE}} \leq \underline{W}^{\Delta}_{\mathrm{SSE}} \leq \overline{W}^{\Delta}_{\mathrm{SSE}} \leq \overline{W}^{\Delta}_{\mathrm{PBE}} \leq V^*_N$$

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

$$W^{\Delta}_1 \to V^*_1$$

Equilibrium Payoffs



Theorem

(i) *There exists* $\hat{p} \in [p_N^*, p_1^*]$ *such that*

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{SSE}} = V_{N,\hat{p}}$$

$$\lim_{\Delta \to 0} \underline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \underline{W}^{\Delta}_{\text{SSE}} = V_1^*$$

uniformly on $[0, 1]$, *where* $V_{N,\hat{p}}$ *is the players' continuous-time payoff function when they all use the cut-off* $\hat{p}$.

Theorem

(i) *There exists* $\hat{p} \in [p_N^*, p_1^*]$ *such that*

$$\lim_{\Delta \to 0} \overline{W}_{\text{PBE}}^{\Delta} = \lim_{\Delta \to 0} \overline{W}_{\text{SSE}}^{\Delta} = V_{N,\hat{p}}$$

$$\lim_{\Delta \to 0} \underline{W}_{\text{PBE}}^{\Delta} = \lim_{\Delta \to 0} \underline{W}_{\text{SSE}}^{\Delta} = V_1^*$$

uniformly on $[0,1]$, *where* $V_{N,\hat{p}}$ *is the players' continuous-time payoff function when they all use the cut-off* $\hat{p}$.

(ii) *If* $\alpha_1 \neq \alpha_0$, *then* $\hat{p} = p_N^*$ *(and hence* $V_{N,\hat{p}} = V_N^*$*).*

Theorem

(i) *There exists* $\hat{p} \in [p_N^*, p_1^*]$ *such that*

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{SSE}} = V_{N,\hat{p}}$$

$$\lim_{\Delta \to 0} \underline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \underline{W}^{\Delta}_{\text{SSE}} = V_1^*$$

uniformly on $[0,1]$, *where* $V_{N,\hat{p}}$ *is the players' continuous-time payoff function when they all use the cut-off* $\hat{p}$.

(ii) *If* $\alpha_1 \neq \alpha_0$, *then* $\hat{p} = p_N^*$ *(and hence* $V_{N,\hat{p}} = V_N^*$*).*

(iii) *If* $\alpha_1 = \alpha_0$, *then* $\hat{p}$ *is the unique belief in* $[p_N^*, p_1^*]$ *satisfying*

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p});$$

Theorem

(i) *There exists* $\hat{p} \in [p_N^*, p_1^*]$ *such that*

$$\lim_{\Delta\to 0} \overline{W}^{\Delta}_{\mathrm{PBE}} = \lim_{\Delta\to 0} \overline{W}^{\Delta}_{\mathrm{SSE}} = V_{N,\hat{p}}$$

$$\lim_{\Delta\to 0} \underline{W}^{\Delta}_{\mathrm{PBE}} = \lim_{\Delta\to 0} \underline{W}^{\Delta}_{\mathrm{SSE}} = V_1^*$$

uniformly on $[0,1]$*, where* $V_{N,\hat{p}}$ *is the players' continuous-time payoff function when they all use the cut-off* $\hat{p}$*.*

(ii) *If* $\alpha_1 \neq \alpha_0$*, then* $\hat{p} = p_N^*$ *(and hence* $V_{N,\hat{p}} = V_N^*$*).*

(iii) *If* $\alpha_1 = \alpha_0$*, then* $\hat{p}$ *is the unique belief in* $[p_N^*, p_1^*]$ *satisfying*

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p});$$

$\hat{p} = p_N^*$ *if and only if* $j(p_N^*) \leq p_1^*$*;*

Theorem

(i) *There exists* $\hat{p} \in [p_N^*, p_1^*]$ *such that*

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\mathrm{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\mathrm{SSE}} = V_{N,\hat{p}}$$

$$\lim_{\Delta \to 0} \underline{W}^{\Delta}_{\mathrm{PBE}} = \lim_{\Delta \to 0} \underline{W}^{\Delta}_{\mathrm{SSE}} = V_1^*$$

uniformly on $[0, 1]$*, where* $V_{N,\hat{p}}$ *is the players' continuous-time payoff function when they all use the cut-off* $\hat{p}$*.*

(ii) *If* $\alpha_1 \neq \alpha_0$*, then* $\hat{p} = p_N^*$ *(and hence* $V_{N,\hat{p}} = V_N^*$*).*

(iii) *If* $\alpha_1 = \alpha_0$*, then* $\hat{p}$ *is the unique belief in* $[p_N^*, p_1^*]$ *satisfying*

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p});$$

$\hat{p} = p_N^*$ *if and only if* $j(p_N^*) \leq p_1^*$*;*
$\hat{p} = p_1^*$ *if and only if* $\lambda_0 = 0$*.*

Learning with a Brownian Component ($\alpha_1 \neq \alpha_0$)

Asymptotic efficiency:

$$\hat{p} = p_N^*$$

$$\lim_{\Delta \to 0} \overline{W}_{\text{PBE}}^{\Delta} = \lim_{\Delta \to 0} \overline{W}_{\text{SSE}}^{\Delta} = V_N^*$$

Learning with a Brownian Component ($\alpha_1 \neq \alpha_0$)

Asymptotic efficiency:

$$\hat{p} = p_N^*$$

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\mathrm{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\mathrm{SSE}} = V_N^*$$

Intuition:

Learning with a Brownian Component ($\alpha_1 \neq \alpha_0$)

Asymptotic efficiency:

$$\hat{p} = p_N^*$$

$$\lim_{\Delta\to 0} \overline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta\to 0} \overline{W}^{\Delta}_{\text{SSE}} = V_N^*$$

Intuition:

- For $\hat{p} > p_N^*$, the function $V_{N,\hat{p}}$ has a convex kink at $\hat{p}$

Learning with a Brownian Component ($\alpha_1 \neq \alpha_0$)

Asymptotic efficiency:

$$\hat{p} = p_N^*$$

$$\lim_{\Delta \to 0} \overline{W}_{\text{PBE}}^{\Delta} = \lim_{\Delta \to 0} \overline{W}_{\text{SSE}}^{\Delta} = V_N^*$$

Intuition:

- For $\hat{p} > p_N^*$, the function $V_{N,\hat{p}}$ has a convex kink at $\hat{p}$
- Diffusion part of beliefs then induces an incentive to experiment beyond $\hat{p}$

Learning with a Brownian Component ($\alpha_1 \neq \alpha_0$)

Asymptotic efficiency:

$$\hat{p} = p_N^*$$

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{SSE}} = V_N^*$$

Intuition:

- For $\hat{p} > p_N^*$, the function $V_{N,\hat{p}}$ has a convex kink at $\hat{p}$
- Diffusion part of beliefs then induces an incentive to experiment beyond $\hat{p}$
  - Informational benefit vanishes no faster than $\Delta^{\frac{3}{4}}$
  - Opportunity cost vanishes like Δ

Learning with a Brownian Component ($\alpha_1 \neq \alpha_0$)

Asymptotic efficiency:

$$\hat{p} = p_N^*$$

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{SSE}} = V_N^*$$

Intuition:

- For $\hat{p} > p_N^*$, the function $V_{N,\hat{p}}$ has a convex kink at $\hat{p}$
- Diffusion part of beliefs then induces an incentive to experiment beyond $\hat{p}$
  - Informational benefit vanishes no faster than $\Delta^{\frac{3}{4}}$
  - Opportunity cost vanishes like Δ
- Similar to smooth pasting in stopping problems

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency possible:

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{SSE}} = V_{N,\hat{p}}$$

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p})) - s\right] - (N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p})) - s\right] = rc(\hat{p})$$

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency possible:

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{SSE}} = V_{N,\hat{p}}$$

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p})$$

Intuition:

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency possible:

$$\lim_{\Delta\to 0} \overline{W}^{\Delta}_{\mathrm{PBE}} = \lim_{\Delta\to 0} \overline{W}^{\Delta}_{\mathrm{SSE}} = V_{N,\hat{p}}$$

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p})$$

Intuition:

- Informational benefit and opportunity cost of an experiment both vanish like Δ

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency possible:

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{SSE}} = V_{N,\hat{p}}$$

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p})$$

Intuition:

- Informational benefit and opportunity cost of an experiment both vanish like Δ
- Rates at which they vanish are equalized at $\hat{p}$

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency possible:

$$\lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{PBE}} = \lim_{\Delta \to 0} \overline{W}^{\Delta}_{\text{SSE}} = V_{N,\hat{p}}$$

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p})$$

Intuition:

- Informational benefit and opportunity cost of an experiment both vanish like Δ
- Rates at which they vanish are equalized at $\hat{p}$
- What matters is the ability to punish a deviator in the event that good news arrives

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic efficiency for 'small' news:

$$\hat{p} = p_N^* \iff j(p_N^*) \leq p_1^* \iff V_1^*(j(p_N^*)) = s$$

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic efficiency for ‘small’ news:

$$\hat{p} = p_N^* \iff j(p_N^*) \leq p_1^* \iff V_1^*(j(p_N^*)) = s$$

Compare

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p})) - s\right] - (N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p})) - s\right] = rc(\hat{p})$$

with

$$N\lambda(p_N^*)\left[V_N^*(j(p_N^*)) - s\right] = rc(p_N^*)$$

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic efficiency for 'small' news:

$$\hat{p} = p_N^* \iff j(p_N^*) \leq p_1^* \iff V_1^*(j(p_N^*)) = s$$

Compare

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p})$$

with

$$N\lambda(p_N^*)\left[V_N^*(j(p_N^*))-s\right] = rc(p_N^*)$$

Intuition:

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic efficiency for 'small' news:

$$\hat{p} = p_N^* \quad \Longleftrightarrow \quad j(p_N^*) \le p_1^* \quad \Longleftrightarrow \quad V_1^*(j(p_N^*)) = s$$

Compare

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p})$$

with

$$N\lambda(p_N^*)\left[V_N^*(j(p_N^*))-s\right] = rc(p_N^*)$$

Intuition:

- A deviating player can be held down to s even after news arrives

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic efficiency for 'small' news:

$$\hat{p} = p_N^* \quad \Longleftrightarrow \quad j(p_N^*) \leq p_1^* \quad \Longleftrightarrow \quad V_1^*(j(p_N^*)) = s$$

Compare

$$N\lambda(\hat{p})\left[V_{N,\hat{p}}(j(\hat{p}))-s\right]-(N-1)\lambda(\hat{p})\left[V_1^*(j(\hat{p}))-s\right] = rc(\hat{p})$$

with

$$N\lambda(p_N^*)\left[V_N^*(j(p_N^*))-s\right] = rc(p_N^*)$$

Intuition:

- A deviating player can be held down to s even after news arrives
- Players thus set the benefit of N experiments against the costs of one, as the social planner would

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency for ‘big’ news:

$$\hat{p} > p_N^* \iff j(p_N^*) > p_1^* \iff V_1^*(j(p_N^*)) > s$$

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency for ‘big’ news:

$$\hat{p} > p_N^* \quad \Longleftrightarrow \quad j(p_N^*) > p_1^* \quad \Longleftrightarrow \quad V_1^*(j(p_N^*)) > s$$

Intuition:

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency for ‘big’ news:

$$\hat{p} > p_N^* \iff j(p_N^*) > p_1^* \iff V_1^*(j(p_N^*)) > s$$

Intuition:

- Now a deviating player does better than s after news arrives

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency for ‘big’ news:

$$\hat{p} > p_N^* \quad \Longleftrightarrow \quad j(p_N^*) > p_1^* \quad \Longleftrightarrow \quad V_1^*(j(p_N^*)) > s$$

Intuition:

- Now a deviating player does better than s after news arrives
- Players thus set the benefit of N experiments against more than the costs of one

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency for 'big' news:

$$\hat{p} > p_N^* \quad \Longleftrightarrow \quad j(p_N^*) > p_1^* \quad \Longleftrightarrow \quad V_1^*(j(p_N^*)) > s$$

Intuition:

- Now a deviating player does better than s after news arrives
- Players thus set the benefit of N experiments against more than the costs of one

Extreme case:

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency for 'big' news:

$$\hat{p} > p_N^* \quad \Longleftrightarrow \quad j(p_N^*) > p_1^* \quad \Longleftrightarrow \quad V_1^*(j(p_N^*)) > s$$

Intuition:

- Now a deviating player does better than s after news arrives
- Players thus set the benefit of N experiments against more than the costs of one

Extreme case:

- Fully conclusive news ($\lambda_0 = 0$): $j(p_N^*) = 1$

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency for 'big' news:

$$\hat{p} > p_N^* \iff j(p_N^*) > p_1^* \iff V_1^*(j(p_N^*)) > s$$

Intuition:

- Now a deviating player does better than s after news arrives
- Players thus set the benefit of N experiments against more than the costs of one

Extreme case:

- Fully conclusive news ($\lambda_0 = 0$): $j(p_N^*) = 1$
- No scope at all for punishment after news arrives

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

Asymptotic inefficiency for ‘big’ news:

$$\hat{p} > p_N^* \iff j(p_N^*) > p_1^* \iff V_1^*(j(p_N^*)) > s$$

Intuition:

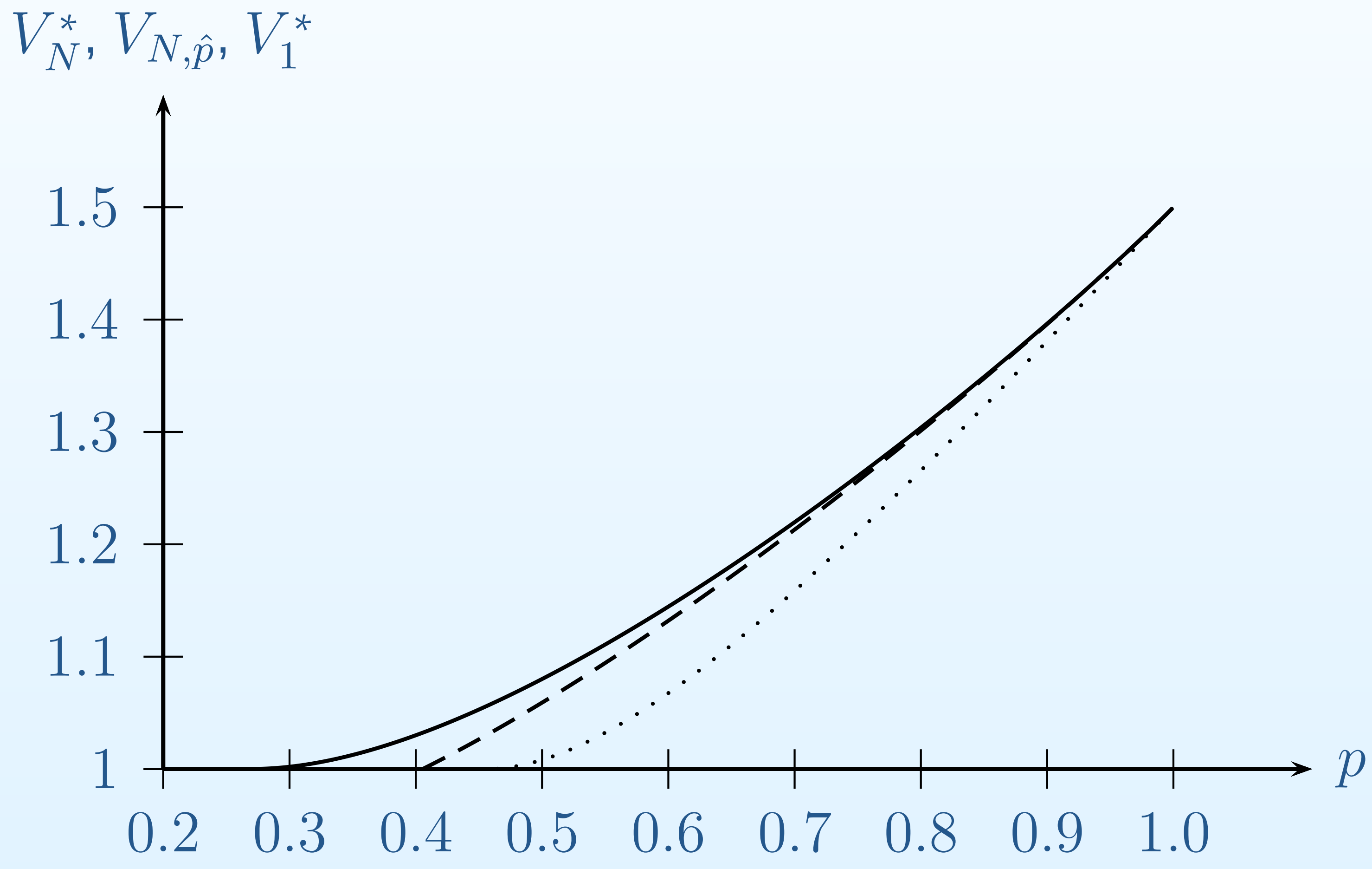
- Now a deviating player does better than s after news arrives
- Players thus set the benefit of N experiments against more than the costs of one

Extreme case:

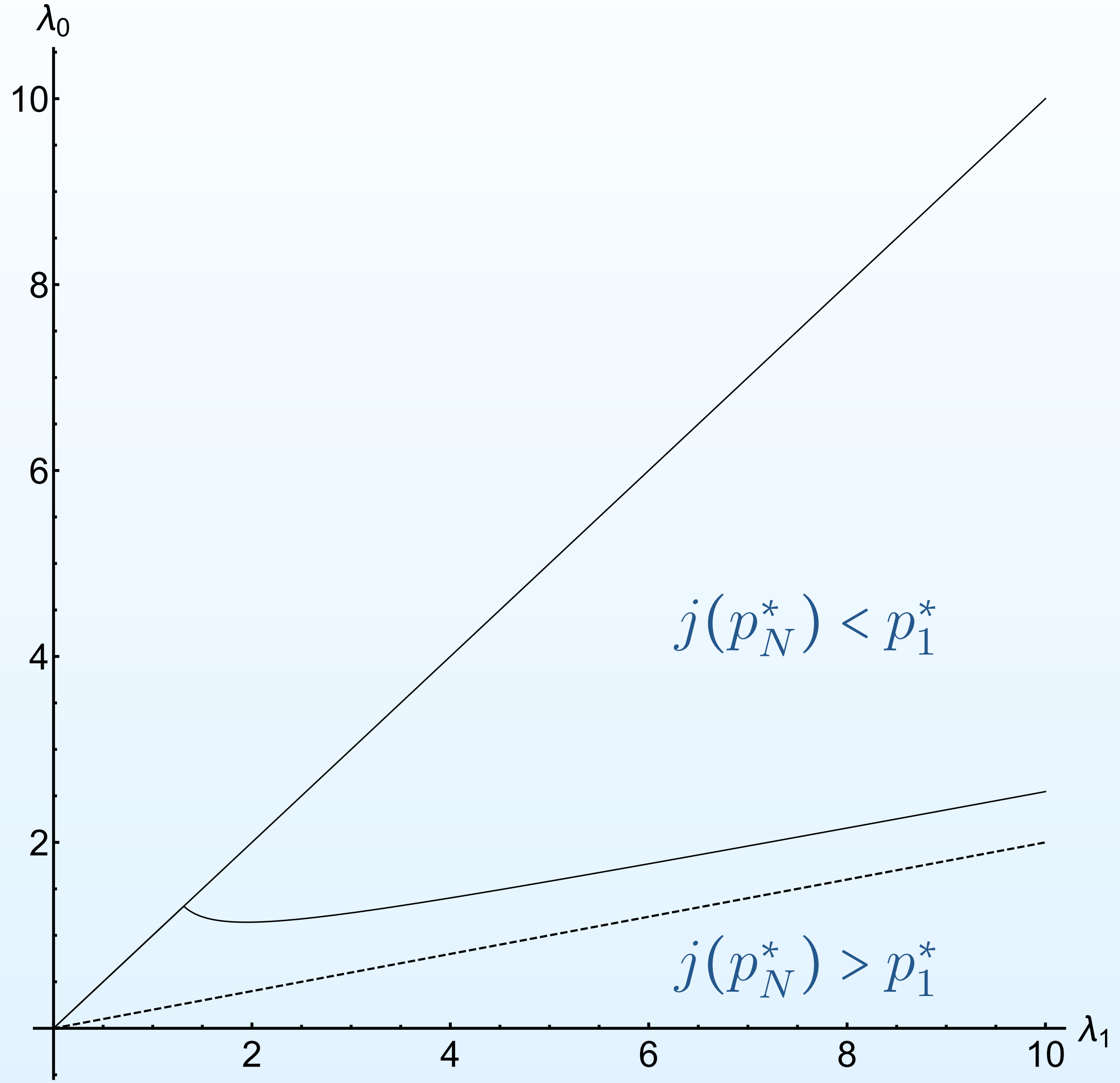
- Fully conclusive news ($\lambda_0 = 0$): $j(p_N^*) = 1$
- No scope at all for punishment after news arrives
- $\hat{p} = p_1^*$

Pure Poisson Learning ($\alpha_1 = \alpha_0$)

$(r, s, h, \lambda_1, \lambda_0, N) = (1, 1, 1.5, 1, 0.2, 5)$,
$(p_N^*, \hat{p}, p_1^*) \simeq (.27, .40, .45)$:



Pure Poisson Learning ($\alpha_1 = \alpha_0$)



Construction of Equilibria

Construct SSE for small Δ that achieve payoffs arbitrarily close to $V_{N,\hat{p}}$ and V_1^*, respectively

Two-state automaton with public randomization

Construction of Equilibria

Normal state:

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

Construction of Equilibria

Normal state:

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

Punishment state:

- Common payoff $\underline{w}^{\Delta}(p)$
- 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

Set

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

for $\underline{p}$ just above $\hat{p}$, and

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

for $\bar{p}$ just below 1

(Later on, let $\underline{p} \to \hat{p}$ and $\bar{p} \to 1$)

Verifying Enforceability

No incentives needed at beliefs $p \leq \underline{p}$ and $p > \bar{p}$

Verifying Enforceability

No incentives needed at beliefs $p \leq \underline{p}$ and $p > \bar{p}$

Incentives through terms of order 0 at beliefs $\underline{p} + \varepsilon < p \leq \bar{p}$:

$$\overline{w}^{\Delta} - \underline{w}^{\Delta} \geq \nu > 0 \text{ on } [\underline{p} + \varepsilon/2, \bar{p} + \varepsilon] \text{ for small } \Delta$$

Verifying Enforceability

No incentives needed at beliefs $p \leq \underline{p}$ and $p > \bar{p}$

Incentives through terms of order 0 at beliefs $\underline{p} + \varepsilon < p \leq \bar{p}$:

$$\overline{w}^{\Delta} - \underline{w}^{\Delta} \geq \nu > 0 \text{ on } [\underline{p} + \varepsilon/2, \bar{p} + \varepsilon] \text{ for small } \Delta$$

Incentives through terms of order > 0 at beliefs $\underline{p} < p \leq \underline{p} + \varepsilon$

Conclusion

Concluding Remarks

- **Lévy bandits**
  - Continuous time
  - News comes gradually and in ‘lumps’

Concluding Remarks

- **Lévy bandits**
  - Continuous time
  - News comes gradually and in 'lumps'
- **Discretized experimentation game**
  - Freeze actions for Δ units of time
  - Characterize range of experimentation and equilibrium payoffs as Δ vanishes
  - Construction via two-state automaton
  - Rely on closed-form solutions for continuous time
  - Asymptotic efficiency whenever there is an informative Brownian component
  - Otherwise, asymptotic efficiency if and only if news events are 'small'

Concluding Remarks

- **What else?**
  - Comparative statics of $\hat{p}$
  - Other payoff-generating processes
    - Informative lump-sum sizes
    - Lumpy bad news
  - Coupled functional equations for best/worst SSE payoffs in discrete and continuous time
  - Asymmetric PBE

Concluding Remarks

- **What else?**
  - Comparative statics of $\hat{p}$
  - Other payoff-generating processes
    - Informative lump-sum sizes
    - Lumpy bad news
  - Coupled functional equations for best/worst SSE payoffs in discrete and continuous time
  - Asymmetric PBE
- **To take away**
  - Markov perfection is truly restrictive in this class of games
  - How much, depends crucially on the payoff-generating processes