# Large Bandit Games

Antoine Salomon

## Introduction

In many situations, economic agents face a dilemma between exploiting a known profitable investment and experimenting others with unknown values. Bandit models provide a good way to study this problem. Each player faces a one-arm bandit machine (or equivalently a two-arm bandit with a risky and a safe arm) which he sequentially decides to operate or not. When the risky arm is pulled, the player gets a payoff from which he can learn about the profitability of its machine. Usually, a machine is one of two types: High and Low. When the type is High, the expected value of the risky action is positive, and negative when the type is Low. Moreover, the player is able to watch others' decisions and/or payoffs, which is another way to get information when the types of the risky arms are correlated.

Efficiency of equilibria is the main problem: will a significant proportion of players be able to learn the type of their machines? Moreover, equilibria may be inefficient because of a delay: even if players eventually know the best action, it could have taken them too much time to guess, as nobody was willing to bear the burden of experimentation or to be the first to reveal his private information. This question is linked to herding effects, which is not restricted to bandit games: if agents are led to know (or to think they probably know) the true state of the market, they will eventually all play the same action.

In the present paper we will focus on large games. Our model is close to the one studied by D. Rosenberg, E. Solan and N. Vieille [2]: time is discrete, types of the machines are perfectly correlated, the decision to move to the safe action is irreversible and each player observes one's payoffs and others' players actions. It was showed in [2] that when the number of players is getting large, equilibria look like as if there were in fact a continuum of players (see [1] for instance). Thus a fraction of agents gets bad news at the first stage and is led to move to the safe action. This reveals the type of the machines to the others as this fraction depends on it. Nevertheless, their model assumes that the information brought by a payoff can be arbitrarily bad news. In particular, it compels some players to leave even if they know that they are then revealing the state. If we relieve this assumption, players could be tempted to delay their exit or to leave far more scarcely. That is the subject of our paper.

## 1 Model and Cutoff Strategies

### 1.1 Model

Each of $N$ players sequentially operates a one-arm bandit. They have to decide when to stop, this decision being irreversible and yielding a payoff normalized to zero. At any stage $n \geq 1$, each player $i$ first decides to drop out or to stay in, then receives a payoff $X_n^i$, and finally observes who stayed in. Note that payoffs are private information, but decisions are publicly observed.
The machines have a common payoff distribution, which can be one of two possible types: High or Low. This type, denoted $\Theta$, is unknown but the players have a common prior $p_0$ which is the probability of the state being High.
We assume that, conditional on $\Theta$, the payoffs $(X_n^i)_{n\geq 1, i\in\{1,\dots,N\}}$ are i.i.d.
$\overline{\theta}$ (resp. $\underline{\theta}$) stands for the expected stage payoff of machine of type High (resp. Low) and is wlog identified with this type. To avoid trivial cases, we assume that $\underline{\theta} < 0 < \overline{\theta}$. Players discount payoffs at a common rate $\delta \in (0,1)$.
Lastly, we denote by $\mathbf{P}_\theta$ the conditional probability given $\Theta = \theta$ ($\theta \in \{\overline{\theta}, \underline{\theta}\}$).

### 1.2 Cutoff Strategies

Let us define a simple class of strategy.
To make a decision, a player may take into account her past payoffs, which partially disclose the state. To this aim, she can compute her *Private Belief*, denoted $p_n^i$:

$$p_n^i = \mathbf{P}(\Theta = \overline{\theta} | X_n^i, ..., X_1^i).$$

Assuming she has an idea of how the others players are playing, she also has to take other players' decisions into account. Let us set the r.v. $\vec{\alpha}_n$, a vector that gives the status of all players at the end of stage $n$ as follows: $j$-th coordinate $\alpha_n^j = \blacktriangle$ if player $j$ still active, $\alpha_n^j = m$ if $j$ left at stage $m$ ($m \leq n$). Moreover, a significant parameter of the $N$ player game is the number of departures before the end of stage $n$, and we will denote it $k_n^{(N)}$, i.e. $k_n^{(N)} = \#\{j \neq i, \alpha_n^j \neq \blacktriangle\}$.

Now, player $i$ can play as follows: at each stage, she computes $p_n^i$ and decides to stay only if it is above a given cut-off which depends on $n$ and on the status of the other players $\vec{\alpha}_n$.
We define cutoff strategies as a sequence $(\pi_n^i(\vec{t}_n))$ with values in $[0,1]$ indexed by the stages $n \geq 1$ and by $\vec{t}_n$, the possible vectors of status at stage $n$. Player $i$ plays the strategy if he stops at stage $\inf\{n \geq 1 : p_{n-1}^i < \pi_{n-1}^i(\vec{\alpha}_{n-1})\}$.

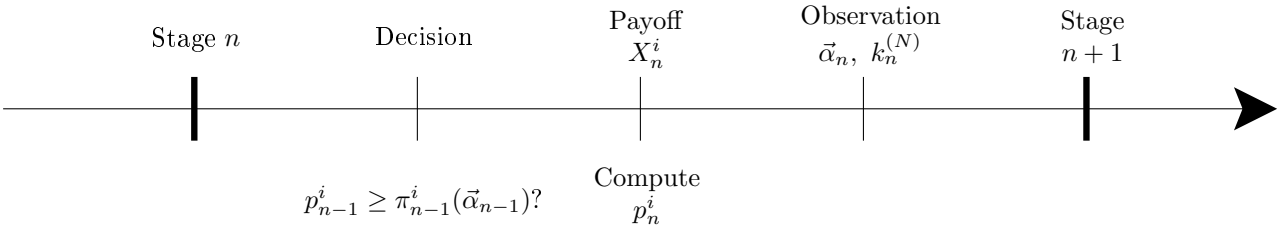


Figure 1: Progress of the game.

**Theorem.** [2] *Assume that $p_1^i$ has a density w.r.t. Lebesgue Measure.*
*Every best reply is a cutoff strategy. There exists symmetric equilibria in cutoff strategies.*

Lastly, we need to introduce the worst possible belief at stage $n$, defined as:

$$\underline{\pi}_n = \inf\{\pi \in [0,1] : F_{n,\theta}(\pi) > 0\}$$

where $F_{n,\theta}$ is the c.d.f. of $p_n^i$ under $\mathbf{P}_\theta$. As $p_1^i$ has a density, $p_n^i$ has the same support under $\mathbf{P}_{\overline{\theta}}$ and $\mathbf{P}_{\underline{\theta}}$, so that $\underline{\pi}_n$ do not depend on $\theta$.

## 2 Asymptotically Deterministic Equilibrium

**Definition 1.** *A sequence of equilibria indexed by the number of players $N$ for which each game is set is Asymptotically Deterministic with delay $n \geq 2$ if*

$$\mathbf{P}\left(k_{n-1}^{(N)} = 0\right) \xrightarrow[N\to+\infty]{} 1,\ \mathbf{P}_{\underline{\theta}}\left(k_{n+1}^{(N)} = N\right) \xrightarrow[N\to+\infty]{} 1 \text{ and } \mathbf{P}_{\overline{\theta}}\left(\forall l \geq n,\ k_n^{(N)} = k_l^{(N)}\right) \xrightarrow[N\to+\infty]{} 1.$$

The idea is that players all stay active until stage $n$, then some of them leave and the number $k_n^{(N)}$ of departures reveals the state to the remaining players. So the latter all leave in the Low state and all stay forever in the High state.

The following theorem gives the conditions for existence of an ADE with delay $n$.

To understand this theorem, we first need to introduce the cutoff $p^*$, defined by the following equation:

$$\frac{p^*\overline{\theta}}{1-\delta} + (1-p^*)\underline{\theta} = 0.$$

This is the cutoff that makes a player indifferent between staying and leaving when she is sure to learn the state at the following stage. Indeed, we have:

$$p^*\overline{\theta} + (1-p^*)\underline{\theta} + \delta\left(p^*\frac{\overline{\theta}}{1-\delta} + (1-p^*)0\right) = 0.$$

$p^*\overline{\theta}+(1-p^*)\underline{\theta}$ is the expectation of the following payoff, and $p^*\frac{\overline{\theta}}{1-\delta}+(1-p^*)0$ is the expectation of the whole payoff thereafter if the player has learned the state, as she will stay forever in the High state and drop out otherwise.

Then let us describe the conditions provided in the theorem.
Firstly, a fraction of players does leave at stage $n$ and this reveals the state to the others. Consequently, the most pessimistic belief is below $p^*$. If not, any leaving player would have better stay active one more stage as it would learn him the state and thus get him a positive average payoff. Conversely, this guarantees that a non negligible fraction of players, whose belief is below $p^*$, is compelled to leave at stage $n$. So we have a first condition.
Secondly, nobody leaves before stage $n$. The second condition below ensures that, at any stage $m < n$, even the most pessimistic player is willing to stay in.

**Theorem 2.1.** *There exists an ADE with delay $n \geq 2$ iff:*

- $\underline{\pi}_{n-1} < p^*$
- $\forall m \in \{0,1,...,n-2\},\ (1+\delta+...+\delta^{n-m-2})\left(\underline{\pi}_m\overline{\theta} + (1-\underline{\pi}_m)\underline{\theta}\right)$
  $+\ \delta^{n-m-1}\left(\underline{\pi}_m\frac{\overline{\theta}}{1-\delta}\mathbf{P}_{\overline{\theta}}(p^i_{n-1} \geq p^* | p^i_m = \underline{\pi}_m) + (1-\underline{\pi}_m)\underline{\theta}\mathbf{P}_{\underline{\theta}}(p^i_{n-1} \geq p^* | p^i_m = \underline{\pi}_m)\right) > 0.$

**Corollary 2.2.**
- *For any $n \geq 2$, there exists a one-arm bandit game for which there exists an ADE with delay $n$.*
- *There exists one-arm bandit games for which there is no ADE.*

# 3 Other asymptotic equilibria and Poisson aggregate behaviour

**Theorem 3.1.** *Let $(\Phi_N)_{N\geq 1}$ be a sequence of symmetric equilibria indexed by the number of players $N$, which is not Asymptotically Deterministic. Even if it means extracting a subsequence, we can assume that there exists a stage $n$ which is asymptotically the first stage of possible exits, i.e.:*

$$\mathbf{P}(k^{(N)}_{n-1} = 0) \xrightarrow[N\to+\infty]{} 1 \text{ and } \liminf_{N\to+\infty} \mathbf{P}(k^{(N)}_n \geq 1) > 0.$$

*Then for any value of $\theta \in \{\underline{\theta}, \overline{\theta}\}$, the average number of exits at stage $n$ is bounded and bounded away from zero. Let $\lambda_{\theta,N} = \mathbf{E}_\theta[k^{(N)}_n]$, then:*

$$0 < \liminf_{N\to+\infty} \lambda_{\theta,N} \leq \limsup_{N\to+\infty} \lambda_{\theta,N} < +\infty.$$

*Moreover, the number of exits is asymptotically a Poisson distribution:*

$$\mathbf{P}_\theta(k^{(N)}_n = k) \underset{N\to+\infty}{\sim} e^{-\lambda_{\theta,N}}\frac{\lambda^k_{\theta,N}}{k!}.$$

So if players delay their departures until stage $n$, we see that the alternative to a revealing wave of exits by a fraction of players is a bounded number of exits of the form of a Poisson distribution.

# References


[1] A. Caplin and J. Leahy (1994): Business as Usual, Market Crashes, and Wisdom After the Fact, *The American Economic Review*, Vol. 84, No. 3, 548-565.

[2] D. Rosenberg, E. Solan and N. Vielle (2007): Social learning in One-Arm Bandit Problems, *Econometrica*, Vol. 75, No. 6, 1591-1611.