Equilibria in large one-arm bandit games

A. Salomon

Université Paris 13
HEC Paris

November 25$^{\mathrm{th}}$, 2008

Model: one-arm bandits

There are N one-arm bandit machines. When operated, a machine yields a random payoff.
Nature chooses a state Θ, which can be *High* ($\Theta = \overline{\theta}$) or *Low* ($\Theta = \underline{\theta}$).
This state determines the distribution of payoff of all the N one-arm bandits, i.e. machines are perfectly correlated.
If the state is *High* the expectation of a payoff, also denoted $\overline{\theta}$, and is positive. When the state is *Low* the expected payoff, denoted $\underline{\theta}$, is negative.
The machines are operated sequentially, and conditionally to the state payoffs are i.i.d. across stages and across machines.

Model: progress of the game

Each of the N players operates her machine in discrete time and must decide when to stop. The decision to stop is irreversible and yields an outside payoff normalized to zero.

Model: progress of the game

Each of the N players operates her machine in discrete time and must decide when to stop. The decision to stop is irreversible and yields an outside payoff normalized to zero.
At any stage $n \geq 1$, for each player i the following sequence of events unfolds.

Model: progress of the game

Each of the N players operates her machine in discrete time and must decide when to stop. The decision to stop is irreversible and yields an outside payoff normalized to zero.
At any stage $n \geq 1$, for each player i the following sequence of events unfolds.

1. She decides to drop out or to stay in.

Model: progress of the game

Each of the N players operates her machine in discrete time and must decide when to stop. The decision to stop is irreversible and yields an outside payoff normalized to zero.

At any stage $n \geq 1$, for each player i the following sequence of events unfolds.

1. She decides to drop out or to stay in.
2. If she stays, she gets and observes a random payoff X_n^i

Model: progress of the game

Each of the N players operates her machine in discrete time and must decide when to stop. The decision to stop is irreversible and yields an outside payoff normalized to zero.

At any stage $n \geq 1$, for each player i the following sequence of events unfolds.

1. She decides to drop out or to stay in.
2. If she stays, she gets and observes a random payoff X_n^i
3. She observes who stayed in.

Model: progress of the game

Each of the N players operates her machine in discrete time and must decide when to stop. The decision to stop is irreversible and yields an outside payoff normalized to zero.
At any stage $n \geq 1$, for each player i the following sequence of events unfolds.

1. She decides to drop out or to stay in.
2. If she stays, she gets and observes a random payoff X_n^i
3. She observes who stayed in.

Pay-offs are private information, but decisions are publicly observed.
Players discount payoffs at rate δ: if player stop at stage τ, her overall payoff is $\sum_{n=1}^{\tau-1} \delta^{n-1} X_n^i$.

Cutoff Strategies

At the beginning, all players have the same prior about the state of the world: $p_0 = \mathbf{P}(\Theta = \overline{\theta})$. After each payoff, they can update this prior to get a *private belief*: $p_n^i = \mathbf{P}(\Theta = \overline{\theta} | X_1^i, X_2^i, ..., X_n^i)$.

Cutoff Strategies

At the beginning, all players have the same prior about the state of the world: $p_0 = \mathbf{P}(\Theta = \overline{\theta})$. After each payoff, they can update this prior to get a *private belief*: $p_n^i = \mathbf{P}(\Theta = \overline{\theta} | X_1^i, X_2^i, ..., X_n^i)$.
They also learn from others' behaviours.

Cutoff Strategies

At the beginning, all players have the same prior about the state of the world: $p_0 = \mathbf{P}(\Theta = \overline{\theta})$. After each payoff, they can update this prior to get a *private belief*: $p_n^i = \mathbf{P}(\Theta = \overline{\theta} | X_1^i, X_2^i, ..., X_n^i)$.
They also learn from others' behaviours.
At a given stage, the status of the players can be summed up in a random vector $\vec{\alpha}_n$, with $\alpha_n^i = \begin{cases} \blacktriangle \text{ if } i \text{ is still active} \\ k \text{ if } i \text{ exited at stage } k \leq n \end{cases}$.

Cutoff Strategies

At the beginning, all players have the same prior about the state of the world: $p_0 = \mathbf{P}(\Theta = \overline{\theta})$. After each payoff, they can update this prior to get a *private belief*: $p_n^i = \mathbf{P}(\Theta = \overline{\theta} | X_1^i, X_2^i, ..., X_n^i)$.
They also learn from others' behaviours.
At a given stage, the status of the players can be summed up in a random vector $\vec{\alpha}_n$, with $\alpha_n^i = \begin{cases} \blacktriangle \text{ if } i \text{ is still active} \\ k \text{ if } i \text{ exited at stage } k \leq n \end{cases}$.
A possible strategy for player i is to set *a priori* a sequence of *cutoffs* $\pi_n(\vec{\alpha})$ and to stop the game as soon as $p_{n-1}^i < \pi_n(\vec{\alpha}_{n-1})$.

Cutoff Strategies

At the beginning, all players have the same prior about the state of the world: $p_0 = \mathbf{P}(\Theta = \overline{\theta})$. After each payoff, they can update this prior to get a *private belief*: $p_n^i = \mathbf{P}(\Theta = \overline{\theta} | X_1^i, X_2^i, ..., X_n^i)$.
They also learn from others' behaviours.
At a given stage, the status of the players can be summed up in a random vector $\vec{\alpha}_n$, with $\alpha_n^i = \begin{cases} \blacktriangle \text{ if } i \text{ is still active} \\ k \text{ if } i \text{ exited at stage } k \leq n \end{cases}$.
A possible strategy for player i is to set *a priori* a sequence of *cutoffs* $\pi_n(\vec{\alpha})$ and to stop the game as soon as $p_{n-1}^i < \pi_n(\vec{\alpha}_{n-1})$.

Theorem [D.Rosenberg, E.Solan, N.Vieille]

Assume that the law of p_1^i has a density. Every best reply is a cutoff strategy.
There exists symmetric equilibria in cutoff strategy.

Question

What happens when the number of players N goes to $+\infty$?

A first large game result.

Let us define the cutoff p^* as: $p^* \frac{\overline{\theta}}{1-\delta} + (1 - p^*)\underline{\theta} = 0$.
This is the cutoff that makes a player indifferent between leaving or staying one more stage and being told the state.

A first large game result.

Let us define the cutoff p^* as: $p^* \frac{\overline{\theta}}{1-\delta} + (1 - p^*)\underline{\theta} = 0$.
This is the cutoff that makes a player indifferent between leaving or staying one more stage and being told the state.

Theorem [D.Rosenberg, E.Solan, N.Vieille]

Assume that p_1^i has full support. In equilibria, as the number of players N gets large, all players tends to play with cutoff p^* after the first payoff:
$\sup_{i \in \{1,2,\dots,N\}} |\pi_1^i(\vec{\blacktriangle}) - p^*| \xrightarrow[N \to +\infty]{} 0$.

A first large game result.

Let us define the cutoff p^* as: $p^* \frac{\overline{\theta}}{1-\delta} + (1 - p^*)\underline{\theta} = 0.$
This is the cutoff that makes a player indifferent between leaving or staying one more stage and being told the state.

Theorem [D.Rosenberg, E.Solan, N.Vieille]

Assume that p_1^i has full support. In equilibria, as the number of players N gets large, all players tends to play with cutoff p^* after the first payoff:
$\sup_{i\in\{1,2,\ldots,N\}} |\pi_1^i(\vec{\blacktriangle}) - p^*| \xrightarrow[N\to+\infty]{} 0.$

We will call these equilibria asymptotically deterministic: as N becomes large, players can determine the state after the first stage. Indeed, the proportion of players who leave after the first payoff reveals the state by the Law of Large Number.

A first large game result.

Let us define the cutoff p^* as: $p^* \frac{\overline{\theta}}{1-\delta} + (1 - p^*)\underline{\theta} = 0$.
This is the cutoff that makes a player indifferent between leaving or staying one more stage and being told the state.

Theorem [D.Rosenberg, E.Solan, N.Vieille]

Assume that p_1^i has full support. In equilibria, as the number of players N gets large, all players tends to play with cutoff p^* after the first payoff:
$\sup_{i\in\{1,2,\ldots,N\}} \left|\pi_1^i(\vec{\blacktriangle}) - p^*\right| \xrightarrow[N\to+\infty]{} 0$.

Idea: without the full support assumption, this process could be delayed.

All equilibria are not asymptotically deterministic: counter-example.

We relax the full support assumption.
Let us denote $\underline{\pi}_n$ the worst possible private belief at stage n, i.e. the smallest possible value of p_n^i.
For example, assume that:

All equilibria are not asymptotically deterministic: counter-example.

We relax the full support assumption.
Let us denote $\underline{\pi}_n$ the worst possible private belief at stage n, i.e. the smallest possible value of p_n^i.
For example, assume that:

- Every player can afford to stay after the first payoff if they are able to learn the state thereafter: $\underline{\pi}_1\overline{\theta} + (1 - \underline{\pi}_1)\underline{\theta} + \delta\underline{\pi}_1\frac{\overline{\theta}}{1-\delta} > 0$ i.e. $\underline{\pi}_1 > p^*$.

All equilibria are not asymptotically deterministic: counter-example.

We relax the full support assumption.
Let us denote $\underline{\pi}_n$ the worst possible private belief at stage n, i.e. the smallest possible value of p_n^i.
For example, assume that:

- Every player can afford to stay after the first payoff if they are able to learn the state thereafter:
$\underline{\pi}_1\overline{\theta} + (1 - \underline{\pi}_1)\underline{\theta} + \delta\underline{\pi}_1\frac{\overline{\theta}}{1-\delta} > 0$ i.e. $\underline{\pi}_1 > p^*$.
- Most pessimistic players can not afford to stay for two stages even if they are then able to learn the state:
$(1 + \delta)\left(\underline{\pi}_1\overline{\theta} + (1 - \underline{\pi}_1)\underline{\theta}\right) + \delta^2\underline{\pi}_1\frac{\overline{\theta}}{1-\delta} < 0$

All equilibria are not asymptotically deterministic: counter-example.

We relax the full support assumption.
Let us denote $\underline{\pi}_n$ the worst possible private belief at stage n, i.e. the smallest possible value of p_n^i.
For example, assume that:

- Every player can afford to stay after the first payoff if they are able to learn the state thereafter:
  $\underline{\pi}_1\overline{\theta} + (1 - \underline{\pi}_1)\underline{\theta} + \delta\underline{\pi}_1\frac{\overline{\theta}}{1-\delta} > 0$ i.e. $\underline{\pi}_1 > p^*$.
- Most pessimistic players can not afford to stay for two stages even if they are then able to learn the state:
  $(1+\delta)\left(\underline{\pi}_1\overline{\theta} + (1 - \underline{\pi}_1)\underline{\theta}\right) + \delta^2\underline{\pi}_1\frac{\overline{\theta}}{1-\delta} < 0$

There can't be ADE in this situation.

Definition

A sequence of equilibria (Φ_N) is an Asymptotically Deterministic Equilibrium with delay n if, as $N \longrightarrow +\infty$:

- There are no exits until stage n.
- After stage n, every remaining players stay forever if the state is *High*, or leave if the state is *Low*.
- $n \geq 2$.

Theorem

There exists ADE with delay n if and only if:

- n is the smallest integer such that $\underline{\pi}_{n-1} < p^*$.
- Before stage n, even the most pessimistic player can afford to wait until stage n when the state will be revealed.

Non-deterministic equilibria

There exists asymptotic equilibria which are not deterministic (see previous counter-example).

The typical progress of the game in such equilibria is as follows:

Non-deterministic equilibria

There exists asymptotic equilibria which are not deterministic (see previous counter-example).

The typical progress of the game in such equilibria is as follows:

- During the first stages, players stay no matter what their beliefs could be.

Non-deterministic equilibria

There exists asymptotic equilibria which are not deterministic (see previous counter-example).

The typical progress of the game in such equilibria is as follows:

- During the first stages, players stay no matter what their beliefs could be.
- Then, some players may have to leave because they get too bad news. The proportion of exits must go to 0 as $N \to +\infty$.

Non-deterministic equilibria

There exists asymptotic equilibria which are not deterministic (see previous counter-example).

The typical progress of the game in such equilibria is as follows:

- During the first stages, players stay no matter what their beliefs could be.
- Then, some players may have to leave because they get too bad news. The proportion of exits must go to 0 as $N \to +\infty$.

In fact, the average number of exits is bounded as $N \to +\infty$.

Non-deterministic equilibria

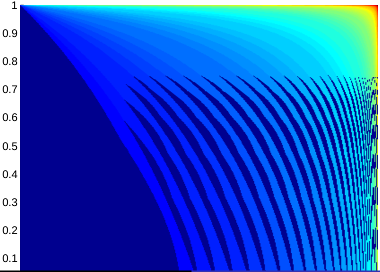
There exists asymptotic equilibria which are not deterministic (see previous counter-example).
The typical progress of the game in such equilibria is as follows:

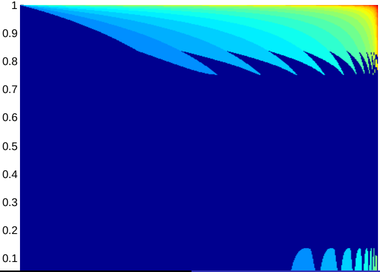
- During the first stages, players stay no matter what their beliefs could be.
- Then, some players may have to leave because they get too bad news. The proportion of exits must go to 0 as $N \to +\infty$.

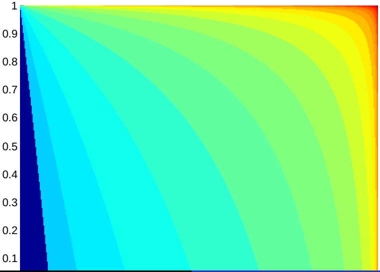
In fact, the average number of exits is bounded as $N \to +\infty$.
In particular, if the equilibria are symmetric, the probability for a player to leave is of order $\frac{1}{N}$. The average numbers of exits converge weakly to a Poisson distributions.

An example.

We compute the conditions of existence of ADE when the payoffs are of the form $X - 1$, $X \sim Exp(\lambda)$.
X-axis is the prior p_0.
Y-axis is the discount rate δ.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1