THE CONTROLLED BIASED COIN PROBLEM

OLIVIER GOSSNER AND TRISTAN TOMALA

Abstract. We study the maxmin value of a zero-sum repeated game where player 1 is restricted to pure strategies but privately observes the realizations of some random variables. This kind of problem was introduced by Gossner and Vieille [GV02] in the case where player 1 observes i.i.d. random variables. The paper solves the case where the law of the random variable (the coin) depends on player 1’s action. We also discuss the general case where the law of the coin is controlled by both players.

1. Model and definitions

1.1. The repeated game. Let \((A, B, g)\) be a zero-sum game where, \(A\) (resp. \(B\)) is the finite set of actions of player 1 (resp.2) and \(g: A \times B \rightarrow \mathbb{R}\) is the payoff function. Let \(S\) be a finite set of signals and \(\mu: A \rightarrow \Delta(S)\) be a transition probability. The repeated game unfolds as follows. At each stage \(t = 1, 2, \ldots\), each player chooses an action in her own set of actions and if \((a, b) \in A \times B\) is the action profile played, the payoff for player 1 is \(g(a, b)\). A signal \(s\) is then drawn according to \(\mu(a)\) and announced to player 1 only.

A history of length \(n\) of the game [resp. for player 2] is an element \(h_n\) of \(H_n = (A \times B \times S)^{\leq n}\), [resp. \(h^2_n\) of \(H^2_n = (A \times B)^{\leq n}\)]. A pure strategy \(\sigma\) for player 1 is a sequence \((\sigma_n)_{n \geq 0}\) with \(\sigma_n: H_n \rightarrow A\). A behavioral strategy \(\tau\) for player 2 is a sequence \((\tau_n)_{n \geq 0}\) with \(\tau_n: H^2_n \rightarrow \Delta(B)\), where \(\Delta(B)\) denotes the set of probabilities on \(B\). A pair of strategies \((\sigma, \tau)\) with \(\sigma\) pure and \(\tau\) behavioral induce a probability distribution \(P_{\sigma, \tau}\) on the set of plays \((A \times B)\) endowed with the product \(\sigma\)-algebra. The average payoff up to stage \(n\) is: \(\gamma_n(\sigma, \tau) = E_{\sigma, \tau}[\frac{1}{n} \sum_{m=1}^{n} g(a_m, b_m)]\) where \((a_m, b_m)\) is the random pair of actions at stage \(n\). The uniform max min payoff of the repeated game is defined as follows (see [GV02] for this model and [MSZ94] for general repeated games):

Definition 1. (1) Player 1 guarantees \(v \in \mathbb{R}\) if:

\[\forall \varepsilon > 0, \exists N, \forall n \geq N, \gamma_n(\sigma, \tau) \geq v - \varepsilon.\]

(2) Player 2 defends \(v \in \mathbb{R}\) if:

\[\forall \varepsilon > 0, \forall \sigma, \exists N, \forall n \geq N, \gamma_n(\sigma, \tau) \leq v + \varepsilon.\]

(3) The max min is \(v_\infty \in \mathbb{R}\) such that I guarantees \(v_\infty\) and II defends \(v_\infty\).

1.2. Information theory tools. Let \(x\) be a finite random variable with law \(P\).

Throughout the paper, we write \(\log\) to denote the logarithm with base 2. By definition, the entropy of \(x\) is:

\[H(x) = -E[\log P(x)] = -\sum_{x} P(x) \log P(x)\]

Note that \(H(x)\) is non-negative and depends only on the law \(P\) of \(x\). One can therefore define the entropy of a probability distribution \(P = (p_k)_{k=1,\ldots,L}\) with finite support by \(H(P) = -\sum_{k=1}^{L} p_k \log p_k\).

Date: 3rd February 2005.
Let \((x, y)\) be a couple of finite random variables with joint law \(P\). Define \(P(x|y) = P(x = x|y = y)\) if \(P(y = y) > 0\) and arbitrarily otherwise. The conditional entropy of \(x\) given \(y\) is:
\[
H(x | y) = -\mathbb{E}[\log P(x | y)]
\]
One has the following chain rule:
\[
H(x, y) = H(y) + H(x | y)
\]

2. The Characterization

Although, player 1 is restricted to pure strategies, since player 2 does not observe the signals of player 1, she holds a belief on the action of player 1, that is a probability distribution on \(A\). Assume this belief to be \(x\) about the action of player 1 at stage \(n\). At stage \(n\), player 2 gets to observe the random action \(a\) of player 1 but not the random signal \(s\) observed by player 1. Define the entropy variation associated to \(x\) as \(H(s | a) - H(a)\) which writes:
\[
\Delta H(x) = \sum_a x(a) H(q(\cdot | a)) - H(x)
\]
This represents the variation of uncertainty for player 2 at stage \(n\).

For \(x \in \Delta(A)\), the optimal payoff associated to \(x\) is:
\[
\pi(x) = \min_b g(x, b)
\]
where \(g\) is extended to mixed actions by linearity. This function represents the expected payoff obtained when player 2 plays her best reply to her belief about the action of player 1. Let then:
\[
V = \{(\Delta H(x), \pi(x)) | x \in \Delta(A)\}
\]
and
\[
\nu^* = \sup\{x_2 \in \mathbb{R} | (x_1, x_2) \in \text{co} V, x_1 \geq 0\}
\]
This is the highest payoff associated to a convex combination of mixed actions under the constraint that the average entropy variation is non-negative. Note that \(\Delta H\) and \(\pi\) are continuous on \(\Delta(A)\) and thus \(V\) is compact. For every \(x\) such that \(H(x) = 0, \Delta H(x) \geq 0\) thus \(V \cap \{x_1 \geq 0\} \neq \emptyset\) and the supremum that defines \(\nu^*\) is a maximum.

We give now other expressions of \(\nu^*\) which are more convenient for computations. First we show how to express the number \(\nu^*\) through simple optimization problems. Define for each real number \(h\):
\[
u(h) = \max\{\pi(x) | x \in \Delta(A), \Delta H(x) \geq h\}
\]
From the definition of \(V\) we have for each \(h\):
\[
u(h) = \max\{x_2 | (x_1, x_2) \in \text{co} V, x_1 \geq h\}
\]
Since \(V\) is compact, \(\nu(h)\) is well defined. Let \(\text{cav} u\) be the least concave function pointwise greater than \(u\). Then:
\[
\text{cav} u(0) = \inf_{\lambda \geq 0} \max x \min b \{g(x, b) + \lambda \Delta H(x)\}
\]
To see this, note that \(u\) is upper-semi-continuous, non-increasing and the hypograph of \(u\) is the comprehensive set \(V - \mathbb{R}^2_+\) associated to \(V\). Then \(\text{cav} u\) is thus also non-increasing, u.s.c. and its hypograph is \(\text{cav} V - \mathbb{R}^2_+\). Now using Fenchel’s duality we get:

Lemma 2.
\[
\text{cav} u(0) = \inf_{\lambda \geq 0} \max x \min b \{g(x, b) + \lambda \Delta H(x)\}
\]
Proof. $\text{cav } u(0) = -(-u)^*(0)$, where $F^*$ is the Fenchel-conjugate of $F$. By definition of the Fenchel conjugate:

$$(-u)^*(\lambda) = \sup_h \{h\lambda - (-u(h))\}$$

$$(-u)^*(h) = \sup_{\lambda} \{h\lambda - (-u)^*(\lambda)\}$$

Thus $\text{cav } u(0) = \inf_{\lambda} \sup_h \{h\lambda + u(h)\}$ and by the definition of $u$:

$$\text{cav } u(0) = \inf_{\lambda} \max_{x} \sup_{h \leq \Delta H(x)} \{h\lambda + \pi(x)\}$$

For $\lambda < 0$, we let $h$ go to $-\infty$ and the sup is infinite. Thus we restrict to $\lambda \geq 0$ which gives:

$$\text{cav } u(0) = \inf_{\lambda \geq 0} \max_{x} \sup_{h \leq \Delta H(x)} \min_b \{h\lambda + g(x, b)\}$$

$$= \inf_{\lambda \geq 0} \min_{x} \sup_b \{\lambda \Delta H(x) + g(x, b)\}$$

concluding the proof. \hfill \square

We have then the following result:

**Theorem 3.** The uniform minmax $v_\infty$ exists and $v_\infty = v^*$.

3. Proof of Theorem 3

3.1. Player 2 defends $v^*$. The argument used here is a generalization of the one found in [GV02]. We introduce some notations. Let $\sigma$ be a strategy of player 1. Define inductively $\tau_1$ as the strategy of player 2 that plays stage-best replies to $\sigma$. At stage 1, $\tau_1(\emptyset) \in \arg\min_b g(\sigma(\emptyset), b)$ where $\emptyset$ is the null history that starts the game. Assume that for $h^2_n$ of player 2, let $x_{n+1}(h^2_n)$ be the distribution of the action of player 1 given $h^2_n$ and let $\tau_1(h^2_n)$ be in $\arg\min_b g(x_{n+1}(h^2_n), b)$.

**Lemma 4.** For each integer $n$ and strategy $\sigma$:

$$\gamma_n(\sigma, \tau_1) \leq v^*$$

**Proof.** Let $\sigma$ be a strategy for the team and set $\tau = \tau_1$. Let $a_m, b_n, s_n$ be the sequences of random action profiles and signals associated to $(\sigma, \tau)$, $h^2_n$ the random history of player 2 and $x_{n+1} = x_{n+1}(h^2_n)$. The expected payoff at stage $m+1$ after $h^2_m$ is $\max_b g(x_{m+1}(h^2_m), b) = \pi(x_{m+1}(h^2_m))$ from the definition of $\tau$ and thus $\gamma_n(\sigma, \tau) = \mathbb{E}_{\sigma, \tau}[\sum_{m=1}^{n} \pi(x_m)]$.

We compute now the entropy variations, $\Delta H(x_{m+1}(h^2_m))$. By definition of $x_{m+1}$,

$$\Delta H(x_{m+1}(h^2_m)) = H(a_{m+1}|h^2_m) - H(s_{m+1}|a_{m+1}, h^2_m)$$

and thus:

$$\mathbb{E}_{\sigma, \tau} \Delta H(x_{m+1}) = H(a_{m+1}|h^2_m) - H(s_{m+1}|a_{m+1})$$

Denote $H_m = H(s_1, \ldots, s_m|a_1, b_1, \ldots, a_m, b_m)$. Applying the additivity formula in two different ways:

$$H(s_1, \ldots, s_m, a_{m+1}, b_{m+1}, s_{m+1}, h^2_m) = H_m + H(b_{m+1}|h^2_m) + H(s_{m+1}|a_{m+1})$$

$$= H_{m+1} + H(b_{m+1}|h^2_m) + H(a_{m+1}|h^2_m)$$

Thus cav
where we use that \((a_{m+1}, b_{m+1})\) are independent conditional on \(h^2_m\), \(a_{m+1}\) is
deterministic given \(h^2_m\) and \(s_{m+1}\) depends on \(a_{m+1}\) only. Thus, \(E_{\sigma, V}\Delta H(x_{m+1}) = H_{m+1} - H_m\) and,
\[
E_{\sigma, V} \frac{1}{n} \sum_{m \leq n} \Delta H(x_m) = \frac{1}{n} H_n \geq 0
\]
Therefore the vector \((E_{\sigma, V} \frac{1}{n} \sum_{m \leq n} \Delta H(x_m), \gamma_n(\sigma, \tau))\) belongs to \(\text{co} V \cap \{x_1 \geq 0\}\).
\(\square\)

3.2. Player 1 guarantees \(v^*\). Call a strategy of player 1 autonomous if it does not depend on player 2’s moves.

**Lemma 5.** Let \(\sigma\) be an autonomous strategy, for each stage \(n\) and strategy \(\tau\) for player 2, \(\gamma_n(\sigma, \tau) \geq \gamma_n(\sigma, \tau)\).

**Proof.** Straightforward since player 2’s moves do not influence the actions of player 1.
\(\square\)

**Corollary 6.** If \(\forall \varepsilon > 0\) there exists an autonomous strategy \(\sigma\) and an integer \(N\) such that \(\forall n \geq N\), \(\gamma_n(\sigma, \tau) \geq v^* - \varepsilon\), then the player 1 guarantees \(v^*\).

An autonomous strategy \(\sigma\) maps \(\cup_n (A \times S)^n\) to \(A\) and thus generates a probability distribution \(P_{\sigma}\) on \((A \times S)^\infty\). Conversely, call a distribution \(P\) on \((A \times S)^\infty\) an \(A\)-distribution if for each \(n\) and for \(P\)-almost every history \(h_n \in (A \times S)^n\), the distribution of \(a_{n+1}\) conditional on \(h_n\) is a Dirac measure. Every \(A\)-distribution induces an autonomous strategy.

Given an autonomous \(\sigma\), consider the belief of player 2 at stage \(n\): given \(h^2_n\), let \(x_{n+1}\) be the distribution of \(\sigma(a_1, s_1, \ldots, a_n, s_n)\) given \(h^2_n\). The empirical distribution of beliefs up to stage \(n\) as the random variable:

\[
d_n = \frac{1}{n} \sum_{m \leq n} \epsilon_{x_m}
\]
where \(\epsilon_x\) denotes the Dirac measure on \(x\). The random variable \(d_n\) has values in \(D = \Delta(\Delta(A))\) and we let \(\delta = E_{\sigma}(d_n)\) be the element of \(D\) such that for every real-valued continuous function \(f\) on \(\Delta(A)\), \(E_{\sigma}(\int f(x)d\delta_n(x)) = \int f(x)d\delta(x)\). We get:

\[
\gamma_n(\sigma, \tau) = E_{\sigma} \pi
\]

Empirical distributions of beliefs are studied in [GT04] were we prove the following result:

**Theorem 7.** [GT04]. For every \(\delta \in \Delta(\Delta(A))\) such that \(E_{\delta} \Delta H \geq 0\), there exists an \(A\)-distribution \(P\) such that \(E_{P}d_n\) weak* converges to \(\delta\).

It follows readily:

**Lemma 8.** Player 1 guarantees \(\sup \{E_{\delta} \pi \mid \delta \in \Delta(\Delta(A)), E_{\delta} \Delta H \geq 0\}\).

We conclude the proof by the following lemma:

**Lemma 9.** \(\sup \{E_{\delta} \pi \mid \delta \in \Delta(\Delta(A)), E_{\delta} \Delta H \geq 0\} = v^*\).

**Proof.** Immediate since the set of vectors \((E_{\delta} \Delta H, E_{\delta} \pi)\) as \(\delta\) varies in \(\Delta(A)\) is co\(V\).  
\(\square\)
4. 2 × 2 Games

We study 2 × 2 games and give an explicit formula for $v^*$. Consider the game $G$ whose matrix is displayed below. The set of signals is $\{0, 1\}$, $\mu(0|T) = p$, $\mu(0|B) = q$, where w.l.o.g. $0 \leq p \leq q \leq \frac{1}{2}$.

\[
\begin{array}{ccc}
T & B \\
\hline \\
L & (a, b) & p \\
R & (c, d) & q \\
\end{array}
\]

Let us assume that the game has no value in pure strategies, that is w.l.o.g. $\mu$ whose matrix is displayed below. The set of signals is $G$ for payoff game. At each stage $\cdot|\cdot$. Player 1 guarantees $\cdot|\cdot$. Heuristic Result. The payoff vector is $(\Delta H(a|b) = 1, b = c = 0, h_0 = \frac{H(q) + H(p)}{2} - 1$ and $v^* = \frac{H(q)}{2 + H(q) - H(p)}$

- If $p = q$, $v^* = \frac{H(q)}{2}$, which is the result of [GV02].
- If $p = 0, q = \frac{1}{2}$, $v^* = \frac{1}{2} = \frac{2}{3}$ which means that player 1 can get the value of $G$ two thirds of the time. This result can be proven directly by considering the following idea of strategy. Play $T$ for 2n stages so as to generate 2n fair coins. This is the trigerring phase. Play then the mixed strategy $(\frac{1}{2}, \frac{1}{2})$ i.i.d. for 2n stages, using the 2n coins. This is the main phase. With high probability, during the main phase player 1 plays $T$ n times and generates $n$ coins. Player 1 then plays $T$ n times so as to get 2n coins again and the main phase can be restarted. With high probability, this process can be repeated a large number of times. The trigerring phase becomes payoff unsignificant and the average payoff is $\frac{1}{2} \frac{2n + 0.0}{3n}$.

5. The bi-controlled case

We analyze here the case where the law of the coin depends on both actions, i.e. the data of the game are the same except for the transition $\mu$ which maps now $A \times B$ to $\Delta(S)$. The definitions of strategies and maxmin of the game are unchanged. For each $x \in \Delta(A)$ and $b \in B$, set:

$$\Delta H(x, b) = \sum_a x(a)H(q, x, b) - H(x)$$

5.1. A heuristic analysis. We define a heuristic vector payoff game as follows: At each stage $u \in [1, 2, \ldots]$, player 1 chooses $x \in \Delta(A)$, player 2 chooses $b \in B$ and the payoff vector is $(\Delta H(x, b), g(x, b))$. The payoff vector is publicly observable.

**Heuristic Result.** Player 1 guarantees $v$ if and only if the set of payoffs $\{a = (u_1, u_2) \in \mathbb{R}^2 | u_1 \geq 0, u_2 \geq v\}$ is approachable ([Bl56]) by player 1 in the vector payoff game.
Proof. (Heuristic). As in the proof of theorem 3, a pair of strategies in the game induces at each stage a belief of player 2 regarding the next action of player 1: let $x_{n+1}$ be the distribution of $\sigma(a_1, s_1, \ldots, a_n, s_n)$ given $h_2^n$. One may again define the empirical distribution of beliefs up to stage $n$ but this quantity will depend on both players’ strategies. The heuristic is to generalize [GT04]’s result to the case where the beliefs of player 2 depend on his own action. This is subject to ongoing research. The main idea is to replace the feasibility condition $E_{\delta} \Delta H \geq 0$ by an approachability condition of the set of distributions $\delta$ that verify such conditions.

□

We deduce a characterization of the maxmin value:

**Corollary 10.** The maxmin value of the game is the largest number $v^*$ such that

$$\{u = (u_1, u_2) \in \mathbb{R}^2 \mid u_1 \geq 0, u_2 \geq v^*\}$$

and:

$$v^* = \inf_{\lambda \geq 0} \max_x \min_b \{g(x, b) + \lambda \Delta H(x, b)\}$$

**Proof.** The first part of the corollary is a direct consequence of the heuristic result. Using Blackwell’s necessary and sufficient conditions for approachability of convex sets we get that $\{u = (u_1, u_2) \in \mathbb{R}^2 \mid u_1 \geq 0, u_2 \geq v^*\}$ is approachable by player 1 if and only if for each $p \in [0, 1]$

$$pv + (1 - p)0 \leq \max_x \min_b \{pg(x, b) + (1 - p)\Delta H(x, b)\}$$

For $p = 0$ this says $\max_x \min_b \Delta H(x, b) \geq 0$ which is verified (take $x$ to be a pure action). Assume thus $p > 0$. Dividing by $p$ and setting $\lambda = \frac{1 - p}{p}$, the condition becomes that for each $\lambda \geq 0$,

$$v \leq \max_x \min_b \{g(x, b) + \lambda \Delta H(x, b)\}$$

The highest $v$ that verifies this is clearly:

$$v^* = \inf_{\lambda \geq 0} \max_x \min_b \{g(x, b) + \lambda \Delta H(x, b)\}$$

□

5.2. Matching pennies. We consider matching pennies and use the above formula to compare two situations: (1) The coin is fully controlled by player 2 and is deterministic if Left and fair if Right; (2) The coin is fully controlled by player 1 and is deterministic if Top and fair if Bottom. The two situations are displayed below. What is the most favourable situation for player 1?

In the second case, we already know that the maxmin value is $\frac{1}{3}$. We use now our formula to compute the value in the first case.

**Proposition 11.** In the first case, $v^*$ is the unique solution in $[0, 1)$ of the equation:

$$\lambda = \frac{1 + \lambda}{2} - \lambda H\left(\frac{1 + \lambda}{2}\right)$$

Numerically: $0.355 < v^* < 0.356$.

It is therefore better for player 1 that the coin is controlled by player 2.
Proof. We wish to compute:
\[
\inf_{\lambda \geq 0} \max_x \min_b \{ g(x, b) + \lambda \Delta H(x, b) \}
\]
which equals:
\[
\inf_{\lambda \geq 0} \max_x \min_b \{ x - \lambda H(x); (1 - x) + \lambda (1 - H(x)) \}
\]
Set \( F(\lambda) = \max_x \min_b \{ g(x, b) + \lambda \Delta H(x, b) \} \). Given \( x \) and \( \lambda \), the minimum is achieved at \( L \) iff \( x \leq 1 + \lambda/2 \) which holds for all \( \lambda \geq 1 \). For each such \( \lambda \), we solve \( \max_x (x - \lambda H(x)) \). This expression being convex, the maximum is achieved at \( x = 1 \), thus \( F(\lambda) = 1, \forall \lambda \geq 1 \).

Assume now \( \lambda \leq 1 \). Since both expressions \( x - \lambda H(x) \) and \( (1 - x) + \lambda (1 - H(x)) \) are convex in \( x \), the maximum of \( \min \{ x - \lambda H(x); (1 - x) + \lambda (1 - H(x)) \} \) is achieved either at 0, at 1 or at \( x \) that equalizes these expressions. One easily deduces that:
\[
F(\lambda) = \max \left\{ \lambda; \frac{1 + \lambda}{2} - \lambda H\left( \frac{1 + \lambda}{2} \right) \right\}
\]
We let \( f(\lambda) = \frac{1 + \lambda}{2} - \lambda H\left( \frac{1 + \lambda}{2} \right) \) and studying the variations of \( f \) we find that there exits a unique \( \lambda_0 < 1 \) such that \( f(\lambda_0) = \lambda_0 \) and \( f(\lambda) \geq \lambda \Leftrightarrow \lambda \geq \lambda_0 \). Numerically, \( 0.355 < \lambda_0 < 0.356 \). Thus,\[
F(\lambda) = \begin{cases} 
  f(\lambda) & 0 \leq \lambda \leq \lambda_0 \\
  \lambda & \lambda_0 \leq \lambda \leq 1
\end{cases}
\]
It follows that \( \inf_{\lambda \geq 0} F(\lambda) = \min_{[0, \lambda_0]} f(\lambda) \). Numerical resolution of \( f'(\lambda) = 0 \) shows \( f'(\lambda_0) < 0 \), hence the result. \( \square \)

References


