Minimizing regret: the general case. ¤

Aldo Rustichini
CentER
Tilburg University
email:aldo@kub.nl

First version January 1997;
this version April 1998

Abstract

In repeated games with differential information on one side, the labelling "general case" refers to games in which the action of the informed player is not known to the uninformed, who can only observe a signal which is the random outcome of his and his opponent's action. Here we consider the problem of minimizing regret (in the sense first formulated by Hannan [8]) when the information available is of this type. We give a simple condition describing the approachable set.

JEL Classification: D81, D82, D83.

Keywords: Minimize Regret, Differential Information, Approachability.

¤This paper is forthcoming in Games and Economic Behavior. The author wishes to thank Jean-François Mertens for illuminating conversations. I tried to make the most of them, but the mistakes are my responsability. I also thank Rakesh Vohra, for his competent and friendly help; and two referee for extremely helpful comments, and for their patience. They forced me to try and make this paper intelligible. Any remaining shortcoming or worse is my responsability. Finally I thank the NSF for financial support of this research.

1 Introduction

In a very early paper Hannan [8] showed the following well known result. In a repeated game a player can use a strategy that, against any strategy of the opponent gives a regret which becomes arbitrarily close to zero as the length of the game increases. Regret is de¯ned as the di®erence between the average payo® along a play and the maximum payo® that the player could get in the one shot game if he could choose his action, when in turn the opponent is choosing his action according to the frequency on the sequence of the play. In other words the player in the repeated game in the limit is just as well o® as someone who is informed in advance of the frequency of choices of the opponent. In Hannan's theorem the player is informed in each period of the action chosen by the opponent, that is he has full monitoring.

This result has been extended and improved in several later papers. In particular Blackwell gave a very simple proof of this result, as a corollary of his approachability result: see [3]. Other contributions are: [2], [12], [4]. In particular the basic intuition of regret has been extended to game theory (see Fudenberg and Levine, [7]). For a detailed an informative discussion of this results, and the related literature on calibrated forecasting, see Foster and Vohra [6].

In a recent paper Auer, Cesa-Bianchi, Freund and Schapire (see [1]) have extended Hannan's result to the case in which the player is informed only of his own payo®: so the choice of action of the opponent can only be inferred from the realized payo®. This result can be considered a surprising extension of the multi-armed bandit, to a problem where no assumption is made on the stochastic property of the arms.

In this paper we extend the mimimum regret result to the general case in which the action of the player and of the opponent generates a random signal which is communicated to the player, i.e. the case of partial monitoring. Some basic concepts and de¯nitions are needed to state the result precisely.

2 Description of the game

We consider a game which is played over in¯nitely many periods.

There are two players, denoted by I and II for the ¯rst and second player respectively. Each player has a ¯nite action set, $S = \{1, \ldots, s, \ldots, S\}$ for player I and $T = \{1, \ldots, t, \ldots, T\}$ for player II. To simplify the notation we write, in subscripts, st rather than (s, t).

The payo® matrix for player II is denoted by $G^0 \in R^{ST}$; an element of the matrix is denoted by $G^0(s, t)$. We are going to build upon the literature on zero-sum games with di®erential information on one side, with player II uninformed: in this literature, the second player is usually minimizing his payo®. To make the connection with that literature easier, we also assume that the player II is minimizing his payo®. The payo® of player I is irrelevant for our purposes, so we

do not introduce notation for it; to fix ideas think of the game as a zero-sum game.

The action chosen by player I is not reavealed to player II directly, but only through the mediation of a signaling structure, which we now describe. For any finite set X, the set of probability measures on X is denoted by $\Delta(X)$. Q is a function from $S \times T$ to $\Delta(B)$, where B is a finite set of signals; a signaling structure is the pair $(B;Q)$. For every pair of actions $(s;t)$ of the two players, a signal is chosen according to the probability $Q(s;t)$ and then reported to player II. The probability measure concentrated on b is denoted by $\delta_{\{b\}}$.

A first example of signaling structure is: $B^* = S \times T$, $Q^*(s;t) = \delta_{\{st\}}$. Here the action of the two players is announced to both; this signaling structure $(B^*;Q^*)$ is called full monitoring.

A second example: the set of signals is the set of possible values of the payoff to player II, that is

$$B = \bigcup_{(s;t)\in S\times T} G^0(s;t); \tag{2.1}$$

and for every $(s;t)$,

$$Q(s;t) = \delta_{\{G^0(s;t)\}}; \tag{2.2}$$

so the signaling structure informs player II exactly of his payoff in that period.

Player I has better information: he knows the actions of both players as well as the signal announced to the second player. In a zero-sum game, this is clearly the worst possible case for player II.

The game is played over infinitely many periods. At each period n player I chooses action s_n and player II chooses action t_n. Then player I is informed of the action of the other player. At the same time a signal is chosen according to the distribution $Q(s_n;t_n)$ and is announced to the both players.

A history at period n for player I is the history of the actions of the two players, and of the signals, up to time n, that is an element in $(S \times T \times B)^n$. A history at period n for player II is a history of his actions and the signals observed, namely an element in $(T \times B)^n$.

A strategy for player I is a sequence $\sigma = (\sigma_0;\sigma_1;\dots;\sigma_n;\dots)$ of functions, where $\sigma_n : (S \times T \times B)^n \to \Delta(S)$ for every $n \geq 1$. A strategy for player II is a sequence $\tau = (\tau_0;\tau_1;\dots;\tau_n;\dots)$ of functions, where $\tau_n : (T \times B)^n \to \Delta(T)$ for every $n \geq 1$. σ_0 and τ_0 are constant strategies.

For each n, a pair of strategies $(\sigma;\tau)$, together with the signaling structure $(B;Q)$ induces a probability distribution $P_{(\sigma;\tau)}$ on the product space $(S \times T \times B)^n$.

The average payoff for player II at time N is

$$g_N^0 = \frac{1}{N+1}\sum_{n=0}^{N} G^0(s_n;t_n); \tag{2.3}$$

and the frequency $f_N \in R^S$ of action s of player I at time N is

$$f_N^s = \frac{1}{N+1}]\{n : s_n = s\}; \tag{2.4}$$

where] denotes the cardinality of the set. Our main result will relate the difference between the average payoff until stage N and the minimum payoff that player II could get in the stage game. The stage game is the zero-sum game with player I and II as players, action sets S and T respectively, payoff matrix G^0, and player II is minimizing. We use the symbols $x; x'$ and y for the mixed strategies in this game of player I and II. To avoid heavy notation, we agree that a vector on the left of a matrix is a row vector, and a vector on the right is a column vector; so xG^0y denotes the sum $\sum_{s;t} x^s G^0(s;t) y^t$; also since $Q \in \Delta(B)^{S \times T}$, $xQ \in \Delta(B)^T$.

Note finally that any signaling structure such that:

$$\text{if } xQ = x'Q \text{ then } x = x' \tag{2.5}$$

reveals to player II the vector x, if he knows the vector xQ. Signaling structures with the property (2.5) may be strictly weaker than the full monitoring. For instance, if $S = T = 3$, $B = \{0; 1\}$, and the matrix $(Q(s;t)(0))_{(s;t) \in S \times T}$ is the identity matrix, then (2.5) is satisfied, but for any fixed $t \in T$ the element $xQ(\cdot; t) \in \Delta(B)$ is not enough to determine x. Our main result (see theorem 3.2) implies that from the point of view of minimizing regret, the signaling structure that satisfy the even weaker condition:

$$\text{if } xQ = x'Q \text{ then } \min_{y \in \Delta(T)} xG^0y = \min_{y \in \Delta(T)} x'G^0y \tag{2.6}$$

(hence in particular those that satisfy 2.5) are as good for player II as the full monitoring one.

Examples of signaling structures that satisfy (2.6) are those that give full monitoring, or those that satisfy (2.5)); a second example is provided by those that inform player II of his own payoff (defined in (2.1) and (2.2)). The latter ones are considered by Auer, Cesa-Bianchi, Freund and Schapire [1]. We say that the signaling structures that satisfy (2.6) give statistically perfect monitoring. We now have all the preliminary notions necessary to present the main result.

3 Minimizing Regret: the Main Result

In loose terms in the full monitoring case regret at N for a given history is defined as the difference between the average payoff at N and the best (that is, the minimum) payoff for player II had he known the frequency of actions of player I. In the case of partial monitoring this idea has to be refined, since the frequency of actions is not directly observed.

Our main result states that in the general case of imperfect monitoring the regret is, in the worst case, the difference between the average payoff until stage N and the minimum payoff that player II could get against any mixed strategy of player I given the same frequency of signals, in the stage game. To make this statement more precise we define the function $\mathbb{C}_Q$ on $\Delta(S)$ as

$$\mathbb{C}_Q(x) \;=\; \max_{\{x' \in \Delta(S):\, x'Q = xQ\}} \; \min_{y \in \Delta(T)} \; x' G^0 y. \tag{3.7}$$

Clearly for signaling structures $(B;Q)$ that give statistically perfect monitoring:

$$\mathbb{C}_Q(x) = \min_{y \in \Delta(T)} xG^0y = \mathbb{C}_{Q^*}(x).$$

As in the literature on minimizing regret, we are going to prove that there is strategy for player II which is effective against any strategy of player I. To clarify the meaning of the word effective in the present context we begin with the case in which the frequency of the actions of player I converges to a limit value, x^* say.

In the case of full monitoring the strategy gives to player II, in the limit, a payoff which is at most equal to the minimum he can achieve in the stage game, where the first player is playing the mixed strategy x^*. This is Hannan's theorem. In the general case player II cannot observe the frequency x^* but only, at best, the induced frequency of signals x^*Q. We say "at best" because he will be able to do so only if he plays with sufficient frequency different actions, since the signal he observes depends on his action and the other player's action. Our main theorem implies that player II can achieve in the limit the worst case payoff among the strategies that give the same frequency of signals. This is precisely the value $\mathbb{C}_Q(x^*)$. More generally when the frequency of actions of player I does not converge, the theorem says that the difference between the average payoff at n and the value $\mathbb{C}_Q(f_n)$ tends to zero as n tends to infinity.

We now state Hannan's theorem and our main result in an informal way: for a precise statement, see the proposition (5.6) below.

Theorem 3.1 (Hannan) *For the full monitoring structure there is a strategy for player II such that for any strategy of player I the difference between the average payoff at n, g_n^0 and $\mathbb{C}_{Q^*}(f_n)$ tends to zero as n tends to infinity.*

Auer, Cesa-Bianchi, Freund and Schapire [1] generalize this to signaling structures that inform player II of his own payoff. We generalize this result to general signaling structure, as follows:

Theorem 3.2 (Main Result) *For all signaling structures there is a strategy for player II such that for any strategy of player I the difference between the average payoff at n, g_n^0, and $\mathbb{C}_Q(f_n)$ tends to zero as n tends to infinity.*

We illustrate the result with some examples.

3.1 Examples

In these examples $S = T$. Let Id be the $S \times S$ identity matrix. For a first example, take $S = T = \{1;2\}$, and consider the simple prediction game: the player II has to guess the next move of player I; so $G^0 = -Id$. The signal set is $B = \{0;1\}$; so the

matrix Q is completely described by the numbers $Q(s;t)(0)$ giving the probability of the signal 0.

If such matrix is:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

then the signaling structure is equivalent to full monitoring (for each action he plays, player II knows which action player I has played). In this case Hannan's theorem implies that if the frequency of the two actions converges to $(x^1; x^2)$, then player 1 can get a payoff of $\min\{-x^1; -x^2\}$, which is our function φ_Q in this case. In other words, if we let the set $C \subset R \times \Delta(S)$ be defined as

$$C = \{(y; x^1; x^2) : y \leq \min\{-x^1; -x^2\}\} \tag{3.8}$$

then the set C is approachable by player II.

Now consider a case of partial monitoring: let the matrix $Q(s;t)(0)$ be:

$$\begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix}$$

This example is similar to the one discussed in Helmold, Littlestone and Long, [10]. In this case when the player II plays the second action he receives no useful information on the move of player I in that period, while by playing the first action he knows exactly what action the first player has chosen.

Our main result implies that the same set C as defined in (3.8) is approachable by player II; in particular he can get a payoff of at most $\min\{-x^1; -x^2\}$ when the frequency of the actions of player I converges.

For the next example let $S = T = 3$, $G^0 = -Id$, $B = \{0; 1\}$, and the matrix $Q(s;t)(0)$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

Now player II can determine (by choosing his actions appropriately) the frequency of choice of action 1 by player I, but he has no way of determining the relative frequency of action 2 compared to action 3. If we again consider the case in which the frequency of the actions of player I converge to a limit $(x^1; x^2; x^3)$, then:

i. if $x^1 \leq 1/3$, then player II gets a payoff of at most $-1/3$, by playing by playing an appropriate strategy, with the frequency on his actions equal to 1 in the limit;

ii. otherwise, he can get a payoff of at most $-\frac{1-x^1}{2}$, by playing an appropriate strategy, with the frequency on his actions equal to $(0; 1/2; 1/2)$ in the limit.

4 Preliminaries

The reduction of our basic problem to an approachability result derived for games of incomplete information follows standard lines (as in Blackwell's proof [3] of Hannan's theorem [8]). This proof is a classical result: a clear exposition may be found in Luce and Rai®a, [11], pages 479-483). However, several important modi¯cations will be necessary. In this section we review the basic concepts and results needed in the sequel.

4.1 The full monitoring case

The idea of Blackwell's proof of Hannan's theorem is to de¯ne a game with vector payo®s, $(G^j)_{j \in J}$, where $J = \{0;1;\ldots;S\}$ has cardinality $S+1$; (see [13] chapter II, section 4, for details on games with vector payo®s: to understand what follows the reader only needs to know that these are games where the payo® to each player is a vector rather than a scalar.)

The payo® matrix for the ¯rst coordinate is the payo® matrix G^0. The payo® matrix for the coordinate $j \in \{1;\ldots;S\}$ is the matrix $G^j \in \mathbb{R}^{ST}$, with $G^j(s;t) = 1$ if $s = j$, and $G^j(s;t) = 0$ otherwise. The role of this matrix G^j, for $j \in S$, is simply to count the number of times the action j has been chosen by the ¯rst player; the average payo® in period n at the $j \in S$ coordinate is the frequency of action j in the ¯rst n periods.

Blackwell then uses his approachability theorem. The average payo® of the vector payo® game at time N is the vector $g_N \in \mathbb{R}^{S+1}$ de¯ned by

$$g_N^j = \frac{1}{N+1} \sum_{n=0}^{N} G^j(s_n;t_n), \text{ for } j = 0;\ldots;S.$$

For any integer m and any vector $y \in \mathbb{R}^m$ we de¯ne $\|y\|_i = \sum_{k=1}^{m} (|y^k|^i)^{1/i}$ for $i = 1;2$, where $|y^k|$ is the absolute value of y^k. The distance between the two vectors x and y is denoted by $d_i(x;y) = \|x - y\|_i$. We are going to use the same notations d_i independently of the dimension m; $d_i(x;C)$ is the d_i-distance between $x \in \mathbb{R}^m$ and a set C. Recall that for an m-dimensional vector x

$$\|x\|_2 \leq \|x\|_1 \leq \|x\|_2 \sqrt{m}. \tag{4.9}$$

As usual a set C is said to be approachable by player II if this player has a strategy σ such that for all $\varepsilon > 0$ there is an integer N such that for all strategies τ of the opponent

$$P_{\tau,\sigma}(\sup_{n \geq N} d_2(g_n;C) \leq \varepsilon) \geq 1 - \varepsilon. \tag{4.10}$$

Now de¯ne the set in $\mathbb{R} \times \Delta(S)$:

$$f(z;x) : z \geq \min_{y \in \Delta(T)} xG^0y\}$$

This set is convex, it is also easy to show directly that the Blackwell's condition for this set to be approachable by player II is satisfied; which gives Hannan's result. Since we do not use this condition directly, but rather a reformulation which we are going to discuss immediately, we refer the reader to pages 479 to 483 of [11] and continue.

4.2 Approachable sets

Here is the reformulation we mentioned. First define for every $z \in R^{S+1}$ a zero-sum game where players and action set are as before, and the matrix payoff is

$$D(z) = \sum_{j \in J} z^j G^j \tag{4.11}$$

in which the second player is minimizing. Denote by $v(z)$ its value. Then it follows from Blackwell's theorem and the minimax theorem (for a proof see [13], Corollary 4.6, section 4, chapter II; but see also [11], page 480, for the main idea) that a convex set $C \subset R^{S+1}$ is approachable by player II if and only if

$$\sup_{c \in C} \langle c, z \rangle \geq v(z) \text{ for every } z \in R^{S+1}. \tag{4.12}$$

where $\langle c, z \rangle$ denotes the inner product.

For the class of convex sets of the form (4.14) below, this condition has a simpler form. Let $w \in R^{S+1}$ be such that:

$$\langle w, p \rangle \geq v(p), \text{ for all } p \in \Delta(J). \tag{4.13}$$

Then it is a known result that the set:

$$\{c \in R^{S+1} \mid c^j \geq w^j, \text{ for all } j \in J\} = w + R^{S+1}_+ \tag{4.14}$$

is approachable by player II.

The proof of this result is in [13], (Proposition 2.13, section 2, chapter V) but the idea of the proof is simple. For any integer m, the set R^m_+ is the set of elements in R^m with non-negative components, and for any two subsets A and B of R^m,

$$A + B = \{a + b : a \in A, b \in B\}. \tag{4.15}$$

The set $w + R^{S+1}_+$ is convex, so by the characterization presented in (4.12), it is approachable by player II if and only if:

$$\sup_{c \in w + R^{S+1}_+} \langle c, z \rangle \geq v(z) \text{ for every } z \in R^{S+1}. \tag{4.16}$$

But for vectors z which are either zero, or have at least one component negative the conclusion is trivial (in the second case because the supremum in (4.16) is $+\infty$); for non-negative non-zero vectors the conclusion follows dividing the two sides of

the inequality in (4.16) (which are linearly homogeneous) by $k\,z\,k_1$, and using (4.13).

The reformulation in (4.13) introduces a simplification, at some cost. The simplification is that we can restrict attention to vectors in the simplex $\Delta(J)$ rather than considering all the vectors $z \in R^{S+1}$, as in (4.12). The cost is that we can prove approachability in this way only for convex sets C with the property that $C + R^{S+1}_+ \subseteq C$. We can however deal with the case of general convex sets easily, by rewriting our convex set appropriately. In the proof of the main result this is done in the paragraph following the equation (5.22).

4.3 Non-revealing strategies with partial monitoring

In this subsection we discuss the meaning of non-revealing strategies. No technical result, or concept needed later will be presented. Rather, we try to provide an intuitive reason for the definition of non-revealing strategies given later in equation (5.31). For details, see [13], chapter V, in particular page 229.

Consider first the case of games with full monitoring. The value of the game with payoff matrix $D(p)$ defined in (4.11) is also the expected value of a game with $S + 1$ states of nature, and where

i. player I is informed about the state, and might play a different strategy x_j for every state, but

ii. he plays non-revealing strategies, so that the posterior p (of the second player) does not change.

Formally, a non-revealing strategy at p is a vector of strategies $(x_j)_{j \in J}$ such that $x_j = x_{j'}$ for every j for which $p_j > 0$ and $p_{j'} > 0$.

In the case of games with partial monitoring, the concept of non-revealing has to be modified: a non-revealing strategy of the informed player is not a strategy which is constant across states, but a strategy which gives the same vector in $\Delta(B)$ independently of the state.

Formally, a non-revealing strategy at p with signaling structure $(B; Q)$ is a vector of strategies $(x_j)_{j \in J}$ such that $x_j Q = x_{j'} Q$ for every j such that $p_j > 0$ and $p_{j'} > 0$. The corresponding value v_Q is now the value of the zero-sum game where the informed player is playing non revealing strategies. The set $c + R^{S+1}_+$ is a set of approachable vectors if:

$$\{c \in R^{S+1} \mid \langle c; p \rangle \geq v_Q(p), \text{ for all } p \in \Delta(J)\}. \tag{4.17}$$

4.4 Games with a continuum of types.

In the previous section we have introduced games with J states of nature, and player I is informed about j. In this section we introduce an extension of this

concept to games with a continuum of types, as presented in [13], (chapter V, section 3.h.)

We consider the following repeated game with incomplete information. K is a finite set of states of nature (in the proof of the main result the cardinality of K will be 2S + 1); for each element k 2 K there is a payoff matrix G^k. Endow $\Delta(K)$ with the Borel structure, consider the set of probabilities $\Delta(\Delta(K))$ over it, and fix an element μ in this set. This element is known to both players. Nature chooses first a probability p 2 $\Delta(K)$ according to μ and communicates it to player I; then chooses the state k 2 K according to p, but does not reveal it to either player. Then the two players play the infinitely repeated game, where actions, payoffs and information are as described previously; its value is a function on $\Delta(\Delta(K))$ denoted u. We do not make the dependence on (B; Q) explicit, since it will be fixed in the rest.

Now the informed player does not know the state, but simply the probability distribution according to which the state was chosen. So the value of the game is a function u defined on $\Delta(\Delta(K))$; a strategy of player I in the stage game is a function from $\Delta(K)$ to $\Delta(S)$; a non-revealing strategy is a strategy which does not provide information on the distribution q 2 $\Delta(K)$; and the inequality $\langle c; p\rangle \leq v_Q(p)$ in (4.17) is replaced by

$$\langle h; \mu\rangle \leq u(\mu) \text{ for every } \mu \in \Delta(\Delta(K)) \tag{4.18}$$

with the finite dimensional inner product $\langle c; p\rangle$ replaced by the inner product $\langle h; \mu\rangle$ of a function h and a measure μ. In particular the approachability result for this setup that we are going to use is in the second part of proposition 3.47, section 3, chapter V of [13]. This result is recalled in proposition 5.4 below for convenience.

5 Proof of the Main result

We have now all the elements to proceed with the proof of the main result. Here is an outline of the main steps of the argument.

Outline of the proof

The three main steps in the proof are presnted in section 5.1, 5.2, 5.3 as follows:

i. in section 5.1 we define the set C ½ R^{S+1} we want to prove is approachable by player II; C is defined in equation (5.22). We prove that C is a convex set; then we let K ´ f0; 1; : : : ; 2Sg, and transform C into a set $\hat{C}$ ½ R^{2S+1} defined in (5.24). This set is also convex, and invariant under translation by the negative orthant (see equation (5.25) below). Lemma 5.1 proves that C is approachable if and only if $\hat{C}$ is approachable;

ii. in section 5.2 we associate to the set $\hat{C}$ a special function h, and after some preliminary result we prove in corollary 5.3 the basic inequality

$$\int_{\Phi(K)} h(q)\mu(dq) \geq u(\mu); \text{ for every } \mu \in \Phi(\Phi(K)) \tag{5.19}$$

which corresponds to the equation (4.13) for the set defined in equation (4.14) in the full monitoring case;

iii. in the final section 5.3 we use a result in the approachability theory for games with a continuum of types reported as Proposition 5.4, which implies that $\hat{C}$ is approachable if the function h satisfies (5.19); we then conclude by proving that the set C is approachable.

We now present the three main steps in detail.

5.1 The set C

We first prove that the function φ_Q is concave. The function f defined by

$$f(x) = \min_{y \in \Phi(T)} xG^0y; \tag{5.20}$$

is concave because it is a minimum of linear functions; so (see the second part of theorem 5.7 in Rockafellar [14]), the function

$$(Qf)(y) = \max\{f(x) : xQ = y\}$$

is also concave. But then:

$$\begin{aligned} \varphi_Q(x) &= \max\{f(x') : x'Q = xQ\} \\ &= (Qf)(xQ) \end{aligned} \tag{5.21}$$

(the first equality follows from the definitions of φ_Q and of f, the second from the definition of Qf), and $(Qf)(\cdot Q)$ is a concave function by the first part of theorem 5.7 of Rockafellar [14].

The set which we claim is approachable by player II is $C \subset R \times \Phi(S)$, the hypograph of φ_Q:

$$C = \{(b; x) \in R \times \Phi(S) : b \leq \varphi_Q(x)\}; \tag{5.22}$$

C is a convex set because φ_Q is concave. We now transform the set C into a more convenient form: the purpose is to get a set that satisfies the two properties (5.25) and (5.26) below. For any $(b; x) \in R \times \Phi(S); (b; x) = (b; x^1; \ldots; x^S)$ we let:

$$\Psi(b; x) = (b; x^1; 1 - x^1; \ldots x^S; 1 - x^S): \tag{5.23}$$

and:

$$\Phi(C) \equiv \{\Phi(b,x) : (b,x) \in C\}.$$

Finally we let

$$\hat{C} \equiv \Phi(C) - \{\{0\} \times R_+^{2S}\}. \tag{5.24}$$

Since C is convex and Φ is linear, $\Phi(C)$ is convex and so is $\hat{C}$. Also it is easy to check that

$$\hat{C} - R_+^{2S+1} = \hat{C}. \tag{5.25}$$

We want to show that the distance between the average payoff in the vector payoff game, which is a vector of the form $(b,x) \in R \times \Delta(S)$, and the set C is eventually small. But we find it convenient to show that the distance between $\Phi(b,x)$ and the set $\hat{C}$ is eventually small. These two statements turn out to be equivalent in view of the following lemma.

Lemma 5.1 For any $(b,x) \in R \times \Delta(S)$

$$d_1(\Phi(b,x),\hat{C}) = d_1((b,x),C). \tag{5.26}$$

Proof. An element of $\hat{C}$ has the form $(c^0,\ldots,c^s - z_1^s, -c^s - z_2^s,\ldots)$, with $c \equiv (c^0,c^1,\ldots,c^S)$ an element of C, and $z \equiv (z_1^1,\ldots,z_2^S) \in R_+^{2S}$. So

$$\begin{aligned}
d_1(\Phi(b,x),\hat{C}) &= \min_{c\in C, z\in R_+^{2S}} [|b - c^0| \\
&+ \sum_{s\in S} (|x^s - (c^s - z_1^s)| + |-x^s - (-c^s - z_2^s)|)] \\
&= \min_{c\in C} (|b - c^0| \\
&+ \sum_{s\in S} \min_{(z_1^s,z_2^s)\in R_+^2} (|x^s - c^s + z_1^s| + |x^s - c^s - z_2^s|)) \\
&= \min_{c\in C} (|b - c^0| + \sum_{s\in S} |x^s - c^s|) \\
&\equiv d_1((b,x),C).
\end{aligned} \tag{5.27}$$

■

5.2 The functions h and u

Let $K \equiv \{0,1,\ldots,2S\}$. It will prove convenient to write an element q of the simplex $\Delta(K)$ as follows: $(q^0,q_1^1,q_2^1,\ldots,q_1^S,q_2^S)$. Now define for any such q:

$$h(q) \equiv \max_{y\in\Phi(C)} \langle y,q\rangle; \tag{5.28}$$

so h is the support function of the set $\Phi(C)$, restricted to $\Delta(K)$, hence a convex function; also since the vector q has non-negative components,

$$h(q) = \max_{y \in \hat{C}} \langle y, q \rangle;$$

and, since $q^0 \geq 0$:

$$h(q) = \max_{x \in \Delta(S)} [q^0 \cdot G_Q(x) + \sum_{s \in S} x^s (q_1^s - q_2^s)] \tag{5.29}$$

Now we define a game with a continuum of types (see section 4.4 for the definition): the state space is K, and for every $s \in S$, we define the payoff matrix

$$G^s(s', t) = 1 \text{ if } s = s'; \text{ and 0 otherwise;}$$

and the matrix of the $2S + 1$-dimensional vector payoff

$$G \equiv (G^0, G^1, -G^1, \ldots, G^S, -G^S). \tag{5.30}$$

For any $\mu \in \Delta(\Delta(K))$, $\text{supp}(\mu)$ is the support of μ. We also denote by X the set of all functions from $\Delta(K)$ to $\Delta(S)$. The set of non-revealing strategies at μ, denoted $NR(\mu)$, is defined to be the set:

$$\{\sigma \in X : \exists \pi \in \Delta(B)^T \text{ s. t. } \sigma(q)Q = \pi \text{ for all } q \in \text{supp}(\mu)\} \tag{5.31}$$

We denote by $\sigma(q)^s$ the probability assigned to the action s by $\sigma(q)$. For every $\mu \in \Delta(\Delta(K))$ consider the zero-sum game, with players I and II, choosing a strategy in $\sigma \in NR(\mu)$ and $y \in \Delta(T)$ respectively, and payoff function

$$\int_{\Delta(K)} \sigma(q)[q^0 G^0 + \sum_{s \in S} (q_1^s - q_2^s) G^s] y \, \mu(dq)$$

The value of this game is:

$$\sup_{\sigma \in NR(\mu)} \min_{y \in \Delta(T)} \int_{\Delta(K)} [q^0 \sigma(q) G^0 y + \sum_{s \in S} \sigma(q)^s (q_1^s - q_2^s)] \mu(dq); \tag{5.32}$$

well defined by Sion's theorem [15], and is denoted by $u(\mu)$ (since the signaling structure (B, Q) is fixed we do not make the dependence explicit in the notation). For convenience we denote the integrand in its definition as

$$F(x, y, q) \equiv q^0 x G^0 y + \sum_{s \in S} x^s (q_1^s - q_2^s)$$

and, for any $\pi \in \Delta(B)^T$, the function u^π on $\Delta(\Delta(K))$:

$$u^\pi(\mu) \equiv \sup_{\{\sigma \in X : \sigma(q)Q = \pi; \text{ for all } q \in \text{supp}(\mu)\}} \min_{y \in \Delta(T)} \int_{\Delta(K)} F(\sigma(q), y, q) \mu(dq). \tag{5.33}$$

Clearly

$$u(\mu) = \sup_{\pi \in \Delta(B)^T} u^\pi(\mu), \text{ for every } \mu. \tag{5.34}$$

We now prove :

Lemma 5.2 For h de¯ned as in (5.28), every $¹ \in ¢(¢(K))$ and every $¯ \in ¢(B)^T$,

$$\int_{¢(K)} h(q)\,¹(dq) \geq u^{¯}(¹):$$

Proof. We begin with an estimate of the ¯rst term:

$$\begin{aligned}
&\int_{¢(K)} h(q)\,¹(dq)\\
&= \int_{¢(K)} \max_{x \in ¢(S)} [q^0 \max_{fx^0 \in ¢(S):x^0Q = xQg} \min_{y \in ¢(T)} x^0G^0y + \sum_{s \in S} x^s(q_1^s ¡ q_2^s)]\,¹(dq)\\
&= \max_{¢(K) \to ¯ \in ¢(B)^T} \max_{fx \in ¢(S):xQ = ¯g} [q^0 \max_{fx^0 \in ¢(S):x^0Q = ¯g} \min_{y \in ¢(T)} x^0G^0y\\
&+ \sum_{s \in S} x^s(q_1^s ¡ q_2^s)]\\
&= \max_{¢(K) \to ¯ \in ¢(B)^T} [q^0 \max_{fx^0 \in ¢(S):x^0Q = ¯g} \min_{y \in ¢(T)} x^0G^0y\\
&+ \max_{fx \in ¢(S):xQ = ¯g} \sum_{s \in S} x^s(q_1^s ¡ q_2^s)]\,¹(dq)\\
&\geq \max_{¯ \in ¢(B)^T} \int_{¢(K)} [q^0 \max_{fx^0 \in ¢(S):x^0Q = ¯g} \min_{y \in ¢(T)} x^0G^0y\\
&+ \max_{fx \in ¢(S):xQ = ¯g} \sum_{s \in S} x^s(q_1^s ¡ q_2^s)]\,¹(dq)\\
&= \max_{¯ \in ¢(B)^T} [\int_{¢(K)} (q^0 \max_{fx^0 \in ¢(S):x^0Q = ¯g} \min_{y \in ¢(T)} x^0G^0y)\,¹(dq)\\
&+ \int_{¢(K)} \max_{fx \in ¢(S):xQ = ¯g} \sum_{s \in S} x^s(q_1^s ¡ q_2^s)\,¹(dq)]: \qquad (5.35)
\end{aligned}$$

The ¯rst equality follows from (5.29) and the de¯nition (3.7) of the function $©_Q$, the second equality follows because

$$\max_{x \in ¢(S)} [q^0 \max_{fx^0 \in ¢(S):x^0Q = xQg} \min_{y \in ¢(T)} x^0G^0y + \sum_{s \in S} x^s(q_1^s ¡ q_2^s)]$$

$$= \max_{¯ \in ¢(B)^T} \max_{fx \in ¢(S):xQ = ¯g} [q^0 \max_{fx^0 \in ¢(S):x^0Q = ¯g} \min_{y \in ¢(T)} x^0G^0y + \sum_{s \in S} x^s(q_1^s ¡ q_2^s)]$$

for every $q \in ¢(K)$; the third equality follows because the term

$$q^0 \max_{fx^0 \in ¢(S):x^0Q = ¯g} \min_{y \in ¢(T)} x^0G^0y$$

only depends on $¯$ and not on x; the ¯rst inequality is clear (the $¯$ is now chosen independently of q); the fourth equality is simply the linearity of the integral.

In particular we conclude:

$$\int_{¢(K)} h(q)\,¹(dq) \geq \int_{¢(K)} (q^0 \max_{fx^0 \in ¢(S):x^0Q = ¯g} \min_{y \in ¢(T)} x^0G^0y)\,¹(dq)$$

$$+ \int_{\Delta(K)} \max_{\{x \in \Delta(S): xQ = \bar{\mu}\}} \sum_{s \in S} x^s (q_1^s - q_2^s) \lambda(dq), \text{ for every } \bar{\mu} \in \Delta(B)^T. \tag{5.36}$$

Next, we estimate the term $u^{-}(\lambda)$, for an arbitrary $\bar{\mu} \in \Delta(B)^T$:

$$\begin{aligned}
u^{-}(\lambda) &= \sup_{\{\sigma: \sigma(q)Q = \bar{\mu}, \text{ for all } q \in \text{supp}(\lambda)\}} \int_{\Delta(K)} \inf_{y \in \Delta(T)} F(\sigma(q); y; q) \lambda(dq) \\
&= \sup_{\{\sigma: \sigma(q)Q = \bar{\mu}\}} \int_{\Delta(K)} \inf_{y \in \Delta(T)} F(\sigma(q); y; q) \lambda(dq) \\
&= \sup_{\{\sigma: \sigma(q)Q = \bar{\mu}\}} \Big[\int_{\Delta(K)} \inf_{y \in \Delta(T)} q^0 \sigma(q) G^0 y \lambda(dq) \\
&\quad + \int_{\Delta(K)} \sum_{s \in S} \sigma(q)^s (q_1^s - q_2^s) \lambda(dq) \Big] \\
&\quad \sup_{\{\sigma: \sigma(q)Q = \bar{\mu}\}} \Big[\int_{\Delta(K)} \inf_{y \in \Delta(T)} q^0 \sigma(q) G^0 y \lambda(dq) \Big] \\
&\quad + \sup_{\{\sigma: \sigma(q)Q = \bar{\mu}\}} \int_{\Delta(K)} \sum_{s \in S} \sigma(q)^s (q_1^s - q_2^s) \lambda(dq) \\
&= \inf_{y \in \Delta(T)} \sup_{\{\sigma: \sigma(q)Q = \bar{\mu}\}} \int_{\Delta(K)} q^0 \sigma(q) G^0 y \lambda(dq) \\
&\quad + \sup_{\{\sigma: \sigma(q)Q = \bar{\mu}\}} \int_{\Delta(K)} \sum_{s \in S} \sigma(q)^s (q_1^s - q_2^s) \lambda(dq) \\
&= \inf_{y \in \Delta(T)} \int_{\Delta(K)} \max_{\{x \in \Delta(S): xQ = \bar{\mu}\}} q^0 x G^0 y \lambda(dq) \\
&\quad + \int_{\Delta(K)} \max_{\{x \in \Delta(S): xQ = \bar{\mu}\}} \sum_{s \in S} x^s (q_1^s - q_2^s) \lambda(dq).
\end{aligned} \tag{5.37}$$

The first equality is the definition of u^{-}; the second equality follows from the definition of support of a measure; the third equality by definition of the function F, the linearity of the integral, and the fact that the y variable does not appear in the second integral; the first inequality follows from a basic property of the sup; the fourth equality follows from Sion's theorem [15]; and the last equality follows from the measurable selection theorem (see for instance [5], theorem 3.1.1, page 111).

We now combine the two inequalities in (5.36) and (5.37) to prove the claim in the lemma; since the second term both sums is the same, we only need to prove that for a fixed $\bar{\mu} \in \Delta(B)^T$,

$$\int_{\Delta(K)} q^0 \max_{\{x \in \Delta(S): xQ = \bar{\mu}\}} \min_{y \in \Delta(T)} x G^0 y \lambda(dq) = \min_{y \in \Delta(T)} \int_{\Delta(K)} q^0 \max_{\{x \in \Delta(S): xQ = \bar{\mu}\}} x G^0 y \lambda(dq). \tag{5.38}$$

Consider the zero-sum game with payoff matrix G^0 where the player I, who maximizes, has to choose x such that $xQ = \bar{\mu}$. Let $\hat{x}$ and $\hat{y}$ be equilibrium strategies in this game, that is respectively maximin and minimax strategies, and $v(G^0; \bar{\mu})$ be

the value of this game. We claim that both sides of (5.38) are equal to

$$v(G^0;\bar{\ }) \int_{\Delta(K)} q^0 \mu(dq):$$

$\hat{x}$ and $\hat{y}$ are in fact, by de¯nition of equilibrium, also equilibrium strategies in the game q^0G^0 for any non-negative q^0. Hence the result. ■

Obviously:

Corollary 5.3 For h de¯ned as in (5.28) and every $\mu \in \Delta(\Delta(K))$,

$$\int_{\Delta(K)} h(q)\mu(dq) \geq u(\mu): \tag{5.39}$$

The equation (5.39) is the main conclusion of this section.

5.3 Final Step

We can now apply the second part of proposition 3.47, section 3, chapter V of [13]. We reproduce here the statement of that proposition, adapted to our case, for convenience of the reader. Let $L = \max\{\max_{(s;t)\in S\times T} |G(s;t)|, 1\}$. Recall that the average payo® g_n^0 and the frequency f_n are de¯ned in (2.3) and (2.4) respectively, and

$$g_n = \Phi((g_n^0; f_n)): \tag{5.40}$$

Then:

Proposition 5.4 (Mertens-Sorin-Zamir) For every h that satis¯es (5.39), and any sequence θ_n converging to zero there is a sequence ε_n with $2 \geq \varepsilon_n \geq \varepsilon_{n+1}$ converging to zero and a strategy σ_h of player II such that for every $n \geq 0$ and every strategy τ of the player I,

$$P_{\tau,\sigma_h}(N \geq n) \leq \exp(-n\theta_n); \tag{5.41}$$

where $N = \sup\{n \mid E_n > L\varepsilon_n\}$ (with $\sup(\emptyset) = 0$), where

$$E_n = \max_{q\in\Delta(K)} [\langle g_n; q\rangle - h(q)]: \tag{5.42}$$

The function h as de¯ned in (5.28) is a function that satis¯es (5.39) thanks to corollary (5.3). From this proposition and the preliminary results derived in the previous sections we now derive our main result. First we de¯ne a set to be approachable in the sense of (5.4):

De¯nition 5.5 The set C (de¯ned in (5.22)) is approachable if for every sequence θ_n converging to zero there is a sequence ε_n with $2 \geq \varepsilon_n \geq \varepsilon_{n+1}$ converging to zero

and a strategy σ of player II such that for every $n \geq 0$ and every strategy τ of the player I,

$$P_{\tau,\sigma}(N \geq n) \leq \exp(-n^2/n); \tag{5.43}$$

where $N = \sup\{n \mid \hat{E}_n > L\pm_n\}$ (with $\sup(\emptyset) = 0$), where

$$\hat{E}_n = d_2((g_n^0, f_n), C). \tag{5.44}$$

Finally we give the precise formulation of our main theorem (3.2):

Proposition 5.6 *The set C is approachable by player II.*

Proof. We have seen in section 5.2 that the function h defined in (5.28) satisfies the basic property (5.39). So by the proposition 5.4 we conclude that there exists a strategy σ_h such that holds for E_n, defined in (5.42).

We now claim that

$$\text{if } E_n \leq \rho \text{ then } \hat{E}_n \leq \rho(2S+1). \tag{5.45}$$

From this it follows that the basic estimate of the probability (5.43) given by (5.41) in proposition (5.4) holds also for $\hat{E}_n$, defined in (5.44). This will conclude the proof of the lemma. So we can now turn to the proof of the claim (5.45).

For any y, denote by $p_{\hat{C}}(y)$ the projection of y on $\hat{C}$, that is the unique element of $\hat{C}$ at minimum d_2-distance from y.

We prove two preliminary results, (5.46) and (5.49) below. First, we claim that for any element $y \in \mathbb{R}^{2S+1} \setminus \hat{C}$,

$$q^\alpha = \frac{y - p_{\hat{C}}(y)}{\|y - p_{\hat{C}}(y)\|_1} \in \Phi(K). \tag{5.46}$$

The only thing we have to prove is that $y - p_{\hat{C}}(y)$ is a vector with non-negative components. To see this, recall that the normal cone to a set A at a point x is:

$$N_A(x) = \{v : \langle x - c, v\rangle \geq 0 \text{ for all } c \in A\}. \tag{5.47}$$

First we have:

$$y - p_{\hat{C}}(y) \in N_{\hat{C}}(p_{\hat{C}}(y)); \tag{5.48}$$

(this is well known: the statement follows for instance from the two theorems 2.5.4, page 66 and 2.4.2, page 51, of Clarke [5]; or from theorem 3.1.1 in chapter III of Hiriart-Urruty and Lemarechal [9].) From the basic property (5.25) of the set $\hat{C}$ we have that $p_{\hat{C}}(y) - \mathbb{R}^{2S+1}_+ \subset \hat{C}$ and therefore $N_{\hat{C}}(p_{\hat{C}}(y)) \subset N_{p_{\hat{C}}(y) - \mathbb{R}^{2S+1}_+}(p_{\hat{C}}(y))$; but $N_{p_{\hat{C}}(y) - \mathbb{R}^{2S+1}_+}(p_{\hat{C}}(y)) = \mathbb{R}^{2S+1}_+$, hence the claim (5.46) is proved.

Second, the definition of $N_{\hat{C}}(p_{\hat{C}}(y))$ implies that

$$\langle p_{\hat{C}}(y) - c, q^\alpha\rangle \geq 0$$

for all $c \in \hat{C}$, hence

$$h(q^{\alpha}) = \langle p_{\hat{C}}(y); q^{\alpha} \rangle : \tag{5.49}$$

We now conclude the proof of the claim (5.45). Take now any $\varepsilon > 0$, and assume that

$$\langle y; q^{\alpha} \rangle - h(q^{\alpha}) \leq \varepsilon : \tag{5.50}$$

Using (5.49) one can compute:

$$\langle y; q^{\alpha} \rangle - h(q^{\alpha}) = \langle y - p_{\hat{C}}(y); \frac{y - p_{\hat{C}}(y)}{\| y - p_{\hat{C}}(y) \|_1} \rangle \leq \varepsilon ; \tag{5.51}$$

and we conclude after an elementary rearrangement and using (4.9)

$$\begin{aligned} \| y - p_{\hat{C}}(y) \|_2^2 &\leq \varepsilon \| y - p_{\hat{C}}(y) \|_1 \\ &\leq \varepsilon \| y - p_{\hat{C}}(y) \|_2 (2S+1)^{1/2}; \end{aligned} \tag{5.52}$$

hence:

$$\| y - p_{\hat{C}}(y) \|_2 \leq \varepsilon (2S+1)^{1/2} \tag{5.53}$$

Now if $E_n \leq \varepsilon^2$ then the inequality (5.50) holds if we set $y = g_n$; so if we set $y = g_n$ also in (5.46) then (5.52) holds, and therefore $\| g_n - p_{\hat{C}}(g_n) \|_2 \leq \varepsilon (2S+1)^{1/2}$; we conclude $d_2(g_n; \hat{C}) \leq \varepsilon (2S+1)^{1/2}$. But $d_1(g_n; \hat{C}) = d_1((g_n^0; f_n); C)$ by lemma 5.1 and (5.40). Therefore (recall (4.9)):

$$d_2((g_n^0; f_n); C) \leq \varepsilon (2S+1);$$

so our claim (5.45) is proved. ∎

6 Conclusion

The main result of this paper has been a simple characterization of the minimum regret in a game where the player has imperfect monitoring on the actions of the opponent. The worst case payoff is fully described by the function $\mathcal{C}_Q$.

This result suggests several possible measures of the value of information (that is, of the value of different signaling structures). For instance, one can introduce a simple partial ordering on signaling structures, for a given payoff matrix G^0, and say that $(B; Q)$ is better than $(B'; Q')$, for player II, given G^0, if

$$\mathcal{C}_Q \leq \mathcal{C}_{Q'} :$$

A partially open question is the issue of the speed of convergence. It is known from an example due to Zamir (see [13], page 290) that the rate of convergence in the general case is strictly worse than the rate $O(n^{-1/2})$ in the full monitoring case, namely $O(n^{-1/3})$. See also on this Auer, Cesa-Bianchi, Freund and Schapire [1]. It remains to be seen if there are interesting connections between the two matrices of payoff G^0 and of signals Q and the speed of convergence.

Finally, a simpler proof, independent of the approachablity result proved in [13] should be possible, perhaps along the lines of Auer, Cesa-Bianchi, Freund and Schapire [1]. This is left to future research.

References

[1] Auer, P., N. Cesa-Bianchi, Y. Freund and R. E. Schapire, (1995), \Gambling in a rigged casino: The adversarial multi-armed bandit problem", in Proceedings, 36th Annual Symposium on Foundations of Computer Science.

[2] Banos, A., (1968), \On Pseudo-Games" Annals of Mathematical Statistics, 39, 1932-1945.

[3] Blackwell, D., (1956), Controlled Random Walks. In Proceedings of the International Congress of Mathematicians, 1954 Vol.III, 336-338, Amsterdam, North-Holland, and Invited Address, Institute of Mathematical Statistics, Seattle, August 1956.

[4] Cesa-Bianchi, N., Y. Freund, H. Helmbold, D. Haussler, R. E. Schapire and M. K. Warmuth, (1993), \How to use expert advice", Proceedings of the 25th Annual ACM Symposium on the Theory of Computing, 382-391.

[5] Clarke, F. H., (1983), Optimization and Nonsmooth Analysis Canadian Mathematical Society, Wiley-Interscience Publication.

[6] Foster, D. and R. Vohra, (1996), Regret in the On-line Decision Problem, Discussion Paper.

[7] Fudenberg, D. and D. Levine, (1995), \Universal Consistency and cautious ¯ctitious play" Journal of Economic Dynamics and Control, 19, 1065-1089.

[8] Hannan, J., (1956), Approximation to Bayes risk in repeated play, In Dresher, Tucker and Wolfe (1957), Contributions to the Theory of Games, Vol. III, Annals of Mathematics Studies, vol. 39; Princeton, NJ, Princeton University Press.

[9] Hiriart-Urruty, J-B. and C. Lemarechal, (1993), Convex Analysis and Minimization Algorithms, I, Springer-Verlag, Berlin, New York.

[10] Helmold, D. P., Littlestone N. and Long P. M., (1992), \Apple tasting and nearly one-sided learning", 33rd Annual Symposium on Foundations of Computer Science.

[11] Luce, R. D. and H. Rai®a, (1957), Games and Decisions, published by Dover, 1985.

[12] Megiddo, N., (1980), \On repeated games with incomplete information played by non-Bayesian players" International Journal of Game Theory 9, 157-167.

[13] Mertens, J-F., S. Sorin and S. Zamir, (1994), Repeated Games. CORE (Center for Operations Research and Econometrics) Discussion Paper, number 9420-9421-9422.

[14] Rockafellar, R. T., (1970), Convex Analysis, Princeton University Press, Princeton, NJ.

[15] Sion, M., (1958), \On General Minimax Theorems", Paci¯c Journal of Mathematics, 8, 171-176.

Send Galley Proofs to

Aldo Rustichini
CentER for Economic Research
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands
fax: 00 31 13 466 3066
tel: 00 31 13 466 2770
email: aldo@kub.nl