A Wide Range No-Regret Theorem

by

Ehud Lehrer[1]
School of Mathematical Sciences
Sackler Faculty of Exact Sciences
Tel Aviv University, Ramat Aviv
Tel Aviv 69978, Israel
e-mail: lehrer@math.tau.ac.il

First version: February 2000
December 30, 2002

Abstract. In a sequential decision problem at any stage a decision maker, based on the history, takes a decision and receives a payoff which depends also on the realized state of nature. A strategy, f, is said to be as good as an alternative strategy g at a sequence of states, if in the long run f does, on average, at least as well as g does. It is shown that for any distribution, μ, over the alternative strategies there is a strategy f which is, at any sequence of states, as good as μ-almost any alternative g.

Journal of Economic Literature classification: C72, D81, D83

[1]I am grateful to Ehud Kalai, Rann Smorodinsky and Eilon Solan for fruitful discussions on this subject. I am also grateful to two anonymous referees of *Games and Economic Behavior*. This research was partially supported by the Israel Science Foundation, Grant no. 178/99

1. Introduction

In a sequential decision problem at any stage a decision maker, based on the history, takes a decision and receives a payoff which depends on the decision as well as on the realized state of nature. A strategy, f, determines which (possibly mixed) action to take after any history. After using f for a while, the decision maker may examine the track record of f. For instance, the decision maker may compare the performance of f with the performance of any stationary strategy. Hannan Theorem (1957) states that there is a strategy which does as good as any stationary strategy, independently of the history of states.

Another way to examine the performance of a strategy is to check whether on some subsets of stages the average payoff is as high as it could have been had a fixed action been played instead. Fudenberg and Levine (1999) showed that for any finite division of the set of stages which is known in advance, there is a strategy that outperforms any strategy that dictates playing a fixed action in each one of the subsets.

Suppose that a and b are two actions. An $(a,b)-$replacing scheme is a scheme that dictates playing action b whenever f dictates playing a. Note that a replacing scheme depends, at any given stage, on the action dictated by f and not only on past information, and is therefore not referred to as a strategy but as a scheme. Hart and Mas-Colell (1996) showed that there is a strategy which is simultaneously as good as all $(a,b)-$replacing schemes. In other words, there exists a strategy f with the property that, independently of the sequence of realized states, a decision maker has no regret for not playing action b instead of action a whenever the latter is the one determined by f.

A decision maker may want to play a strategy that is as good as all the strategies or schemes mentioned above and even as others. In this paper we consider a wide range of alternatives. An alternative is characterized by a

replacing scheme and a function, called an activeness function, that indicates the active periods of the strategy. A strategy is as good as an alternative if, over the periods declared by the activeness function as active, the average payoff actually received is as high as the average payoff that could have been received had the replacing scheme been played. In this case the decision maker has no regret against the alternative.

The notion of the activeness function enables us to model a decision maker who wants to examine the performance of the strategy used in subsets of periods. For instance, in case the decision maker would like to have a good track record over a set of specially important days, like weekend days, the activeness function would indicate only the weekend days as active. The result of Fudenberg and Levine (1999) can be therefore rephrased as follows. Consider a finite partition of the set of periods and for each subset consider the corresponding activeness function (the one that indicates only those times in the subset as active). Such an activeness function and a fixed action constitute an alternative. Fudenberg and Levine (1999) showed that there exists a strategy which is, in the long run, as good as any alternative of this kind independently of the history of states.

The model involved replacing schemes and activeness functions unifies the existing regret free results under a general theorem which is the main result of this paper. The latter states that for any distribution over alternatives there is a strategy which is as good as any member of a set of alternatives having probability 1. That is, there is a strategy which is regret free against a wide range of alternatives. This wide range no-regret theorem implies, in particular, that against any countable set of alternatives there is a strategy which is regret free, regardless of the sequence of states.

Hannan theorem can be also expressed in terms of finite automata. Stationary strategies are the strategies that can be generated by automata with one state. Thus, Hannan theorem states that there is a strategy which is, in the long run, as good as any strategy generated by an automaton with one

state. The wide range no-regret theorem implies, in particular, that there is a regret free strategy against **any** strategy in the countable set of the alternatives that can be generated by deterministic finite automata. This result may be extended verbatim to all the stochastic automata with rational numbers transition probabilities and mixed actions. Since the set of all stochastic automata (not only those restricted to rational numbers) is not countable, we can only have an "almost surely" statement.

The proof of existence uses the approachability theorem of Lehrer (1997) which extends Blackwell's theorem (1956) to infinite dimensional spaces.

The paper is organized as follows. The model is introduced in Section 2. Examples are given in Section 3. Existing no-regret theorems are presented in Section 4. The wide range no-regret theorem is introduced in Section 5. A short review of the extended approachability theorem, adapted to the current context, is given in Section 6. Section 7 contains the proof of existence. The paper is concluded with some final remarks in Section 8.

2. The Model

Let Ω be a finite state space. At any period nature chooses a state from Ω and the decision maker takes an action from a finite set A. At period, say, n, after nature chose a state, say, ω_n, and the decision maker took the action, say, a_n, the decision maker receives the payoff $u(\omega_n, a_n)$.

A history of length n ($n = 1, 2, ...$), is a list of n states and n actions, $h_n = (\omega_1, ..., \omega_n, a_1, ..., a_n)$. Let $\mathcal{H}$ be the set of all finite histories and of the null history, h_0. A strategy, f, is a function from the set of all histories to the set of distributions over A, denoted $\Delta(A)$. That is, $f : \mathcal{H} \rightarrow \Delta(A)$. Note that for any infinite sequence of states, $\omega = (\omega_1, \omega_2, ...)$, the strategy f and ω induce a probability distribution over the set of infinite sequences of actions, $A^{\mathbb{N}}$: At the first stage f determines a distribution according to which the first action, say, a_1, is to be chosen. Then at the second stage, based on the realized state, ω_1, and on a_1, f determines a distribution according to which the second action, a_2, is to be chosen, etc. Sometimes, when no ambiguity

arises, we use (ω, f) to denote the distribution induced by (ω, f).

Definition. A *replacing scheme* is a function, g from $\mathcal{H} \times A$ to A.

In words, a replacing scheme replaces an action (in A) with another action. This replacement may depend on the history. Note that we use the term scheme rather than strategy. This is so because a replacing scheme may depend not only on the history but also on an additional action. In case the strategy f is compared to the replacing scheme, the additional action is the one determined by f.

Definition. An *activeness function* is a function I from $\mathcal{H} \times A$ to $\{0, 1\}$.

For an activeness function I, we denote by $\overline{I}(h_{n-1}, a_n)$, where $h_{n-1} \in \mathcal{H}$ is a history of length $n-1$ and $a_n \in A$, the number of times I was active up to period n along h_{n-1}. That is, $\overline{I}(h_{n-1}, a_n) = \sum_{t=1}^{n} I(h_{t-1}, a_t)$, where $h_{n-1} = (\omega_1, ..., \omega_{n-1}, a_1, ..., a_{n-1})$.

We say that a strategy f is *as good as* the pair (g, I) at the sequence $(\omega = (\omega_1, \omega_2, ...)$, if $(\omega, f)-$almost surely $\overline{I}(h_{n-1}, a_n) \to_n \infty$ implies

$$\liminf_n \frac{\sum_{t=1}^{n} I(h_{t-1}, a_t)\big[u(\omega_t, a_t) - u(\omega_t, g(h_{t-1}, a_t))\big]}{\overline{I}(h_{n-1}, a_n)} \geq 0, \tag{1}$$

where $h_{t-1} = (\omega_1, ..., \omega_{t-1}, a_1, ..., a_{t-1})$, $t = 1, ..., n$. The meaning of (1) is the following. Suppose that according to strategy f the action taken at period t is a_t. The payoff then is $u(\omega_t, a_t)$. The decision maker compares it with the payoff resulted from the replacing scheme g, $u(\omega_t, g(h_{t-1}, a_t))$. The difference between the two payoffs is therefore $u(\omega_t, a_t) - u(\omega_t, g(h_{t-1}, a_t))$. Whether or not the decision maker wants to compare f with g at period t, is determined by the activeness function I. The comparison with g takes place only in the periods where I attains the value 1. The average, up to time n, of the difference in payoffs over the periods where the comparison takes place is therefore,

$$\frac{\sum_{t=1}^{n} I(h_{t-1}, a_t)\big[u(\omega_t, a_t) - u(\omega_t, g(h_{t-1}, a_t))\big]}{\overline{I}(h_{n-1}, a_n)}.$$

For a fixed sequence of states, if on almost any sequence of actions where the activeness function is active infinitely many times, the limit inferior of these averages is at least zero, we say that f is as good as (g, I). This means that the performance of f is on average (over the active times) not worse than that of g.

In the sequel we refer to a pair (g, I) also as an alternative to f.

3. Examples

It is convenient to give a short hand to some frequently used replacing schemes and activeness functions. For any action a, a^* denotes the stationary strategy that dictates playing a all the time, regardless of history. Let a, b be a pair of actions. There are many schemes that replace a with b. The symbol $g_{a,b}$ stands for the scheme that replaces a with b and keeps all other actions unchanged. That is, $g_{a,b}(h_{t-1}, a) = b$ and $g_{a,b}(h_{t-1}, c) = c$ for any $c \in A$, $c \neq a$ and history h_{t-1}. By I_a we denote the activeness function which is 1 only if the action in the argument is a. That is, $I_a(h_{t-1}, b) = 1$ only if $b = a$.

Let $\mathbf{1}$ be the activeness function which is always 1. A strategy f is as good as $(a^*, \mathbf{1})$, at the sequence $\omega_1, \omega_2, ...$, if in the long run the average payoff obtained by employing f is at least the payoff that could be achieved by playing constantly the action a. In other words, the decision maker does not regret playing f instead of playing constantly the action a. The average payoff obtained by the strategy a^* at time, say, t, equals the payoff obtained by playing the action a against the empirical distribution of states at time t. Thus, strategy f is as good as $(a^*, \mathbf{1})$ at a sequence of states, if over almost any sequence of actions, there is a time from which on f outperforms the action a against the empirical distribution of states. Strategy f is, therefore, as good as $(a^*, \mathbf{1})$ for every $a \in A$, if f is not worse than playing always a best response to the long-run empirical distribution of states.

The use of the activeness functions is better exemplified in what follows. Suppose that a decision maker would like to examine the performance of the strategy he employs separately on the even number periods and on the

odd number periods. Let E and O be two activeness functions, such that $E(t) = 1$ only if t is an even number and $O = 1 - E$. In particular, E and O are complementary: E is active when O is not. A strategy f is as good as (a^*, E), if in the long run, the average payoff over the even number periods is greater than what the action a could ensure against the empirical distribution of the states over these periods. Strategy f may be as good as many pairs of replacing schemes and activeness functions. For instance, strategy f is as good as both (a^*, E) and (a^*, O), if, at the long run, first, the average payoff over the even number periods is greater than what the action a could ensure against the empirical distribution of the states over these periods and, second, the average payoff over the odd number periods is greater than what the action a could ensure against the empirical distribution of the states over the odd number periods.

Strategy f is simultaneously as good as $(g_{a,b}, I_a)$ and as (a^*, O), if first, over the times where a was played, a is not worse than b; and, second, over the odd number periods f is as good as the action a.

Remark 1. Note that the fact that f is as good as $(g_{a,b}, I_a)$ implies that f is as good as $(g_{a,b}, \mathbf{1})$. The reason is that the denominator of (1) that corresponds to $(g_{a,b}, I_a)$ (the number of times that a was played up to period n) is not greater than n, the denominator of (1) that corresponds to $(g_{a,b}, \mathbf{1})$.

A decision maker may also want to examine his strategy in a Markovian fashion. For instance he may want to ensure that the strategy he employs is as good as another strategy over stages that follow a specific action or a specific state. In order to illustrate it, fix $\omega \in \Omega$ and $a \in A$ and define the activeness function $M_{\omega,a}$ as 1 only if the last state and action were ω and a, respectively. Strategy f is as good as, say, $(b^*, M_{\omega,a})$, if f is as good as the action b over the periods that follow the occurrence of ω, a.

4. Existing Related Results

4.1 Hannan Theorem

Hannan Theorem (1957) refers only to stationary strategies. It states that there is a strategy that is as good as any stationary strategy at any sequence of realizations. That is, without knowing anything about the selection of states by nature, there is a way to ensure, based on past information only, that the average payoffs is as high as the average payoff obtained by playing a constant strategy.

Formally,

Theorem (Hannan, 1957). *There is a strategy f which is as good as $(a^*, \mathbf{1})$ for any $a \in A$ at any sequence $\omega_1, \omega_2, ...$.*

4.2 Hart and Mas-Colell No-Regret Theorem

Hart and Mas-Colell deals with replacing schemes of the kind $g_{a,b}$. Similarly to Hannan theorem it ensures the existence of a strategy which is regret free when only the strategies of the kind $g_{a,b}$ are considered.

Theorem (Hart and Mas-Colell, 1996). *There is a strategy f which is as good as $(g_{a,b}, \mathbf{1})$ at any sequence $\omega_1, \omega_2, ...$ and for any $a, b \in A$.*

4.3 Fudenberg and Levine No-Regret Theorem

Suppose that $B_1, ..., B_k$ is a partition of the set of periods. That is, the sets $B_1, ..., B_k$ are disjoint and cover the set of all integers. Denote by $a^*|B_j$ the scheme that replaces any action on B_j with the action a and leaves any action out of B_j unchanged.

Fudenberg and Levine (1999) showed that Hannan theorem can be generalized to subsets of periods. In other words, there is a strategy which is better than any fixed strategy over each one of the subsets.

Theorem (Fudenberg and Levine, 1999). *For any partition of the set of stages $B_1, ..., B_k$ there is a strategy f which is as good as $(a^*|B_j, \mathbf{1})$ at any sequence $\omega_1, \omega_2, ...$ for any $a \in A$ and $1 \leq j \leq k$.*

5. A Wide Range No-Regret Theorem

This section presents the main result of this paper. Each one of the results that have been quoted above restricts itself to a specific finite set of replacing schemes. Here we consider the set of all replacing schemes.

Denote by $\mathcal{R}$ the set of all pairs (g, I), where g is a replacing scheme and I is an activeness function. The set $\mathcal{R}$ is measurable[2]. Let μ be a probability measure over $\mathcal{R}$. The measure μ can be interpreted as the distribution over all possible replacements that the decision maker compares the strategy he employs to. It can also be interpreted as the subjective weight of importance the decision maker attributes to any possible alternative.

The measure μ can assign a positive probability to only countably many pairs. As an example one may think of all the schemes, denoted as $c^t_{a,b}$, that are constantly a up to stage t and are constantly b thereafter. The set consisting of all pairs $(c^t_{a,b}, \mathbf{1})$ is a countable set. Another example of a countable set is that of all the alternatives (recall, these consist of a scheme and an activeness function) that can be generated by finite automata.

The result of this paper is that for any μ there is a strategy f which is regret free against μ-almost any alternative considered. That is, f is as good as μ-almost any pair (g, I). In other words, f is immunized against regret when a wide range of alternatives is considered. Formally,

Theorem *Given a distribution μ over $\mathcal{R}$, there is a strategy f, such that for any sequence $\omega_1, \omega_2, \ldots$, f is as good as μ-almost any pair $(g, I) \in \mathcal{R}$ at $\omega_1, \omega_2, \ldots$.*

Remark 2. In case μ has a countable support, the theorem states that there is a strategy which is regret free against any alternative in the support. In particular it means that there is a strategy which is as good as any alternative (g, I) that can be produced by a finite automaton.

[2]Both, g and I, are functions defined on $\mathcal{H} \times A$ and are therefore similar to pure strategies. Thus, $\mathcal{R}$ is similar to the set of a repeated game pure strategies, and is therefore, measurable.

Remark 3. Note that due to Remark 1, the theorem implies a stronger result than that of Hart and Mas-Colell (see Section 4.2): There is a strategy f which is as good as $(g_{a,b}, I_a)$ at any sequence $\omega_1, \omega_2, ...$ and for any $a, b \in A$.

Remark 4. The theorem implies, in particular, that if $I_1, I_2, ...$ is a sequence of activeness functions, not necessarily complementary, then there exists a strategy f which is as good as any (a^*, I_j), $j = 1, 2, ...$, $a \in A$. In other words, for any j, the average payoff obtained by playing f is, over the periods where I_j is equal to 1, at least what any constant action can ensure at the same set of periods.

This result extends that of Fudenberg and Levine (1999) (see Section 4.3) in two ways. First, the fact that f is as good as (a^*, I_j) implies that f is as good as $(a^*|B_j, \mathbf{1})$, where B_j is the set of periods where I_j is equal to 1. Second, this result allows for more than finitely many activeness functions, and furthermore, these functions should not necessarily induce a partition (i.e., $\sum I_j$ can be greater or smaller than 1.)

Remark 5. In this discussion we restrict ourselves to pure replacing schemes. A behavioral replacing scheme can be defined like a replacing strategy with the only change that the range of the behavioral replacing scheme is the set of mixed actions (instead of the set of pure actions). In a similar way one can define a behavioral activeness function. A pair of a behavioral replacing scheme and a behavioral activeness function is called a behavioral alternative. Since the game played is with perfect recall, any behavioral alternative is equivalent to a distribution over pairs of replacing schemes and activeness functions. Thus, a distribution over behavioral alternatives is equivalent to a distribution over pure pairs. Therefore, the restriction to pure alternatives does not create any loss in the generality of the theorem. In other words, the theorem can be rephrased as follows. For any distribution over behavioral alternatives there is a strategy which is regret free against almost all pure alternatives at any sequence of states. The "almost all" here means with probability 1 according to both, the distribution over behavioral

alternatives and the distribution over pure alternatives induced by the behavioral alternatives. This statement implies that for any distribution over behavioral alternatives there is a strategy which is regret free against almost all behavioral alternatives at any sequence of states.

6. Approachability in Large Spaces

It is well known that the various no-regret theorems mentioned in Section 4 can be proven by an approachability theorem. Our result is not exceptional. We prove it by an approachability theorem in infinite dimensional space. For this purpose we resort to Lehrer (1997) which extends the approachability theorem of Blackwell (1956). This section is devoted to a short review of this result.

We think of the decision maker and of nature as two competing players. Instead of having only one objective as usual (usually to maximize the expected payoff), here the decision maker has a multiple objective. He wants to ensure that he has no regret against many alternatives.

Considering one alternative, the decision maker regrets not playing this alternative, if the average payoff he actually received is smaller than the average payoff that could have been received had he played the alternative. In other words, he has no regret if the actual average payoff is at least the average payoff ensured by the alternative. Thus, for a given alternative his goal is to ensure that the average realized payoff minus the "could have been" payoff is at least zero. This goal extends to many alternatives. That is, in case many alternatives are considered, the goal is to ensure that for each alternative the average of the actual payoff minus the "could have been" payoff is at least zero.

The formal presentation of the game is as follows. Let μ be a probability measure over $\mathcal{R}$. At any stage the decision maker chooses (possibly in a random fashion) an action in A. Nature chooses a state in Ω. After $t-1$ stages the history of actions is $a_1, a_2, ..., a_{t-1}$ and the history of states is $\omega_1, \omega_2, ..., \omega_{t-1}$. At this time the decision maker decides to play an action, a_t,

and nature decides to choose a state, ω_t. The payoff is a function over $\mathcal{R}$, denoted $X^{(a_t,\omega_t)}$. That is, for any $(g,I) \in \mathcal{R}$ there is a value attached. In this context, we refer to (g,I) as a coordinate. The value of $X^{(a_t,\omega_t)}$ at the coordinate (g,I), $X^{(a_t,\omega_t)}(g,I)$, is

$$I(h_{t-1},a_t)\big[u(\omega_t,a_t)-u(\omega_t,g(h_{t-1},a_t))\big], \tag{2}$$

where $h_{t-1}=(\omega_1,...,\omega_{t-1},a_1,...,a_{t-1})$. The goal of the decision maker is to ensure that the (long run) average payoff in almost every coordinate is at least zero.

Note that the payoff in (2) depends on the history. Stated differently, in the games played between the decision maker and nature, the payoffs (recall, these are functions over $\mathcal{R}$) depend on the history h_{t-1}.[3] As stated above, whether the coordinate (g,I) is active or not depends on whether $I(h_{t-1},a_t)$ is 0 or 1.

Up to period n, the coordinate (g,I) was active $\overline{I}(h_{n-1},a_n)$ times. Thus, the average payoff up to period n, over the active periods, at the (g,I) coordinate is precisely the quotient of (1). That is, considering the alternative (g,I), the decision maker has no regret against (g,I), if the average payoff at the (g,I) coordinate is, asymptotically, at least 0.

Denote by $\overline{X}_n(g,I)$ the average payoff in the coordinate (g,I) up to time n along the history $h_n=(\omega_1,...,\omega_n,a_1,...,a_n)$. That is,

$$\overline{X}_n(g,I)=\frac{\sum_{t=1}^{n}X^{(a_t,\omega_t)}(g,I))}{\overline{I}(h_{n-1},a_n)}.$$

Using this notation, strategy f is as good as (g,I) at the sequence $\omega=(\omega_1,\omega_2,...)$, if $\liminf \overline{X}_n(g,I)\geq 0$ at (ω,f)-almost any sequence $a_1,a_2,...$, provided that $\overline{I}\to\infty$. In order to link it with approachability, a few more notations are needed.

Let C be the set of all non negative functions over $\mathcal{R}$. That is, $C=\{\psi:\mathcal{R}\to I\!R;\ \ \psi(g,I)\geq 0 \text{ for } \mu-\text{almost all } (g,I)\}$. Fix two infinite sequences,

[3] It is actually a stochastic game with deterministic transition probabilities (i.e., 0 or 1).

$a_1, a_2, ...$ of actions and $\omega_1, \omega_2, ...$ of states. Denote by $\mathbb{1}_{\{\overline{I}\to\infty\}}$ the indicator of the set of alternatives, (g, I), satisfying $\overline{I} \to \infty$. That is, $\mathbb{1}_{\{\overline{I}\to\infty\}}(g, I)$ takes the value 1 if $\overline{I} \to \infty$ and the value 0 otherwise. Note that the indicator $\mathbb{1}_{\{\overline{I}\to\infty\}}$ depends on the two sequences $a_1, a_2, ...$ and $\omega_1, \omega_2, ...$.

Let $\overline{X}_n^-(g, I) = \min\{\overline{X}_n(g, I), 0\}$. We say that $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n \to C$, if the difference between C and $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n$ goes to zero μ-almost surely. That is, if $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n^-(g, I) \to 0$ μ-almost surely. Strategy f is as good as μ-almost every (g, I), at the sequence $\omega = (\omega_1, \omega_2, ...)$, if $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n \to C$ (ω, f)-almost surely (i.e., for (ω, f)-almost all sequences of actions). Alternatively, if $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n$ approaches the set C, then the decision maker has no regret playing f, against almost all alternative strategies.

Lehrer (1997) provides a condition that guarantees that the decision maker has a strategy that ensures that for any strategy of nature $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n$ approaches the set C. The adaptation to the current context of Theorem 3 in Lehrer (1997) is as follows.

Proposition 1 *1. For any $\omega \in \Omega$, if at any period n and after any history of length $n-1$, $h_{n-1} = (\omega_1, ..., \omega_{n-1}, a_1, ..., a_{n-1})$, there is a mixed action p of the decision maker (a distribution over A) such that,*

$$(3) \qquad \int_{\mathcal{R}} \overline{X}_n^-(g, I) \sum_{a\in A} \left(\frac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)} p(a) \big[u(\omega, a) - u(\omega, g(h_{n-1}, a)) \big] \right) d\mu \leq 0,$$

then there is a strategy f, such that (ω, f)-almost surely $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n \to C$.

In the statement of the proposition there are two probabilities involved. The first is the probability over $\mathcal{R}$, μ. The second is the probability over sequences of actions induced by (ω, f). The theorem claims that, if (3) holds, then there exists a strategy f such that with (ω, f)-probability 1, $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n$ converges to C with μ-probability 1. Note that the theorem ensures that $\overline{X}_n(g, I) \to 0$ (ω, f)-almost surely only subject to the condition that I is active infinitely many times. There is no way to ensure the convergence of the quotient of (1) to 0 in case the denominator does not converge to infinity.

The goal of the decision maker is to bring the function $\overline{X}_n$ as close as possible to the target set (of functions), C. At period n the gap between $\overline{X}_n$ and the set C, referred to later as the error, is the difference between $\overline{X}_n$ and the closest point in C to $\overline{X}_n$. This is precisely $\overline{X}_n^-$.

In case the decision maker chose the action a and nature chose the state ω, the payoff at time n corresponding to coordinate (g, I), is $I(h_{n-1}, a)\big[u(\omega, a) - u(\omega, g(h_{n-1}, a))\big]$. The contribution of this single stage payoff to the average $\overline{X}_n(g, I)$ is greater as the coordinate is less active. That is, the contribution of a single stage payoff to the average $\overline{X}_n(g, I)$ decreases with $\overline{I}(h_{n-1}, a)$. Furthermore, the contribution of a single stage payoff to the average at the coordinate (g, I) depends whether $I(h_{n-1}, a)$ is 1 or not. Taking these facts into account, we consider a weighted next-time payoff, $\frac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}\big[u(\omega, a) - u(\omega, g(h_{n-1}, a))\big]$, which depends on $I(h_{n-1}, a)$ and on $\overline{I}(h_{n-1}, a)$. Note that the weighted next-time payoff decreases with $\overline{I}(h_{n-1}, a)$. Furthermore, it is equal to 0 when $I(h_{n-1}, a) = 0$. The expected next-time weighted payoff, when the mixed action p is played, is therefore, $\sum_{a\in A}\left(\frac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}p(a)\big[u(\omega, a) - u(\omega, g(h_{n-1}, a))\big]\right)$.

Inequality (3) says that the integral of the product of the error and the expected weighted next-time payoff is less than or equal to zero. In terms of the geometry of the space of functions over $\mathcal{R}$ this means that the expected next-time weighted payoff and the average payoff, $\overline{X}_n$, are located in two different sides of a hyper space that separates between $\overline{X}_n$ and C.[4] In this sense the next-time payoff "corrects" the past error. Proposition 1 states that if at any period n, there is a mixed action that ensures that the expected payoff "corrects" the error accumulated up to period $n-1$, then the error can be diminished to zero. That is, there is a strategy f, such that $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n$ approaches C (ω, f)-almost surely.

[4]In fact this is not just a hyper space that separates $\overline{X}_n$ from C. This is the one that passes through the closest point in C to $\overline{X}_n$, $\overline{X}_n - \overline{X}_n^-$, and is perpendicular to $\overline{X}_n^-$, the difference between $\overline{X}_n$ and C .

Remark 6. In fact Lehrer (1997) not only provides the condition, it also provides, like in Blackwell (1956), the strategy f that ensures that $\mathbb{1}_{\{\overline{I}\to\infty\}}\overline{X}_n(g,I)$ approaches C (ω,f)-almost surely. The strategy defined there dictates at any stage to choose an action according to the mixed action p that satisfies (3).

7. The Proof of the Wide-Range No-Regret Theorem

What remains to be done is to show that the condition of Proposition 1 is satisfied. Let h_{n-1} be a history of length $n-1$. Denote for any $a,b\in A$ $\mathcal{R}_{a,b}=\{(g,I);\ g(h_{n-1},a)=b\}$. For any a $\{\mathcal{R}_{a,b}\}_{b\in A}$ is a partition of $\mathcal{R}$ into finitely many disjoint sets. Thus, for any a

$$\int_{\mathcal{R}}\overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}\big[u(\omega,a)-u(\omega,g(h_{n-1},a))\big]d\mu=$$

$$\sum_{b\in A}\int_{\mathcal{R}_{a,b}}\overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}[u(\omega,a)-u(\omega,b)]d\mu.$$

Thus, for any mixed action, say, q, (compare with (3))

$$\int_{\mathcal{R}}\overline{X}_n^-(g,I)\sum_{a\in A}\Big(\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}q(a)\big[u(\omega,a)-u(\omega,g(h_{n-1},a))\big]\Big)d\mu=$$

$$\sum_{a\in A}\Big(\int_{\mathcal{R}}\overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}q(a)\big[u(\omega,a)-u(\omega,g(h_{n-1},a))\big]d\mu\Big)=$$

$$\sum_{a\in A}\int_{\mathcal{R}}\overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}q(a)u(\omega,a)d\mu-$$

$$\sum_{a\in A}\sum_{b\in A}\int_{\mathcal{R}_{a,b}}\overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}q(a)u(\omega,b)d\mu=$$

$$\sum_{a\in A}\int_{\mathcal{R}}\overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}q(a)u(\omega,a)d\mu-$$

$$\sum_{b\in A}\sum_{a\in A}\int_{\mathcal{R}_{b,a}}\overline{X}_n^-(g,I)\tfrac{I(h_{n-1},b)}{\overline{I}(h_{n-1},b)}q(b)u(\omega,a)d\mu=$$

$$\sum_{a\in A} u(\omega,a)\Big(\int_{\mathcal{R}} \overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}q(a)d\mu - \sum_{b\in A}\int_{\mathcal{R}_{b,a}} \overline{X}_n^-(g,I)\tfrac{I(h_{n-1},b)}{\overline{I}(h_{n-1},b)}q(b)d\mu\Big) =$$

$$\sum_{a\in A} u(\omega,a)\Big(q(a)\int_{\mathcal{R}} \overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}d\mu -$$

$$\sum_{b\in A} q(b)\int_{\mathcal{R}_{b,a}} \overline{X}_n^-(g,I)\tfrac{I(h_{n-1},b)}{\overline{I}(h_{n-1},b)}d\mu\Big).$$

The condition of Proposition 1 is that there exists p such that

$$\sum_{a\in A} u(\omega,a)\Big(p(a)\int_{\mathcal{R}} \overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}d\mu -$$

$$\sum_{b\in A} p(b)\int_{\mathcal{R}_{b,a}} \overline{X}_n^-(g,I)\tfrac{I(h_{n-1},b)}{\overline{I}(h_{n-1},b)}d\mu\Big) \le 0.$$

This inequality will be guaranteed if we show that there exists p such that for any a the term in the parenthesis is equal to zero. That is,

$$(4)\qquad p(a)\int_{\mathcal{R}} \overline{X}_n^-(g,I)\tfrac{I(h_{n-1},a)}{\overline{I}(h_{n-1},a)}d\mu - \sum_{b\in A} p(b)\int_{\mathcal{R}_{b,a}} \overline{X}_n^-(g,I)\tfrac{I(h_{n-1},b)}{\overline{I}(h_{n-1},b)}d\mu = 0.$$

Consider the matrix $\{W_{b,a}\}_{b,a\in A}$, where $W_{b,a} = -\int_{\mathcal{R}_{b,a}} \overline{X}_n^-(g,I)\frac{I(h_{n-1},b)}{\overline{I}(h_{n-1},b)}d\mu$ for $b\neq a$ and $W_{b,b} = \int_{\mathcal{R}-\mathcal{R}_{b,b}} \overline{X}_n^-(g,I)\frac{I(h_{n-1},b)}{\overline{I}(h_{n-1},b)}d\mu$. Note that $\sum_a W_{b,a}$. Thought of as a zero-sum game, where the row player (the player who chooses the action b) is the minimizer, the matrix W has a value zero. To see this, note that the column player (here the maximizer) can ensure getting at least 0 by playing the uniform distribution over A. Furthermore, against any distribution over the a's (of the column player), say, s, the row player can choose a row, say, b, whose s probability is maximal. Thus, if $s(b)$ is maximal, the expected payoff that corresponds to b and s is, $\sum_{a\neq b} s(a)W_{b,a} + s(b)W_{b,b} = \sum_{a\neq b}\Big((s(b)-s(a))\int_{\mathcal{R}_{b,a}} \overline{X}_n^-(g,I)\frac{I(h_{n-1},b)}{\overline{I}(h_{n-1},b)}d\mu\Big) \le 0$ because $\overline{X}_n^-(g,I)\frac{I(h_{n-1},b)}{\overline{I}(h_{n-1},b)} \le 0$. Since the row player can ensure at most zero against any mixed action s of the column player, we obtain, by the minmax theorem, that the row player

has an action p that ensures the value. That is, there exists p such that for any column the payoff is at most zero.

In order to conclude the argument we need to show that with p the payoff is precisely zero for any column. However, this is obvious since there exists an optimal action of the column player which assigns a positive probability to any column: the uniformly distributed mixed action. That is, the mixed action p satisfies (4), and therefore satisfies the condition of Proposition 1.

Proposition 1 ensures the existence of a strategy f that guarantees that the average payoff, $\mathbb{1}_{\{\bar{I}\to\infty\}}\overline{X}_n \to C$ (ω, f)-almost surely, regardless of nature's choices. We conclude that there is a strategy f which is regret free against μ-almost all alternatives (g, I) at any sequence of realizations $\omega_1, \omega_2, ...$.

8. Final Remarks

8.1 An Alternative Proof[5]

The main result of this paper can be proven in an alternative way. Sandroni, Smorodinsky and Vohra (2000) proved that there exists a forecasting rule that calibrates with a large set of checking rules. For any replacing scheme one can find a few corresponding checking rules that have the following property. A best response to the forecasting rule that calibrates with the corresponding checking rules is a regret free strategy against the replacing scheme under consideration. When a large set of replacing schemes is considered one can find, using Sandroni, Smorodinsky and Vohra (2000), a forecasting rule that calibrates with almost all the corresponding checking rules. Like in Hart and Mac-Colell (1996), a best response to this forecasting rule is regret free against almost all the replacing schemes under consideration.

[5]The contents of this section has been suggested independently by Rann Smorodinsky and by an anonymous referee.

¿From a computational point of view the proof provided here is significantly simpler than the alternative proof. The reason is that here the approachability theorem requires solving, at any stage, a fixed size, $|A| \times |A|$, zero-sum game (recall the matrix $\{W_{b,a}\}_{b,a \in A}$). In contrast, in the proof of Sandroni, Smorodinsky and Vohra (2000) the size of the zero-sum game that the forecaster needs to solve at any stage (in order to apply the approachability theorem) increases with time. Furthermore, in the alternative proof the construction of the regret free strategy must pass through a forecasting rule which is conceptually rather complicated. The forecasting rule provided by Sandroni, Smorodinsky and Vohra (2000) is, at any stage, a random choice from some distributions over the set Ω.

8.2 No-regret and Correlated Equilibrium

In a game with a few players, each one may consider the other players as nature. In this case, other players' chosen actions are, in terms of the previous sections, the realized state of nature. Denote by A_i player i's set of actions. As stated in Hart and Mas-Colell (1996), if each player i plays a strategy which is as good as $(g_{a,b}, I_a)$ for any $a, b \in A_i$, then the empirical frequency of the joint actions played converges to the set of correlated equilibria. It may well happen that the empirical frequency over the even times is meaningless. That is, it may happen that the empirical frequency of the joint actions over the entire set of times converges to the set of correlated equilibria, while the statistics over the even number times does not converge to anything meaningful.

The wide range no-regret theorem ensures that each player has a strategy f which is regret free against each one of the alternatives $(g_{a,b}, I_a)$, $a, b \in A_i$, and simultaneously against (recall the activeness function E mentioned in Section 2) $(g_{a,b}, EI_a)$, $a, b \in A_i$. That is, over the even number periods where f dictates playing a, f performs at least as well as the action b. The same argument as in Hart and Mas-Colell (1996) implies that if each player employs such a strategy, the empirical frequencies over the entire history

and that restricted to the even number periods, both converge to the set of correlated equilibria. There is no guarantee that the two frequencies are in any sense similar to each other.

One can employ many other activeness functions that are active on pre-specified subsets of periods. In case all players do the same and if a subset of periods is infinite, the empirical frequency of the joint actions (over this subset) converges to the set of correlated equilibria.

8.3 Guessing Games

At the end of Section 7 above I defined a matrix W and showed that the row player has an action that guarantees that the expected payoff at any column is exactly zero. The matrix $-W$ belongs to a family of games called Guessing games (see Lehrer (1998)). A guessing game is a zero sum game in which both players have the same set of actions, A. The payoffs on the diagonal of the game matrix are non-negative while the other payoffs are non-positive. Finally, the sum of all payoffs in any row is zero. The interpretation of such a game is as follows. The column player chooses an action a and the row player guesses which action was it. In case the row player guesses correctly, he receives a reward (a payoff in the diagonal) otherwise he is penalized by receiving a non-positive payoff (off the diagonal). In guessing games, for any guess b, of the row player, the sum of all penalties is equal to the reward. Lehrer (1998) uses the fact that any positive combination of guessing games is a guessing game to show that a player who participates in many sequential guessing games at the same time, can ensure that, on average, the payoff in each one of them is asymptotically non-negative.

One can find references (explicit or insinuated) to guessing games in Hart and Schmeidler (1989) Nau and McCradle (1990) and in Foster and Vohra (1999).

8.4 No-Regret Theorem with Imperfect Monitoring

Rustichini (1999) generalized Hannan theorem to the case of imperfect

monitoring. At any stage the decision maker receives a signal which stochastically depends on the state and the action chosen. Since the information about the previous choices of nature is partial, there are many possible distributions over Ω that are informationally consistent with the empirical frequency of the signals received (see also Lehrer (1989)). Rustichini showed that there is a strategy whose long-run average payoff is as high as the payoff obtainable by playing a best response to the worst mixed choice of nature which is informationally consistent with the empirical frequency of signals. In other words, there exists a strategy which is regret free against the worst (in the sense of minmax) mixed choice of nature which is indistinguishable (using the signals received through the information structure) from the actual frequency of states.

I conjecture that Rustichini's result can be generalized to a wide range of alternatives. In other words, it is conjectured that in the case of imperfect monitoring there exists a strategy which is regret free (in a sense adapted to the information structure) against many alternatives.

9. References

Blackwell, D. (1956) "A Vector Valued Analog of the Mini-Max Theorem," *Pacific Journal of Mathematics*, **6**, 1-8.

Foster, D., and R. Vohra (1999), "Regret in the On-Line Decision Problem," *Games and Economic Behavior*, **29**, 7-35.

Fudenberg, D., and D. Levine (1999), "Conditional Universal Consistency," *Games and Economic Behavior*, **29**, 104-130.

Hannan, J. (1957), "Approximation to Bayes Risk in Repeated Plays," in *Contribution to the Theory of Games*, **3**, 97-139. Princeton, NJ: Princeton University Press.

Hart, S., and A. Mas-Colell (1996), "A Simple Adaptive Procedure Leading to Correlated Equilibrium," Center for Rationality and Interactive Decision Theory, DP #126, The Hebrew University, Jerusalem, Israel.

Hart, S., and D. Schmeidler (1989), "Existence of Correlated Equilibria," *Mathematics of Operations Research*, **14**, 18-25.

Lehrer, E. (1989), "Lower Equilibrium Payoffs in Two Players Repeated Games with Non-Observable Actions," *International Journal of Game Theory* , **18**, 57-89.

Lehrer, E. (1997), "Approachability in Infinite Dimensional Spaces and an Application: A Universal Algorithm for Generating Extended Normal Numbers," mimeo.

Lehrer, E. (1998), "Guessing Games," mimeo.

Nau, R.F. and K.F. McCradle (1990), "Coherent Behavior in Noncooperative Game," *Journal of Economic Theory* , **50** , 424-444.

Rustichini, A. (1999), "Minimizing Regret: The General Case," *Games and Economic Behavior*, **29**, 224-243.

Sandroni, A., R. Smorodinsky and R. Vohra (2000),"Calibration with Many Checking Rules," manuscript.

List of Footnotes

1. Both, g and I, are functions defined on $\mathcal{H} \times A$ and are therefore similar to pure strategies. Thus, $\mathcal{R}$ is similar to the set of a repeated game pure strategies, and is therefore, measurable.

2. It is actually a stochastic game with deterministic transition probabilities (i.e., 0 or 1).

3. In fact this is not just a hyper space that separates $\overline{X}_n$ from C. This is the one that passes through the closest point in C to $\overline{X}_n$, $\overline{X}_n - \overline{X}_n^-$, and is perpendicular to $\overline{X}_n^-$, the difference between $\overline{X}_n$ and C.

4. The contents of this section has been suggested independently by Rann Smorodinsky and by an anonymous referee.

Ehud Lehrer
Mailing address:
School of Mathematical Sciences
Sackler Faculty of Exact Sciences
Tel Aviv University
Ramat Aviv
Tel Aviv 69978, Israel

e-mail: lehrer@math.tau.ac.il
tel: 972-3-6408822
fax: 972-3-6409357