Learning Effectiveness and Memory Size

Abraham Neyman*

April 29, 2008

Abstract

We study learning effectiveness as a function of memory size.

We quantify the maximal level that a bounded memory machine (or agent) can match (or reproduce) a long string of inputs as a function of the input length k and the memory size n. The input string is an element of I^k and the output string is an element of J^k and the loss of the agent when matching an input coordinate $i \in I$ with an output coordinate $j \in J$ is $g(i,j)$.

This level is expressed by a function $v(p,\theta)$ of two variables: a probability p on I and a nonnegative $\theta \geq 0$. The function $v(p,\theta)$ is defined as a function of the triple $G = \langle I, J, g\rangle$. It equals the minimum of $E_Q g(i,j)$, where the minimization is over all distributions Q on action pairs with marginal p on I, denoted Q_I, and the mutual information $I_Q(i;j) = H(Q_I) + H(Q_J) - H(Q) \leq \theta$, where H is the entropy function.

If $i_1, \ldots, i_k$ are iid I-valued random variables with distribution p, then for $T \subset J^k$ we have $E \min_{(j_1,\ldots,j_k)\in T} \frac{1}{k}\sum_{t=1}^k g(i_t, j_t) \geq v(p, \frac{\log|T|}{k})$. Moreover, for every finite set T of functions τ from the finite strings I^* of I-elements to J we have $E \min_{\tau\in T} \frac{1}{k}\sum_{t=1}^k g(i_t, \tau(i_1,\ldots,i_{t-1})) \geq v(p, \frac{\log|T|}{k})$. It follows that if σ is the mixed strategy of player 1 in the infinite repetition of the stage game G that plays a k-periodic sequence $i_1, i_2, \ldots$, where $i_1, \ldots, i_k$ are iid random variables with distribution p, then for every strategy τ of player 2 that is defined by an automaton

*Institute of Mathematics and Center for the Study of Rationality, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel (e-mail: aneyman@math.huji.ac.il). This research was supported in part by Israel Science Foundation grants 263/03 and 1123/06.

with n states we have $E_{\sigma,\tau}\frac{1}{k}\sum_{t=1}^{k} g(i_{L+t}, j_{L+t}) \geq v(p, \frac{\log n}{k})$ for every L. This inequality holds also for any strategy τ of player 2 that is defined by the most general machine with n states of memory (for example, the machine can have, in addition to its n memory states, access to a clock and to a randomization device).

On the other hand, for every k and n there is a function $f : I^k \to T \subset J^k$ with $|T| \leq n$ such that $\frac{1}{k}\sum_{t=1}^{k} g(i_t, j_t) \leq v(p, \frac{\log n}{k}) + o(1)$ (as $k \to \infty$), where p is the empirical distribution of $i_1, \ldots, i_k$ and $(j_1, \ldots, j_k) = f(i_1, \ldots, i_k)$. In addition, we prove that there is a deterministic automaton with n states and input alphabet I that when faced with a periodic sequence $i_1, i_2, \ldots$ of I-inputs with period $r \leq k$ outputs a sequence $j_1, j_2, \ldots$ with $\lim_{L\to\infty}\frac{1}{L}\sum_{t=1}^{L} g(i_t, j_t) \leq v(p(\sigma), \frac{\log n}{k}) + o(1)$ as $k \to \infty$, where $p(\sigma)$ is the empirical distribution of the sequence $i_1, i_2, \ldots$.

It follows that the value of the two-person zero-sum repeated game $G[k, n]$ (with stage game $G = \langle I, J, g\rangle$), where player 1's possible strategies are those defined by oblivious automata of size k and player 2's possible strategies are those defined by automata of size n, converges, as k goes to infinity and $\frac{\log n}{k}$ goes to $\theta \geq 0$, to the limit $v(\theta)$, where $v(\theta)$ is the max of $v(p, \theta)$ where the max is over all mixed stage actions p of player 1. Moreover, player 2 has a pure strategy in $G[k, n]$ that is approximately optimal. The result remains intact ?
when player 2's possible strategies are those defined by automata with time-dependent mixed actions and mixed transitions.

The minimal duration of learning is derived from the analysis of the L-stage repeated games $G^L[k, n]$. We prove that if k and $\frac{L}{k \log k}$ go to infinity and $\frac{\log n}{k}$ goes to θ then the values of $G^L[k, n]$ converge to $v(\theta)$.

1 Introduction

Some hedge funds owe their success to recognizing economic patterns that their competitors failed to notice. Their discovery of such trading strategies contradicts the common economic/finance theory wisdom that agents are fully rational with unlimited computational and memory capacities. After all, if one hedge fund is capable of finding such profitable patterns, why should others not be able to trade on the same patterns?

Such contradictions dissipate if one takes into account the limited rationality of agents.

The present paper derives quantitative results about the level at which a limited rational agent can utilize repeated patterns of streaming data.

We consider a decision maker that interacts with a stochastic process (i_t) with values i_t in the finite set I. At stage $t \geq 1$ the decision maker outputs an action j_t in the finite set J, and thereafter observes the realization i_t, and the cost at stage t to the decision maker is $g(i_t, j_t)$ where $g : I \times J \to \mathbb{R}$. If the t-coordinate i_t of the process is a known deterministic function of the past $(i_s)_{1 \leq s < t}$, then a decision maker with unbounded memory and unlimited computational ability can compute i_t as a function of the past $(i_s)_{1 \leq s < t}$ and guarantee at stage t the minimal feasible cost $\min_j g(i_t, j)$. In particular, if for $t > k$ the i_t coordinate of the process is a function of $i_1, \ldots, i_k$, a "supersmart" agent can output the j_t that minimizes the cost $g(i_t, j_t)$. However, if the decision maker has a bounded memory, such a perfect optimization may prove impossible.

The results characterize a threshold (continuous) function v of two variables: a distribution p on I and the positive number $\frac{\log n}{k}$ such that 1) the agent has a simple strategy that uses n states of memory such that for every k-periodic sequence the average per-stage cost is $v(p, \frac{\log n}{k}) + o(1)$ (as $k \to \infty$) where p is the empirical distribution of (i_t); and 2) if (i_t) is a k-periodic sequence with $i_1, \ldots, i_k$ iid with distribution p, then for every strategy with n states of memory the expected average per-stage cost is at least $v(p, \frac{\log n}{k})$.

The strategy in 1) is the simplest strategy with n states of memory: a deterministic (stationary) automaton with n states. Moreover, the automaton's program is a natural simple program and finding it requires little sophistication: a natural random choice leads with high probability to the desired approximate optimal automaton. On the other hand, the inequality in 2) holds for the most general strategy with n states of memory: a probabilistic time-dependent automaton with n states. Therefore the results are robust

with respect to the choice of modeling strategies with n states of memory.

An important ingredient of the model is that when the agent outputs j_t he is in one of n states of memory and this state can be a function of $i_1, \ldots, i_{t-1}$. The state of memory then is the only record of the history $i_1, \ldots, i_{t-1}$. At stage t an additional input i_t is observed. Therefore the state of memory recording the history $i_1, \ldots, i_t$, $m_{t+1} = m(i_1, \ldots, i_t)$ is a function of $m_t = m(i_1, \ldots, i_{t-1})$ and i_t. The classical model of an automaton requires this function to be deterministic and stationary (independent of t). The most general model of an n-state memory will allow the function mapping m_t and i_t to m_{t+1} to depend on time and a randomization device.

The periodicity of the sequence is essential for the statement of the formal result but is not conceptually important. An alternative interpretation (or statement) of the result is that the agent examines a long stream of data $i_1, \ldots, i_k$, and outputs a string $j_1, \ldots, j_k$ so as to minimize $\frac{1}{k}\sum_{t=1}^{k} g(i_t, j_t)$. The agent with n states of memory is modeled here as an automaton with n states. The states of the automaton are partitioned into transition states and terminal states. As long as the automaton is in a transition state it continues to examine the sequence. Once reaching a terminal state it no longer has access to the sequence and starts to output $j_1, \ldots, j_k$ (as a function of the terminal state).

We turn now to the statement and the discussion of the results from a game-theoretic perspective. Let $G = \langle I, J, g \rangle$ be a two-person zero-sum game; I and J are the finite action sets of players 1 and 2 respectively, and $g : I \times J \to \mathbb{R}$ is the payoff function to player 1. The repeated game, where player 1's, respectively player 2's, possible strategies are those defined by automata of size k, respectively size n, and the payoff is the average per-stage payoff, is denoted $G(k, n)$. Ben-Porath (1986, 1993) proves that the value of $G(k, n)$ converges to the value of the stage game G, as k goes to infinity and $\frac{\log n}{k} + \frac{\log k}{n}$ goes to 0 (namely, the size of the larger automata is subexponential of the size of the smaller automata).

It follows that in order to have an asymptotic nonvanishing advantage in the repeated game with finite-state automata an exponentially larger automata size is needed. [12] (respectively, [18]) proves that if $\liminf_{k\to\infty} \frac{\log n_k}{k}$ is $> \min\{\log|I|, \log|J|\}$ (respectively, $\geq \min\{\log|I|, \log|J|\}$), then the value of $G(k, n_k)$ converges, as k goes to infinity, to the maxmin of the stage game, where player 1 maximizes over his pure stage actions $i \in I$ and player 2 minimizes over his pure stage actions $j \in J$. The asymptotic behavior of the values of $G(k, n_k)$ as $n_k/k \to \theta > 0$ is unknown for $0 < \theta < \min\{\log|I|, \log|J|\}$.

The approximate optimal strategies used in [1] are mixtures of oblivious strategies (namely, strategies that are nonreactive to the actions of the other player) defined by an automaton of the specified size. Therefore, if $G[k,n]$ denotes the repeated game, where player 1's possible strategies are those defined by oblivious automata of size k and player 2's possible strategies are those defined by automata of size n, then [1] proves that the value of $G[k,n_k]$ converges, as k goes to infinity and $\frac{\log n_k}{k}$ goes to 0, to the value of the stage game.

One possible interpretation of a stochastic process of actions generated by a mixture of oblivious strategies is a stochastic process of states of nature. The set I denotes the temporal states of nature and $g(i,j)$ is the cost of the decision maker (player 2 in the game-theoretic interpretation) as a function of his/her action j and the state of nature i. Therefore, the result of [1] demonstrates that a memory size that is a subexponential function of the length of the (minimal) cycle of the states of nature is insufficient to utilize the fact that the states of nature follow a cyclic play.

A repeated game model that leads to oblivious strategies of a player is the case where the player does not observe the actions of the other players.

The present paper proves that for sufficiently large k the value of $G[k,n]$ is approximated by a function v of $\frac{\log n}{k}$. The function v depends on the data of the stage game and its definition uses the entropy function. For a probability distribution p in $\Delta(I)$ (the set of probability distributions over I) and $\theta \geq 0$, we denote by $\mathcal{Q}(p,\theta)$ the set of all probability distributions $Q \in \Delta(I \times J)$ with marginal Q_I on I coinciding with p and $H(Q_I) + H(Q_J) - H(Q) \leq \theta$ (where H is the entropy function). Note that the set $\mathcal{Q}(p,\theta)$ is closed and convex, and $\alpha\mathcal{Q}(p,\theta_1)+(1-\alpha)\mathcal{Q}(p,\theta_2) \subset \mathcal{Q}(p,\alpha\theta_1+(1-\alpha)\theta_2)$ for $0 \leq \alpha \leq 1$. Define

$$\begin{aligned} v(p,\theta) &= \min_{Q\in\mathcal{Q}(p,\theta)} E_Q g(i,j) \\ v(\theta) &= \max_{p\in\Delta(I)} v(p,\theta) \end{aligned}$$

Let $p \in \Delta(I)$. We prove that if σ is the mixed strategy of player 1 in the repeated game $G[k,n]$ that plays a k-periodic sequence $i_1, i_2, \ldots$ where $i_1, \ldots, i_k$ are iid with $P(i_1 = i) = p(i)$, then for every strategy τ of player 2 in $G[k,n]$, we have

$$E_{\sigma,\tau}\frac{1}{k}\sum_{t=1}^{k} g(i_{L+t}, j_{L+t}) \geq v(p, \frac{\log n}{k}) \tag{1}$$

for every L. In particular, $E_{\sigma,\tau} \lim_{L\to\infty} \frac{1}{L} \sum_{t=1}^{L} g(i_t, j_t) \geq v(p, \frac{\log n}{k})$. Therefore the value of $G[k,n]$ is $\geq v(p, \frac{\log n}{k})$ for every $p \in \Delta(I)$, and thus $\geq v(\frac{\log n}{k})$.

On the other hand, we prove that in $G[k,n]$ player 2 has a pure strategy τ such that for every strategy σ of player 1 the long-run average payoff, $\lim_{L\to\infty} \frac{1}{L} \sum_{t=1}^{L} g(i_t, j_t)$, is close to $v(p(\sigma), \frac{\log n}{k})$ (explicitly, $\leq v(p(\sigma), \frac{\log n}{k}) + o(1)$ as $k \to \infty$), where $p(\sigma)$ is the empirical distribution of the actions of player 1 when using the strategy σ. Therefore, the value of $G[k,n]$ is $(\geq v(\frac{\log n}{k})$ and$) \leq v(\frac{\log n}{k}) + \varepsilon_k$ where $\varepsilon_k \to 0$ as k goes to infinity. It follows that the values of $G[k, n_k]$ converge to $v(\theta)$ as $\frac{\log n_k}{k} \to \theta \geq 0$ and k goes to infinity. Moreover, the limit is uniform over all stage games $\langle I, J, g\rangle$ with $\|g\| := \max_{i,j} |g(i,j)| \leq 1$.

A finite automaton of player 2 with n states is a machine with n states of memory. The memory state m_t is changing from stage t to stage $t+1$ as a deterministic function of the input i_t. Following a play in stages $1 \leq t < L$ the automaton's summary of the past play/data is captured by his present state m_L. Two characteristics of such an automaton are the deterministic and stationary transitions. In [4] a striking difference between a time-dependent probabilistic automaton and a time-independent probabilistic automaton[1] emerges in hypothesis testing. It is therefore of interest to ask whether inequality (1) holds also for any strategy τ that is defined by a time-dependent probabilistic automaton. It turns out that it holds also for any strategy τ defined by a time-dependent mixed actions and mixed transitions. It follows that the result about the limit of the values of $G[k,n]$ remains intact when player 1's possible strategies are those defined by oblivious automata of size k and player 2's possible strategies are those defined by n-state automata with time-dependent mixed actions and mixed transitions.

The duration of learning is analyzed by studying the value of the L-stage repeated game $G^L[k,n]$, where the possible strategies of player 1 are those defined by oblivious automata of size k and player 2's possible strategies are those defined by automata of size n, and the payoff is the average of the payoffs in the first L stages of the repeated game. We prove that the value of $G^{L_k}[k, n_k]$ is $\geq v(\frac{\log n_k}{k})$ and $\leq v(\frac{\log n_k}{k}) + \varepsilon(L_k, k, n_k)$ where $\varepsilon(L_k, k, n_k) \to 0$ as k and $\frac{L_k}{k \log k}$ go to infinity. It follows that the value of $G^{L_k}[k, n_k]$ converges to $v(\theta)$ as $\frac{n_k}{k} \to \theta$ and both k and $\frac{L_k}{k \log k}$ go to infinity.

[1] Interesting comments on this issue appear in [2, 3, 5, 8, 9, 10].

2 Preliminaries

A *pure strategy* of player 1, respectively player 2, in the repeated game G^* (with stage game $G = \langle I, J, g \rangle$) is a function $\sigma : (I \times J)^* \to I$, respectively $\tau : (I \times J)^* \to J$, where $(I \times J)^*$ is the set of all finite strings (including the empty string $\emptyset$) of elements of $I \times J$. A pair of pure strategies, σ of player 1 and τ of player 2, defines a play $i_1, j_1, i_2, j_2, \ldots$ of the repeated game as follows: $i_1 = i_1(\sigma, \tau) = \sigma(\emptyset)$, $j_1 = j_1(\sigma, \tau) = \tau(\emptyset)$, $i_t = i_t(\sigma, \tau) = \sigma(i_1, j_1, \ldots, i_{t-1}, j_{t-1})$, and $j_t = j_t(\sigma, \tau) = \tau(i_1, j_1, \ldots, i_{t-1}, j_{t-1})$.

The L-stage average payoff as a function of the pure strategy pair (s_1, s_2) is $g_L(s_1, s_2) = \frac{1}{L} \sum_{t=1}^{L} g(i_t, j_t)$, where $i_t = i_t(s_1, s_2)$ and $j_t = j_t(s_1, s_2)$. If (σ, τ) is a mixed strategy pair, then $g_L(\sigma, \tau) = E_{\mu,\sigma} g_L(s_1, s_2)$. Whenever the limit of $g_L(\sigma, \tau)$ as $L \to \infty$ exists, it is denoted by $g(\sigma, \tau)$, and termed the average per-stage payoff.

An *automaton* of player 2 consists of

- a set of *states* M
- an *action function* $\alpha : M \to J$
- a *transition function* $\beta : M \times I \to M$
- an *initial state* $m^* \in M$

The *size* of an automaton is the number $|M|$ of states.

An *automaton* $A = \langle M, m^*, \alpha, \beta \rangle$ for player 2 defines a strategy $\tau = \tau^A$ as follows. Define the sequence of states $(m_t)_{t \geq 1}$

- $m_1 = m^*$
- $m_{t+1} = \beta(m_t, i_t)$

Note that m_t is a function of $i_1, j_1, \ldots, i_{t-1}, j_{t-1}$. Define

$$\tau(i_1, j_1, \ldots, i_{t-1}, j_{t-1}) = \alpha(m_t)$$

Analogously, one defines an automaton for player 1.

An *oblivious automaton* is an automaton $A = \langle M, m^*, \alpha, \beta \rangle$, where the transition function β is independent of the action of the other player. Explicitly, an oblivious automaton of player 1 consists of a set of *states* M, an *action function* $\alpha : M \to I$, a *transition function* $\beta : M \to M$, and an *initial*

state $m^* \in M$. The *size* of an oblivious automaton is the number $|M|$ of states.

An oblivious automaton $A = \langle M, m^*, \alpha, \beta \rangle$ defines a strategy $\sigma = \sigma^A$ as follows. Define the sequence of states $(m_t)_{t \geq 1}$

- $m_1 = m^*$
- $m_{t+1} = \beta(m_t)$

Note that m_t is a function of $i_1, \ldots, i_{t-1}$ and thus, in particular, a function of $i_1, j_1, \ldots, i_{t-1}, j_{t-1}$. Define

$$\sigma(i_1, j_1, \ldots, i_{t-1}, j_{t-1}) = \alpha(m_t)$$

A sequence of actions $(i_1, i_2, \ldots)$ is defined by an oblivious automaton of size $|M|$ if and only if it enters at a stage $1 \leq t_0 \leq |M|$ a cycle of length $1 \leq t_1 \leq |M|+1-t_0$; namely, there are $1 \leq t_0 \leq |M|$ and $1 \leq t_1 \leq |M|+1-t_0$ such that for every $t \geq t_0$ we have ($m_t = m_{t+t_1}$ and thus) $i_t = i_{t+t_1}$. For example, set $t_0 = \min\{t \geq 1 \mid \exists t' > t \text{ with } m_{t'} = m_t\}$ and $t_1 = \min\{t \geq 1 \mid m_{t_0} = m_{t_0+t}\}$.

The set of all automata of size n of player 2 (as well as the set of all strategies of player 2 that are defined by automata of size n) is denoted $\mathcal{A}(n)$. The set of all oblivious automata of player 1 of size k (as well as the set of all strategies of player 1 that are defined by oblivious automata of size k) is denoted by $\mathcal{A}^o(k)$. For a finite set $\mathcal{A}$ we denote by $\Delta(\mathcal{A})$ all probability measures on $\mathcal{A}$. $[k]$ denotes the set $\{1, \ldots, k\}$. $\lceil \alpha \rceil$ denotes the smallest integer that is $\geq \alpha$.

For two probability measures P and Q on a measure space X we denote by $\|P - Q\|$ the supremum over measurable $Y \subset X$ of $2(P(Y) - Q(Y))$. If X is a finite or discrete space, $\|P - Q\| = \sum_{x \in X} |P(x) - Q(x)|$. If Q_I and Q_J are measures on the (finite sets or) spaces I and J respectively, then $Q_I \otimes Q_J$ denotes the product measure on $I \times J$ with marginal measures Q_I on I and Q_J on J.

3 The results

The deterministic play induced by a pure strategy $\sigma \in \mathcal{A}(k)$ of player 1 and a pure strategy $\tau \in \mathcal{A}(n)$ of player 2 enters a cycle (of length $\leq kn$) and therefore the average per-stage payoff, $g(\sigma, \tau)$, is well defined. A mixed

strategy $\sigma \in \Delta(\mathcal{A}(k))$ of player 1 and a mixed strategy $\tau \in \Delta(\mathcal{A}(n))$ induce a random play, which is a mixture of periodic plays, and therefore the expected average per-stage payoff is well defined and denoted $g(\sigma, \tau)$.

For $p \in \Delta(I)$ and $\theta \geq 0$ we denote by $\mathcal{Q}(p, \theta)$ the set of all probability measures Q on $I \times J$ such that $Q_I = p$ and $H(Q_I) + H(Q_J) - H(Q) \leq \theta$.

Note that for every distribution $Q \in \Delta(I \times J)$ we have $0 \leq H(Q_I) + H(Q_J) - H(Q)$ with equality iff Q is a product distribution, and $H(Q_I) + H(Q_J) - H(Q) \leq \min(\log|I|, \log|J|)$. The term $H(Q_I) + H(Q_J) - H(Q)$, called the mutual information of i and j, is an information-theoretic quantity that measures the dependence of i and j when Q is the distribution of (i, j).

For $p \in \Delta(I)$ and $\theta \geq 0$ we denote by $v_g(p, \theta)$, or $v(p, \theta)$ for short, the minimum of $E_Q g(i, j)$ where the min ranges over all $Q \in \mathcal{Q}(p, \theta)$, namely,

$$v(p, \theta) = \min_{Q \in \mathcal{Q}(p,\theta)} E_Q g(i, j)$$

and $v_g(\theta)$, or $v(\theta)$ for short, denotes the maximum of $v(p, \theta)$, where the max ranges over all $p \in \Delta(I)$, namely,

$$v(\theta) = \max_{p \in \Delta(I)} v(p, \theta)$$

Note that for every $p \in \Delta(I)$ and $\theta \geq 0$, $v(p, \theta)$ and $v(\theta)$ are functions of the data of the stage game $\langle I, J, g \rangle$. A useful relation is

$$v(p, 0) = \min_j \sum_{i \in I} p(i) g(i, j) \geq v(p, \theta) \geq v(p, 0) - \|g\| \sqrt{2\theta \ln 2}$$

Indeed (see, e.g., Cover and Thomas (1991), p. 300), for every distribution $Q \in \Delta(I \times J)$, $H(Q_I) + H(Q_J) - H(Q) \geq \frac{1}{2 \ln 2} \|Q - Q_I \otimes Q_J\|^2$. Therefore, the inequality $H(Q_I) + H(Q_J) - H(Q) \leq \theta$ implies that $\|Q - Q_I \otimes Q_J\| \leq \sqrt{2\theta \ln 2}$. Thus, $E_Q g(i, j) \geq E_{Q_I \otimes Q_J} g(i, j) - \sqrt{2\theta \ln 2} \|g\| \geq v(Q_I, 0) - \|g\| \sqrt{2\theta \ln 2}$, implying that $v(p, \theta) \geq v(p, 0) - \|g\| \sqrt{2\theta \ln 2}$.

Another property of the function $v(p, \theta)$ is its convexity in θ, which follows from the concavity of the function $Q \mapsto H_Q(i \mid j)$.

The first proposition bounds from below the values of $G[k, n]$ and $G^L[k, n]$ as a function of $\frac{\log n}{k}$.

Proposition 1

$$\mathit{Val}\ G^L[k, n] \geq v\left(\frac{\log n}{k}\right) \quad \mathit{and} \quad \mathit{Val}\ G[k, n] \geq v\left(\frac{\log n}{k}\right)$$

and $\mathit{Val}\ G^L[k, n] \to_{L \to \infty} \mathit{Val}\ G[k, n]$.

The proof is given in Section 4.

The next proposition generalizes Proposition 1. To motivate it, we describe here the family, parameterized by $p \in \Delta(I)$, of oblivious strategies $\sigma(p) \in \Delta(\mathcal{A}^o(k))$ such that $\min_{\tau \in \mathcal{A}(n)} g_L(\sigma(p), \tau) \geq v(p, \frac{\log n}{k})$. For $p \in \Delta(I)$ the strategy $\sigma(p)$ plays a k-periodic sequence $i_1, i_2, \ldots, i_k, \ldots$ such that $i_1, \ldots, i_k$ are iid and the distribution of i_t is p. It follows that 1) the entropy of $(i_1, \ldots, i_k)$ equals $kH(p)$, and 2) the expected empirical distribution of $i_1, \ldots, i_k$ equals p. Under these two conditions on the oblivious strategy σ, a more general inequality holds: for every family T of n mixed strategies of player 2 we have

$$\int \min_{\tau \in T} g_k(s, \tau) \, d\sigma(s) \geq v\left(p, \frac{\log n}{k}\right) \tag{2}$$

The next proposition generalizes inequality (2) to the case when condition 1) holds but $H(i_1, \ldots, i_k) \leq kH(p)$. Given an oblivious strategy σ of player 1 in the repeated game and a positive integer k we denote by $p^k(\sigma)$ the average empirical distribution of $(i_1, \ldots, i_k)$, namely, $p^k(\sigma)[i] = \frac{1}{k} \sum_{t=1}^{k} \Pr_\sigma(i_t = i)$.

Proposition 2 *Let σ be an oblivious mixed strategy of player 1 such that the per-stage entropy of the process $i_1, \ldots, i_k$ is H. Let T be a set of n mixed strategies of player 2 and set $p = p^k(\sigma)$. Then*

$$\int \min_{\tau \in T} g_k(s, \tau) \, d\sigma(s) \geq v(p, \frac{\log n}{k} + H(p) - H)$$

Note that we can assume without loss of generality that the support S of σ is finite. Thus, $\int \min_{\tau \in T} g_k(s, \tau) \, d\sigma(s) = \sum_{s \in S} \sigma(s) \min_{\tau \in T} g_k(s, \tau)$. Note that if $H = H(p)$ (namely, $i_1, \ldots, i_k$ are iid with $\Pr(i_t = i) = p(i)$), then the term $H(p) - H$ in the argument vanishes.

An interesting special case of Proposition 2 is when $n = 1$. This special case implies a characterization both of the oblivious strategies that are approximately optimal in long finitely repeated games and of the oblivious strategies that are optimal in the infinitely repeated game. Indeed, for $n = 1$, Proposition 2 implies that for every oblivious strategy σ of player 1, every positive integer k, and every mixed strategy τ of player 2, we have

$$g_k(\sigma, \tau) \geq v(p^k(\sigma), H(p^k(\sigma)) - H_k(\sigma)) \tag{3}$$

where $H_k(\sigma)$ is the per-stage entropy of the process $i_1, \ldots, i_k$. It follows that if σ^k is a sequence of oblivious strategies of player 1 in the k-stage repetition of the stage game $G = \langle I, J, g\rangle$ and $H(p^k(\sigma^k)) - H_k(\sigma^k) \to 0$ as $k \to \infty$, then

$$\lim_{k\to\infty} (\min_{\tau} g_k(\sigma^k, \tau) - \min_{j} g(p^k(\sigma^k), j)) = 0$$

(see [17, Lemma 3]). Therefore, if $p \in \Delta(I)$ is an optimal strategy of player 1 in the stage game G and σ^k is a sequence of oblivious strategies of player 1 in the k-stage repeated game (respectively, the restriction of an oblivious strategy σ of the infinitely repeated game to the first k-stages) with[2]

$$p^k(\sigma^k) \to p \quad \text{and} \quad H_k(\sigma^k) \to H(p) \tag{4}$$

then σ^k is approximate optimal in the k-stage repeated game (respectively, σ is optimal in the infinitely repeated game) with stage game $G = \langle I, J, g\rangle$, namely,

$$\lim_{k\to\infty} \min_{\tau} g_k(\sigma^k, \tau) = \text{Val } G$$

On the other hand, if σ^k is a sequence of oblivious strategies of player 1 in the k-stage repeated game (respectively, the restriction of an oblivious strategy σ of the infinitely repeated game to the first k-stages) and $p \in \Delta(I)$ and condition (4) does not hold, namely, $\limsup_{k\to\infty} \|p^k(\sigma^k) - p\| + |H(p) - H_k| > 0$, then there is a stage payoff g and $\varepsilon > 0$ such that p is an optimal strategy of the stage game $G = \langle I, J, g\rangle$ but σ^k is not ε-optimal for sufficiently large k (respectively, σ is not optimal in the infinitely repeated game).

If p^* is a unique optimal strategy of the stage game G, then there is $\delta = \delta(G) > 0$ such that $v(p, *) \le v(p^*, 0) - \delta\|p - p^*\|_1$. Therefore, for every oblivious strategy σ we have

$$\min_{\tau} g_k(\sigma, \tau) \le v(p^k(\sigma), 0) \le v(p^*, 0) - \delta\|p - p^*\|_1 \tag{5}$$

and

$$\min_{\tau} g_k(\sigma, \tau) \le v(p^*, 0) - \delta\sqrt{|H_k - H(p^*)|} \tag{6}$$

Therefore, if p is the unique optimal strategy of player 1 in the stage game, then an oblivious strategy σ in the infinitely repeated game is optimal if and only if condition (4) holds.

Now we turn to the main result of the paper.

[2]Condition (4) states, in other words, that the relative entropy (Kullback-Leibler divergence) between σ^k and $p^{\otimes k}$ is $o(k)$ as $k \to \infty$.

Theorem 1 *Let $\theta \geq 0$ and $(n_k)_{k\geq 1}$ with $\frac{\log n_k}{k} \to \theta \geq 0$ as $k \to \infty$. Then*

$$Val\ G[k,n_k] = \max_{\sigma\in\Delta(\mathcal{A}^o(k))} \min_{\tau\in\mathcal{A}(n_k)} g(\sigma,\tau) \to_{k\to\infty} v(\theta) \tag{7}$$

and

$$\begin{aligned} Val\ G[k,n_k] &= \min_{\tau\in\Delta(\mathcal{A}(n_k))} \max_{\sigma\in\mathcal{A}^o(k)} g(\sigma,\tau) \\ &\leq \min_{\tau\in\mathcal{A}(n_k)} \max_{\sigma\in\mathcal{A}^o(k)} g(\sigma,\tau) \to_{k\to\infty} v(\theta) \end{aligned} \tag{8}$$

and, moreover,[3]

$$\exists\tau\in\mathcal{A}(n_k)\ s.t.\ \forall\sigma\in\mathcal{A}^o(k)\ \ g(\sigma,\tau) \leq v(p(\sigma),\theta) + o(1)\ \ as\ k\to\infty \tag{9}$$

and

$$Val\ G^{L_k}[k,n_k] = \min_{\tau\in\Delta(\mathcal{A}(n_k))} \max_{\sigma\in\mathcal{A}^o(k)} g_{L_k}(\sigma,\tau) \to_{\substack{k\to\infty \\ \frac{L_k}{k\log k}\to\infty}} v(\theta) \tag{10}$$

The theorem has four parts, each one of independent importance. The first part, (7), provides the asymptotic behavior of the values of $G[k,n_k]$. The second part, (8), asserts that player 2 has an approximate optimal pure strategy in the game $G[k,n_k]$. Namely, if $n_k \sim 2^{\theta k}$, then player 2 has a pure strategy $\tau \in \mathcal{A}(n_k)$ such that for every strategy $\sigma \in \mathcal{A}^o(k)$ of player 1, $g(\sigma,\tau) \leq v(\theta) + o(1)$. The third part asserts that moreover the approximate optimal strategy $\tau \in \mathcal{A}(n)$ can also exploit the suboptimality of σ. The fourth part, (10), provides an upper bound for the duration of effective learning. Note that in contrast to (8), the approximate optimal strategy of player 2 in the finitely repeated game is a mixture of automata. We do not know if the limit in (10) also holds when the min is over pure strategies $\tau \in \mathcal{A}(n_k)$, nor if the condition $\frac{L_k}{k\log k} \to \infty$ is necessary.

The following lemma is used in the proof of Theorem 1.

Lemma 1 *Let I and J be finite sets. There is a sequence $\varepsilon_k > 0$ that converges to 0 as $k \to \infty$ and there are subsets $S_k(\theta)$, $\theta \geq 0$, of J^k such that*
1) $\theta \leq \eta \implies S_k(\theta) \subset S_k(\eta)$ and $|S_k(\theta)| \leq 2^{\theta k}$, and
2) for every $g : I \times J \to \mathbb{R}$ with $\|g\| \leq 1$ and $\theta \geq 4\varepsilon_k$, for every $i = (i_1,\ldots,i_k) \in I^k$, there is $j = (j_1,\ldots,j_k) \in S_k(\theta)$ such that

$$g(i,j) := \frac{1}{k}\sum_{t=1}^{k} g(i_t,j_t) \leq v_g(e(i),\theta) + \varepsilon_k$$

where $e(i)$ is the empirical distribution of i.

[3]Stronger optimality properties of τ will be discussed later.

4 Proofs

4.1 Proof of Lemma 1

Let $0 < \varepsilon_k \to_{k\to\infty} 0$ with $2^{-2\varepsilon_k k} = o(\varepsilon_k)$ as $k \to \infty$ be such that $k^{|I|} \le (k+1)^{|I\times J|} \le 2^{\varepsilon_k k/2}$ and $|I|^k \exp(-2^{\varepsilon_k k/4}) = o(\varepsilon_k^{|I\times J|+1})$ and $H(Q) - H(Q') \le \varepsilon_k/8$ whenever $\|Q - Q'\| < |J|/k$.

Let $0 \le \delta_k$ be the max of $v_g(p, \theta - 4\varepsilon_k) - v_g(p,\theta)$ where the max is over all $g : I \times J \to \mathbb{R}$ with $\|g\| \le 1$, $p \in \Delta(I)$, and $\theta \ge 4\varepsilon_k$. The function $(g, p, \theta) \mapsto v_g(p,\theta)$ is continuous, and therefore uniformly continuous on (g, p, θ) with $\|g\| \le 1$ and $\theta \le \log|J|$. For $\theta \ge \log|J|$ we have $v_g(p,\theta) = v_g(p, \log|J|)$. Therefore $\varepsilon_k \to 0$ implies that $\delta_k \to 0$ as $k \to \infty$.

Let X_j, $j = (j_1, \ldots, j_k) \in J^k$, be iid random variables that are uniformly distributed on $[0,1]$, and set $\varepsilon = \varepsilon_k$. First we define random subsets $S_k(\theta)$ of J^k that depend on the values of X_j:

$$j = (j_1, \ldots, j_k) \in S_k(\theta) \text{ iff } \log X_j \le (\theta - 3\varepsilon - H(e(j)))k$$

where $e(j)$ is the empirical distribution of j.

The definition of $S_k(\theta)$ implies that

$$\Pr(j \in S_k(\theta)) \le 2^{(\theta - 3\varepsilon - H(e(j)))k}$$

and that $S_k(\eta) \subset S_k(\theta)$ whenever $\eta < \theta$.

For every $q \in \mathbb{T}_k(J) := \{e(j) : j \in J^k\}$ the number of elements in $T_k(q) := \{j \in J^k : e(j) = q\}$ is $\le 2^{H(q)k}$ and therefore the expected number of elements of $S_k(\theta) \cap T_k(q)$ is $\le 2^{(\theta-3\varepsilon)k}$. The number of elements of $\mathbb{T}_k(J)$ is $\le k^{|J|}$. Therefore, the expected number of elements of $S_k(\theta)$ is $\le k^{|J|} 2^{(\theta-3\varepsilon)k}$. As $k^{|J|} \le 2^{\varepsilon k}$ we deduce that

$$E|S_k(\theta)| \le 2^{(\theta-2\varepsilon)k}$$

By Markov inequality we have that

$$\Pr(|S_k(\theta)| > 2^{\theta k}) < 2^{-2\varepsilon k} = o(\varepsilon)$$

Therefore

$$\Pr(\exists \ell \in \mathbb{N} \text{ s.t. } |S_k(\ell\varepsilon)| > 2^{\ell\varepsilon k}) < 2^{-2\varepsilon k} O(\varepsilon^{-1}) = o(1)$$

Fix $g : I \times J \to \mathbb{R}$ and $i = (i_1, \ldots, i_k) \in I^k$. For notational convenience we define $v_g(p, x)$ to be equal to $v_g(p, 0)$ when $x < 0$. What is the probability that there is $j \in S_k(\theta - \varepsilon)$ with $\frac{1}{k}\sum_{t=1}^{k} g(i_t, j_t) \leq v_g(e(i), \theta - 4\varepsilon) + \frac{|J|}{k}\|g\|$?

The definition of v_g implies that there is a distribution $Q' \in \mathcal{Q}(e(i), \theta - 4\varepsilon)$ s.t. $E_{Q'} g(i, j) \leq v_g(e(i), \theta - 4\varepsilon)$. There is $Q \in \mathbb{T}_k(I \times J)$ with $Q_I = Q'_I$ and $\|Q - Q'\| \leq |J|/k$. Therefore, $E_Q g(i, j) \leq E_{Q'} g(i, j) + \frac{|J|}{k}\|g\| \leq v_q(e(i), \theta - 4\varepsilon) + \frac{|J|}{k}\|g\|$. By the choice of $\varepsilon = \varepsilon_k$ and Q', $H(Q_I) + H(Q_J) - H(Q) \leq H(Q_I) + H(Q'_J) - H(Q') + \varepsilon/4 \leq \theta - 4\varepsilon + \varepsilon/4$. Set $q = Q_J$.

Fix $\theta \geq 4\varepsilon$. Let us compute the probability that there is $j = (j_1, \ldots, j_k) \in S_k(\theta)$ such that the empirical distribution of $(i_1, j_1), \ldots, (i_k, j_k)$, denoted $e(i, j)$, equals Q. The number of elements $j = (j_1, \ldots, j_k) \in T_k(q)$ is $\leq 2^{H(q)k}$. The number of elements $j = (j_1, \ldots, j_k) \in T_k(q)$ with $e(i, j) = Q$ is $|T_k(Q)|/|T_k(e(i))|$. Note that

$$|T_k(Q)|/|T_k(p(i))| \geq 2^{(H(Q) - H(e(i)))k}/(k+1)^{|I \times J|} \geq 2^{(H(Q) - H(e(i)) - \varepsilon/2)k}$$

where the last inequality holds by the choice of $\varepsilon = \varepsilon_k$.

If $\theta - 3\varepsilon - H(q) \geq 0$ then $S_k(\theta) \supset T_k(q)$. Otherwise, for every fixed j with $e(j) = Q_J$, the probability that $j \notin S_k(\theta)$ is the probability that $\log X_j > (\theta - 3\varepsilon - H(Q_J))k$, which equals $1 - 2^{(\theta - 3\varepsilon - H(Q_J))k}$. Therefore the probability that $j \notin S_k(\theta)$ for every $j = (j_1, \ldots, j_k) \in T_k(q)$ with $e(i, j) = Q$ is

$$\begin{aligned} & (\Pr(\log X_j > (\theta - 3\varepsilon - H(Q_J))k))^{|T_k(Q)|/|T_k(e(i))|} \\ = \; & (1 - 2^{(\theta - 3\varepsilon - H(Q_J))k})^{|T_k(Q)|/|T_k(e(i))|} \\ \leq \; & \exp(-2^{(\theta - 3\varepsilon - H(Q_J))k} 2^{(H(Q) - H(e(i)) - \varepsilon/2)k}) \leq \exp\left(-2^{\varepsilon k/4}\right) \end{aligned}$$

where the last inequality uses $Q \in \mathcal{Q}(Q_I, \theta - 4\varepsilon + \varepsilon/4)$ and $e(i) = Q_I$.

Therefore the probability that there is $i \in I^k$ such that for every $j \in S_k(\theta)$ we have

$$\frac{1}{k}\sum_{t=1}^{k} g(i_t, j_t) > v_g(e(i), \theta - 4\varepsilon) + \frac{|J|}{k}\|g\|$$

is less than $|I|^k \exp\left(-2^{\varepsilon k/4}\right) = o(\varepsilon^{|I \times J| + 1})$.

Let Y be an ε grid of all functions $g : I \times J \to \mathbb{R}$ with $\|g\| \leq 1$ with at most $(3/\varepsilon)^{|I \times J|}$ elements. Then, the probability that there is $g \in Y$ and $i \in I^k$ such that for every $j \in S_k(\theta)$ we have

$$\frac{1}{k}\sum_{t=1}^{k} g(i_t, j_t) > v_g(e(i), \theta - 4\varepsilon) + \frac{|J|}{k}\|g\|$$

is

$$\leq (3/\varepsilon)^{|I\times J|}o(\varepsilon^{|I\times J|+1}) = o(\varepsilon) \text{ as } k \to \infty$$

Therefore, the probability that there is g with $\|g\| \leq 1$ and $i \in I^k$ such that for every $j \in S_k(\theta)$ we have

$$\frac{1}{k}\sum_{t=1}^{k} g(i_t, j_t) > v_g(e(i), \theta - 4\varepsilon) + \frac{|J|}{k}\|g\| + \varepsilon$$

is

$$\leq o(\varepsilon) \text{ as } k \to \infty$$

Therefore, the probability that there is g with $\|g\| \leq 1$, $i \in I^k$, and $1 + \log|J| \geq \ell \in \mathbb{N}$ such that the inequality

$$\frac{1}{k}\sum_{t=1}^{k} g(i_t, j_t) > v_g(e(i), \ell\varepsilon - 4\varepsilon) + \frac{|J|}{k}\|g\| + \varepsilon$$

holds for all $j \in S_k(\theta)$ is

$$\leq o(1) \text{ as } k \to \infty$$

Define $\hat{S}_k(\theta) = S_k([\theta/\varepsilon]\varepsilon)$ where $[*]$ is the integr part of $*$. Note that $S_k(\theta - \varepsilon) \subset \hat{S}_k(\theta) \subset S_k(\theta)$. Therefore the sets $\hat{S}_k(\theta)$ satisfy 1) of Lemma 1. The probability that the set $\hat{S}_k(\theta)$ satisfies condition 2) of Lemma 1 for the sequence $\hat{\varepsilon}_k = \varepsilon_k + \frac{|J|}{k}$ is close to 1. In particular, there is a realization of X_j such that $\hat{S}_k(\theta)$ satisfies condition 2) of Lemma 1.

4.2 Proof of Proposition 1

Fix $p \in \Delta(I)$. Let $X_1, X_2, \ldots, X_k, X_{k+1}, \ldots$ be a k-periodic sequence of I-valued random variables with $P(X_1 = i) = p(i)$ and $X_1, X_2, \ldots, X_k$ iid. Therefore, for any positive integer d, $X_{d+1}, \ldots, X_{d+k}$ are iid. The mixed strategy σ of player 1 is to play the action $X_t \in I$ at stage t. Obviously, σ is a mixture of strategies that are defined by oblivious automata of size k. Let $\tau \in \mathcal{A}(n)$. Consider the probability distribution induced on plays $(i_1, j_1, \ldots, i_t, j_t, \ldots)$ by the strategy σ and the pure strategy τ. Fix a positive integer d. The expected average payoff in stages $t = d+1, \ldots, d+k$ is

$$E_{\sigma,\tau}\frac{1}{k}\sum_{t=1}^{k} g(i_{d+t}, j_{d+t})$$

Let Q_t be the expected distribution of (i_{d+t}, j_{d+t}) and $Q = \frac{1}{k}\sum_{t=1}^{k} Q_t$. Then

$$E_{\sigma,\tau}\frac{1}{k}\sum_{t=1}^{k} g(i_{d+t}, j_{d+t}) = E_Q g(i,j)$$

Obviously, $Q_I = p$. Next we prove that $\frac{\log n}{k} \geq H(Q_I) + H(Q_J) - H(Q)$ $(= H_Q(i) + H_Q(j) - H_Q(i,j) = H_Q(i) - H_Q(i \mid j) = H(p) - H_Q(i \mid j))$.

$$\begin{aligned}
H_Q(i \mid j) &\geq \frac{1}{k}\sum_{t=1}^{k} H_{Q_t}(i \mid j) = \frac{1}{k}\sum_{t=1}^{k} H(i_{d+t} \mid j_{d+t}) \\
&\geq \frac{1}{k}\sum_{t=1}^{k} H(i_{d+t} \mid i_{d+1}, \ldots, i_{d+t-1}, m_{d+1}) \\
&= \frac{1}{k} H(i_{d+1}, \ldots, i_{d+k} \mid m_{d+1}) \\
&= \frac{1}{k} H(i_{d+1}, \ldots, i_{d+k}, m_{d+1}) - \frac{1}{k} H(m_{d+1}) \\
&\geq H(p) - \frac{\log n}{k} \;=\; H_Q(i) - \frac{\log n}{k}
\end{aligned}$$

The first inequality (above) follows from the concavity (with respect to the underlying probability distribution) of the conditional entropy. The second inequality follows from the fact that j_{d+t} is a function of $(i_{d+1}, \ldots, i_{d+t-1}, m_{d+1})$. The following two equalities follow from the chain rule (additivity rule) of entropies. The last inequality follows from the fact that the number of possible values of m_{d+1}, the state of the automata of player 2 at stage $d+1$, is $\leq n$, and thus $H(m_{d+1}) \leq \log n$.

Therefore,

$$H_Q(i) - H_Q(i \mid j) = H(Q_I) + H(Q_J) - H(Q) \leq \frac{\log n}{k}$$

Therefore,

$$E_{\sigma,\tau}\frac{1}{k}\sum_{t=1}^{k} g(i_{d+t}, j_{d+t}) = E_Q g(i,j) \geq v(p, \frac{\log n}{k})$$

For every $r \leq k$ the finite sequence $X_1, \ldots, X_r$ is an iid sequence with $\Pr(X_t = i) = p(i)$ and therefore

$$E_{\sigma,\tau}\frac{1}{r}\sum_{t=1}^{r} g(i_t, j_t) \geq v(p,0) \geq v(p, \frac{\log n}{k})$$

Fix L. There are nonnegative integers ℓ and $r < k$ such that $L = r + \ell k$. Note that

$$\begin{aligned} E_{\sigma,\tau} \sum_{t=1}^{L} g(i_t, j_t) &= E_{\sigma,\tau} \sum_{t=1}^{r} g(i_t, j_t) + \sum_{d=0}^{\ell-1} E_{\sigma,\tau} \sum_{t=1}^{k} g(i_{r+dk+t}, j_{r+dk+t}) \\ &\geq Lv(p, \frac{\log n}{k}) \end{aligned}$$

and therefore

$$E_{\sigma,\tau} \frac{1}{L} \sum_{t=1}^{L} g(i_t, j_t) \geq v(p, \frac{\log n}{k})$$

The function $p \mapsto v(p, \theta)$ is continuous in p and therefore attains its max. Therefore there is p^* with $v(p^*, \frac{\log n}{k}) = \max_{p \in \Delta(I)} v(p, \frac{\log n}{k}) = v(\frac{\log n}{k})$, and thus

$$\text{Val } G^L[k, n] \geq v(\frac{\log n}{k})$$

and

$$\text{Val } G[k, n] \geq v(\frac{\log n}{k})$$

The convergence of the sequence of values of $G^L[k, n]$ to the value of $G[k, n]$ follows from the fact that the play defined by strategies $\sigma \in \mathcal{A}(k)$ and $\tau \in \mathcal{A}(n)$ enters a cycle of length $\leq kn$ within k stages. Therefore the L-stage payoff $g_L(\sigma, \tau)$ is within $\frac{2kn}{L}\|g\|$ of the payoff $g(\sigma, \tau)$ in the infinite repeated game.[4] □

4.3 Proof of Proposition 2

First assume that T is a set of n pure strategies of player 2. Let τ be a function from $\{(i_1, \ldots, i_k)\} \to T$. Let Q_t be the distribution of (i_t, j_t) induced by σ and $\tau(i_1, \ldots, i_t)$, and $Q = \frac{1}{k} \sum_{t=1}^{k} Q_t^\tau$. Repeating the argument of the proof of Proposition 1 we have

$$H_Q(i \mid j) \geq H - \frac{\log n}{k}$$

and thus

$$H(Q_I) + H(Q_J) - H(Q) \leq H(p) - H + \frac{\log n}{k}$$

[4]The same argument also shows that VAL $G^L(k, n) \to_{L \to \infty}$ VAL $G^L(k, n)$.

which proves the proposition when T is a set of n pure strategies.

Let S be the support of the mixed strategy σ, $T = \{\tau_1, \ldots, \tau_n\}$, and $f : S \to \{1, \ldots, n\}$ (measurable) such that

$$\int \min_{\tau \in T} g_k(s, \tau) d\sigma(s) = \int g_k(s, \tau_{f(s)}) d\sigma(s)$$

Let $T_1, \ldots, T_n$ be the support of $\tau_1, \ldots, \tau_n$ respectively, and $\tau := \tau_1 \otimes \ldots \otimes \tau_n$ the product probability on $T_1 \times \ldots \times T_n$. For every $\omega = (t_1, \ldots, t_n) \in T_1 \times \ldots \times T_n$ let $T(\omega)$ be the finite set of pure strategies $\{t_1, \ldots, t_n\}$. Note that

$$g(s, \tau_{f(s)}) = \int g(s, t_{f(s)}) \, d\tau(\omega)$$

Therefore

$$\begin{aligned} \int g(s, \tau_{f(s)}) \, d\sigma(s) &= \int g(s, t_{f(s)}) \, d\tau(\omega) \, d\sigma(s) \\ &\geq \int \min_{t \in T(\omega)} g(s, t) \, d\tau(\omega) \, d\sigma(s) \\ &= \int \min_{t \in T(\omega)} g(s, t) \, d\sigma(s) \, d\tau(\omega) \\ &\geq v(p, \frac{\log n}{k} + H(p) - H) \end{aligned}$$

□

4.4 Proof of Theorem 1

Assume that $\frac{\log n_k}{k} \to_{k \to \infty} \theta > 0$. As

$$\min_{\tau \in \mathcal{A}(n_k)} \max_{\sigma \in \mathcal{A}^o(k)} g(\sigma, \tau) \geq \text{Val } G[k, n_k] = \max_{\sigma \in \Delta(\mathcal{A}^o(k))} \min_{\tau \in \mathcal{A}(n_k)} g(\sigma, \tau)$$

in order to prove (7) and (8), it suffices to prove that

$$\liminf_{k \to \infty} \text{Val } G[k, n_k] \geq v(\theta) \tag{11}$$

and

$$\limsup_{k \to \infty} \min_{\tau \in \mathcal{A}(n_k)} \max_{\sigma \in \mathcal{A}^o(k)} g(\sigma, \tau) \leq v(\theta) \tag{12}$$

Note that the function $Q \mapsto E_Q g(i,j)$ is continuous on $\Delta(I \times J)$. For every $p \in \Delta(I)$ and $\theta \geq 0$ the set $\mathcal{Q}(p,\theta)$ is a closed subset of $\Delta(I \times J)$ and the map $(p,\theta) \mapsto \mathcal{Q}(p,\theta)$ is continuous in the Hausdorff topology on the closed subsets of $\Delta(I \times J)$. Therefore, the functions $(p,\theta) \mapsto v(p,\theta)$ and $\theta \mapsto v(\theta)$ are continuous. Thus $v(\frac{\log n_k}{k}) \to_{k\to\infty} v(\theta)$ and $v(p, \frac{\log n_k}{k}) \to_{k\to\infty} v(p,\theta)$.

By Proposition 1 we have Val $G^L[k,n_k] \geq v(\frac{\log n_k}{k}) \to_{k\to\infty} v(\theta)$ and Val $G[k,n_k] \geq v(\frac{\log n_k}{k}) \to_{k\to\infty} v(\theta)$. Therefore,

$$\liminf_{k\to\infty} \text{Val } G^{L_k}[k,n_k] \geq v(\theta)$$

and

$$\liminf_{k\to\infty} \text{Val } G[k,n_k] \geq v(\theta)$$

To prove (12), we construct for every sufficiently large k a pure strategy $\tau^k \in \mathcal{A}(n_k)$ such that if $\sigma \in \mathcal{A}^o(k)$ generates a sequence of actions $i = (i_1, i_2, \ldots)$ with empirical distribution $e(i) \in \Delta(I)$, then

$$g(\sigma, \tau^k) \leq v(e(i), \theta - \eta_k) + \eta_k$$

where $\eta_k \to 0$ as $k \to \infty$.

Set $\theta_k = \frac{\log n_k - (2+|I\times J|)\log k}{k}$. We order the elements of $\cup_{k/2<r\leq k} S_r(\theta_k)$ so that all elements of $S_r(\theta_k)$ precede the elements of $S_{r'}$ whenever $r > r'$. The first element of $S_k(\theta_k)$ (and thus of $\cup_{k/2<r\leq k} S_r(\theta_k)$) is denoted j^*. The successor of an element $j \in \cup_{k/2<r\leq k} S_r(\theta_k)$ (but the last one) is denoted j'.

For every $j \in S_r(\theta_k) \subset J^r$ we define a sub-automaton A_j with $r^{|I\times J|+1}$ states $\{(j,1), \ldots, (j,r)\} \times \{0,1,\ldots,r-1\}^{I\times J}$ (where the second factor stands for all functions from $I \times J$ to $\{0,1,\ldots,r-1\}$.

Informally, the sub-automaton A_j, where $j \in S_r(\theta_k)$, tests whether the sequence i defined by the oblivious automaton $\sigma \in \mathcal{A}^o(k)$ of player 1 is r-periodic and whether it results, when matched with the r-periodic play defined by j (repeatedly), in a payoff $\leq v(e(i),\theta_k) + \varepsilon_r$ where $(\varepsilon_r)_r$ is the sequence given by Lemma 1.

The initial state of the sub-automaton A_j is $(j,1,0)$, where 0 is the constant 0 function. The action function of the sub-automaton A_j is

$$\alpha(j,t,*) = j_t$$

To simplify the definition of the transition, we introduce the following notations. For $j \in J^r$, $t \in [r]$, $x : I \times J \to \{0,\ldots,r\}$, and $i \in I$ we

denote by $x'(j,t,x,i)$, or x' for short, the function $x + 1_{i,j_t}$ where $1_{i,j_t}$ is the indicator function (1 on (i,j_t) and 0 elsewhere). For a nonnegative nonzero real-valued function $x : I \times J \to \mathbb{R}_+$ we define $e(x) \in \Delta(I \times J)$ by $e(x)[i,j] = x(i,j)/\sum_{(i,j)\in I\times J} x(i,j)$, and $e_I(x)$ is the marginal of $e(x)$ on I, namely, the element of $\Delta(I)$ defined by $e_I(x)[i] = \sum_{j\in J} e(x)[i,j]$.

The transition function of the sub-automaton A_j is

$$\beta(j,t,x,i) = \begin{cases} (j,t+1,x') & \text{if } t < r \\ (j,1,0) & \text{if } t = r \text{ and } g(x') \le v(e_I(x'),\theta_k) + \varepsilon_r \\ (j',1,0) & \text{if } t = r \text{ and } g(x') > v(e_I(x'),\theta_k) + \varepsilon_r \end{cases}$$

Recall that $(j',1,0)$ is the initial state of the sub-atomaton $A_{j'}$.

Informally, the automaton $\tau \in \mathcal{A}(n_k)$ of player 2 will first count up to k, by which time an oblivious automaton of player 1 will already have entered a cyclic play. The period of this cyclic play is a positive integer $1 \le r \le k$. However, an r-periodic play is also a $2r$-periodic play. Therefore, this cyclic play has a cycle of length r with $k/2 < r \le k$. The automaton τ of player 2 will look for a periodic play with period $k/2 < r \le k$, by starting testing k-periodic plays, and recursively, if no successful match to the r-periodic assumption is found, moving to searching for an $(r-1)$-periodic play. The construction of the automaton is obtained by gluing together the sub-automata A_j, where j ranges over all elements of $\cup_{k/2<r\le k} S_r(\theta_k)$.

The set of states of A_j is denoted M_j. Set $M_r = \cup_{j\in S_r(\theta_k)} M_j$ and $M = [k] \cup \cup_{k/2<r\le k} M_r$. Note that $|M| \le k + \sum_{k/2<r\le k} |M_r| \le k + \sum_{k/2<r\le k} |S_r(\theta_k)| r^{|I\times J|+1} \le k + \sum_{k/2<r\le k} 2^{\theta_k r} r^{|I\times J|+1} \le k^{2+|I\times J|} 2^{\theta_k k} \le n_k$ (where the second to last inequality holds for $k \ge 2$).

Without loss of generality we assume that $\|g\| \le 1$. Property 2) of Lemma 1 guarantees that the play induced by an oblivious automaton $\sigma \in \mathcal{A}(k)$ and the constructed automaton $\tau \in \mathcal{A}(n_k)$ of player 2 enters a cyclic play with a payoff $\le v_g(e(i), \theta - \eta_k) + \eta_k$, where $\eta_k = \max(\theta - \theta_k, \max_{k/2<r\le k} \varepsilon_r)$.

Therefore, for every strategy $\sigma \in \mathcal{A}^o(k)$ we have

$$g(\sigma,\tau) \le v_g(e(i), \theta - \eta_k) + \eta_k$$

Note that $\eta_k \to_{k\to\infty} 0$ and recall that the function $\theta \to v(\theta)$ is continuous. Therefore

$$g(\sigma,\tau) \le v_g(e(i),\theta) + o(1) \text{ as } k \to \infty$$

which proves (8). It follows in addition that

$$\min_{\tau\in\mathcal{A}(n_k)}\max_{\sigma\in\mathcal{A}^o(k)} g(\sigma,\tau)\le v(\theta-\eta_k)+\eta_k$$

Therefore,

$$\limsup_{k\to\infty}\min_{\tau\in\mathcal{A}(n_k)}\max_{\sigma\in\mathcal{A}^o(k)} g(\sigma,\tau)\le v(\theta)$$

This completes the proofs of (7) and (8). □

The proof of (10) is presented in the next section after a short discussion regarding learning duration.

4.5 The learning duration

The 'minimal' learning duration is derived from the asymptotic behavior of the values of $G^L(k,n)$. The value of $G^L(k,n)$ is $\ge v(\frac{\log n}{k})$ and its limit as $L\to\infty$ exists. This limit is $\le v(\frac{\log n}{k})+o(1)$ as $k\to\infty$.

We study the asymptotic condition on the duration $L=L(k,n)$ such that the value of $G^L(k,n)$ is close to $v(\frac{\log n}{k})$.

The approximate optimal strategy τ that is constructed in Section 3 accomplishes its learning duration in about $2^{\theta k}$ stages. Therefore, if $\liminf\frac{\log L_k}{k}\ge\theta$ then

$$G^{L_k}(k,n_k)\to v(\theta)$$

We will show that the much weaker asymptotic relation, $\lim\frac{L_k}{k\log k}\to\infty$, suffices. The learning comprises two phases. The first phase determines the length of a cycle. The second phase searches for a cyclic sequence of actions that leads to an average per-stage payoff of no more than $v(\theta)+o(1)$, where p is the empirical distribution of $i_1,i_2,\ldots$.

4.6 Learning the length of a cycle

The learning of the length of the cycle is probabilistic. If there is a cycle of length $k/2<r\le k$ the randomly selected (deterministic) automaton of player 2 will find this length r of the cycle with probability close to 1. For this part of the learning a duration $\gg k\log k$ suffices.[5] Once the length r of the cycle is discovered an additional $\gg r$ stages suffice for guaranteeing an average per-stage payoff $\le v(\theta)+o(1)$.

[5]We do not know if it is also necessary.

Assume that if $k/2 < r \leq k$ is not the length (namely, not a multiple) of a (the) cycle of the play of player 1. Then for any $s \geq k$ there is a stage $s < t \leq s + k$ such that $i_t \neq i_{t+r}$. Therefore, if we pick at random a subset T of integers such that for every $s < t \leq s + k$ the probability that $t \in T$ is $\geq \delta > 0$ (e.g., pick at random an integer $s - \delta k \leq t_0 \leq s + k$, each equally likely, and choose $T = \{t_0 + 1, \ldots, t_0 + \lceil 2\delta k \rceil\}$ for k sufficiently large and $\delta > 0$ sufficiently small), the probability that there is $t \in T$ with $i_t \neq i_{t+r}$ is at least δ. Therefore, by performing sufficiently many such checks (say K) we can rule out each non-cyclic length r with probability close to 1 ($\geq (1-\delta)^K$).

Fix a sufficiently small $\delta > 0$ and set $K = \lceil \frac{2}{\delta} \ln k \rceil$ and $\bar{q} = \lceil 1/(2\delta) \rceil$ (the least integer $\geq 1/(2\delta)$).

The random automaton will pick a list of integers t_j^q, with $q = 1, \ldots, \bar{q}$ and $j = 1, \ldots, K$, with

$$2k + 4(q-1)kK + 4(j-1)k < t_j^q < 4(q-1)kK + 4jk$$

all such lists equally likely. Note that $t_j^q \leq 4\bar{q}kK \leq \frac{4}{\delta}kK \leq \frac{9}{\delta^2}k \log k$.

The cycle learning phase is done in stages $t \leq 4kK/\delta$. This learning phase is partitioned into sub-phases, indexed by the values of q and j. Set $T_{qj}^- = \{t_j^q + 1, \ldots, t_j^q + \lceil \delta k \rceil\}$, $C_q = \{k - q\lceil \delta k \rceil + 1, \ldots, k - (q-1)\lceil \delta k \rceil\}$, and $T_{qj}^+ = T_{qj}^- + C_q$ (where $+$ is the algebraic sum) and $T_{qj} = T_{qj}^- \cup T_{qj}^+$. Note that $|T_{qj}| = 3\lceil \delta k \rceil$ and $|C_q| = \lceil \delta k \rceil$. In the qj-th sub-phase, the automaton records the actions of player 1 in stages $t \in T_{qj}$, and eliminates from the set of possible cycles all $r \in C_q$ for which a contradiction is discovered. Thereafter, no further use of the recorded play in this sub-phase. Note that $t_K^{\bar{q}} + 2k \leq 4kK/\delta$ and thus all the $\bar{q}K$ sub-phases of checks end before stage $[4kK/\delta]$.

Note that $\cup_{q=1}^{\bar{q}} C_q \supset \{[k/2], \ldots, k\}$, and the number of subsets of C_q is $2^{\lceil \delta k \rceil}$.

The number of sequences of $|T_{qj}|$ elements of I is $|I|^{|T_{qj}|} = |I|^{3\lceil \delta k \rceil}$, and the number of subsets of the set C_q is $2^{\lceil \delta k \rceil}$.

The set of automata states is the Cartesian product of three sets. The first factor is the set $M_1 = \{1, \ldots, [4kK/\delta]\}$; it has $[4kK/\delta]$ elements and it enables the automaton to count the first $[4kK/\delta]$ stages of the repeated game.

The second factor is the set $M_2 = I^{3\lceil \delta k \rceil}$; it has $|I|^{3\lceil \delta k \rceil}$ elements and it enables the automaton to record the actions of player 1 in stages $t \in T_{qj}$. Note that the natural ordering of the stages $t \in T_{qj}$ enables us to identify $I^{T_{qj}}$ with

the factor $M_2 = I^{3\lceil \delta k \rceil}$ of the automaton; a list $(i_t)_{t \in T_{qj}}$ is identified with the element $(i_{\hat{t}}) \in I^{3\lceil \delta k \rceil}$ where for $t \in T_{qj}$ we set $\hat{t} = |\{1, \ldots, t\} \cap T_{qj}|$. Note that this identification maps an element $y \in M_2 = I^{3\lceil \delta k \rceil}$ to an element $\tilde{y} \in I^{T_{qj}}$ and for $r \in C_q$ and $t \in T_{qj}^-$ we have $\tilde{y}_t = \tilde{y}_{t+r}$ iff $y_{\hat{t}} = y_{\hat{t}+r-(1+k-(q+1)\lceil \delta k \rceil)}$.

The third factor is the set M_3 of all subsets of the set $\{1, \ldots, \lceil \delta k \rceil\}$; it has $2^{\lceil \delta k \rceil}$ elements and for every integer $1 \leq q \leq 1/\delta$ the unique 1-1 order-preserving map from M_3 into C_q identifies M_3 with C_q. This identification enables this factor of the automaton (together with the 'recorded' play in stages T_{qj}) to keep track, following the play in stage $t_{qj} = t_j^q + k - (q-2)\lceil \delta k \rceil$ (the last stage in T_{qj}), of all cycle lengths $r \in C_q$ that are compatible for every $j' \leq j$ with the play in stages $T_{qj'}$.

In sub-phases $1j$, $j = 1, \ldots K$, the automaton searches a cycle length $r \in C_1$ that is compatible with the play in each sub-phase $T_{1j'}$. If such a compatible cycle length r is found, the automaton moves to the first state of an automaton M^r. If $r \in C_1$ is not a length of a cycle, the probability that the cycle length r is compatible with the play in stages T_{1j} (namely, $i_t = i_{t+r}$ for every $t \in T_{1j}^-$) is $\leq 1 - \delta$. The events "the cycle length r is compatible with the play in stages T_{1j}", $j = 1, \ldots, K$, are independent. Therefore the probability that a non-cycle length r is compatible for each $j = 1, \ldots, K$ with the play in T_{1j} (namely, $i_t = i_{t+r}$ for every $t \in \cup_{j=1}^K T_{1j}^-$) is $\leq (1-\delta)^K$, which is about $\frac{1}{k^2}$. Therefore the expected number of 'undiscovered' non-periods $r \in C_1$ is $\leq |C_1|/k^2 < 2\delta/k$.

If all $r \in C_{q-1}$ are excluded as periods, we proceed to test the possible periods $r \in C_q$ in sub-phases qj , $j = 1, \ldots K$.

Let us describe the transitions of the automaton (as a function of the random integers t_j^q) in states $(t, x, D) \in M_1 \times M_2 \times M_3$. The initial state is $(1, x^0, C_1)$ (where x^0 is an element of $I^{3\lceil \delta k \rceil}$).

For $t \notin \cup_{qj} T_{qj}$, the transition from a state (t, x, D) is simply effected by adding 1 to the stage counter t and leaving x and D unchanged:

$$\beta((t, x, D), *) = (t+1, x, D) \text{ if } t \notin \cup_{qj} T_{qj}$$

For $x \in I^{3\lceil \delta k \rceil}$, $t \in T_{qj}$, and $i \in I$ we denote by $x'(x, t, i)$ (or $x'(t, i)$ or x' for short) the element $y \in I^{3\lceil \delta k \rceil}$ with $y_s = x_s$ for $s \neq \hat{t}$ and $y_s = i$ for $s = \hat{t}$.

For $t_{qj} \neq t \in T_{qj}$, the transition from a state (t, x, D) is to the state $(t+1, x', D)$, which leaves the third coordinate D unchanged, adds 1 to the stage counter t, and replaces the sequence of I elements x with the sequence

$x' = x'(t,i)$:

$$\beta((t,x,D),i) = (t+1,x',D) \text{ if } t \in \cup_{qj}(T_{qj} \setminus \{t_{qj}\})$$

For $q = 1,\ldots\bar{q}$, $D \subset C_q$, and $y \in I^{3\lceil\delta k\rceil}$ we denote by $D'(q,D,y)$ the subset of D consisting of all $r \in D$ such that $y_{\hat{t}} = y_{\hat{t}+r-(1+k-(q+1)\lceil\delta k\rceil)}$ for every $t \in T^-_{qj}$.

The transition function from a state (t,x,D) where $t = t_{qj}$ depends on whether $j = K$ or not. If $j < K$, the transition adds 1 to the stage counter t, and modifies the set $D \subset C_q$ of non-excluded cycle lengths with its subset $D'(q,D,x'(t,i))$:

$$\beta((t_{qj},x,D),i) = (t+1,x',D') \text{ if } j < K$$

where $x' = x'(t,i)$ and $D' = D'(q,D,x'(t,i))$.

The transition function from a state (t,x,D) where $t = t_{qj}$ and $j = K$ depends on whether $D' = \emptyset$ or not. If $D' = \emptyset$ we move to state $(t+1,x^0,C_{q+1})$, namely,

$$\beta((t_{qK},x,D),i) = (t+1,x^0,C_{q+1}) \text{ if } D' = \emptyset$$

If $D' \neq \emptyset$ we move to the first state of the sub-automaton M^r, where r is a non-excluded cycle length C_q.

The probability of accepting a non-period $r \in C_q$ as a period of the sequence i_t is $\leq \delta/k$. Therefore the probability of approving an erroneous cycle length $k/2 < r \leq r$ is $\leq \lceil\frac{k/2}{\delta k}\rceil\delta/k \leq 1/k$.

The number of states of the automaton used in the first phase is $|M_1 \times M_2 \times M_3|$, which for sufficiently large k is $\leq 2^{5\delta k \log|I|}$

4.7 Effective learning when cycle length is known

The second phase starts following the specification of a cycle length $k/2 < r \leq k$. For every cycle length r we define an automaton M^r.

We partition the cycle length $k/2 < r \leq k$ into ℓ blocks of size $r_1,\ldots,r_\ell$ (we may assume $r_1 \leq r_2 = \ldots = r_\ell$), and run the learning on the blocks in tandem. We have $\sum_{s=1}^{\ell} r_s = r$ and we denote $R_s = \sum_{s' \leq s} r_{s'}$. We use the notations of Lemma 1 and the proof of Theorem 1. Let $0 < \varepsilon_k \to 0$ as $k \to \infty$ and $S_{r_s}(\theta_{r_s}) \subset J^{r_s}$ be such that for every $i \in I^{r_s}$ there is $j \in S_{r_s}(\theta_{r_s})$ such that

$$g(i,j) \leq v_g(e(i),\theta_{r_s}) + \varepsilon_{r_s}$$

Given a list $j = (j^1, \ldots, j^\ell)$ with $j^s \in S_{r_s}(\theta_{r_s})$ we define a sub-automaton A^j, which is a kind of product of the sub-automata A_{j^s} used in the proof of Theorem 1.

The states of the sub-automaton A^j are the triples (j, t, x) where $1 \leq t \leq r$, and for $R_{s-1} < t \leq R_s$ the term x is a function from $I \times J$ to $\{0, \ldots, r_s - 1\}$. We order the elements of each set $S_{r_s}(\theta_{r_s})$ and its first element is denoted j^{s*}, and $j^* = (j^{1*}, \ldots, j^{\ell*})$. For $j = (j^1, \ldots, j^\ell) \in \times_{s=1}^{\ell} S_{r_s}(\theta_{r_s})$ and $1 \leq s \leq \ell$ we denote by j'^s the list j where only its s-th component j^s is replaced by the successor of j^s in the order of $S_{r_s}(\theta_{r_s})$.

The number of states of A^j is $\leq r^{|I \times J|+1}$. Therefore the number of states of the automaton A that is composed of the sub-automata A^j is $\leq (\prod_{s=1}^{\ell} |S_{r_s}(\theta_{r_s})|) r^{|I \times J|+1} \leq r^{|I \times J|+1} 2^{r \max_s \theta_{r_s}}$.

The initial state of the automaton M^r is $(j^*, 0, 0)$. The action function is defined by

$$\alpha(j, t, x) = j_t$$

The transition is defined as follows.

$$\beta(j, t, x, i) = \begin{cases} (j, t+1, x') & \text{if } R_s < t < R_{s+1} \\ (j, t+1, 0) & \text{if } t = R_s < r \text{ and } g(x') \leq v(e_I(x'), \theta_{r_s}) + \varepsilon_{r_s} \\ (j, 1, 0) & \text{if } t = r \text{ and } g(x') \leq v(e_I(x'), \theta_{r_s}) + \varepsilon_{r_s} \\ (j'^s, t+1, 0) & \text{if } t = R_s < r \text{ and } g(x') > v(e_I(x'), \theta_{r_s}) + \varepsilon_{r_s} \\ (j'^\ell, 1, 0) & \text{if } t = r \text{ and } g(x') > v(e_I(x'), \theta_{r_s}) + \varepsilon_{r_s} \end{cases}$$

After $r|S_{r_s}(\theta_{r_s})|$ stages the s-th component j^s of j stabilizes, and therefore if $C = \max_s |S_{r_s}(\theta_{r_s})|$, then after rC stages the entire list j stabilizes. If $i = (i^1, \ldots, i^\ell)$ is the corresponding partition of the play of player 1 in a cycle, the play in the 'stabilized' r-cycle is (i, j) with average payoff

$$g(i, j) \leq \frac{1}{r} \sum_{s=1}^{\ell} r_s (v(e(i^s), \theta_{r_s}) + \varepsilon_{r_s})$$

Fix $\varepsilon > 0$. There exists $r_0 > 0$ such that for every $r \geq r_0$ we have $\varepsilon_r < \varepsilon$. Recall that $v(*, *)$ is uniformly continuous. Therefore there is $\delta > 0$ such that $v(p, \theta - 6\delta \log |I|) \leq v(p, \theta) + \varepsilon$. If we select $2r_0 > r_s \geq r_0$ and set $\theta_{r_s} = \theta - 6\delta \log |I|$ we deduce that

$$\begin{aligned} g(i, j) &\leq \frac{1}{r} \sum_{s=1}^{\ell} r_s (v(e(i^s), \theta) + 2\varepsilon) \\ &\leq v(\theta) + 2\varepsilon \end{aligned}$$

The number of states in M^r is then $\leq r^{|I\times J|+1}2^{r(\theta-6\delta\log|I|}$ and therefore the number of states of the automaton that collates all the the sub-automata M^r and the initial part that 'computes' the cycle length is $\leq 2^{(\theta-\delta/2\log|I|)k}$, which is $\leq n_k$ for sufficiently large k.

It follows that

$$\limsup_{k\to\infty} \text{Val } G^{L_k}(k,n_k) \leq v(\theta)+2\varepsilon \quad \text{as } \frac{\log n_k}{k}\to\theta \text{ and } \frac{L_k}{k\log k}\to\infty$$

As this inequality holds for every $\varepsilon>0$ we conclude that

$$\limsup_{k\to\infty} \text{Val } G^{L_k}(k,n_k) \leq v(\theta) \quad \text{as } \frac{\log n_k}{k}\to\theta \text{ and } \frac{L_k}{k\log k}\to\infty$$

□

5 Extensions

A finite-memory generalization of a finite-state automaton is a finite-state automaton with time-dependent actions and transitions. A time-dependent automaton of player 2 is a quadruple $A=\langle M,m^*,(\alpha_t,)_{t=1}^{\infty},(\beta_t)_{t=1}^{\infty}\rangle$ where m^* is the initial state, $\beta_t: M\times I\to M$ is the transition function at stage t, and $\alpha_t: M\to J$ is the action function at stage t. Let $\mathcal{A}^*(m)$ denote all time-dependent automata with $|M|=m$ states.

A time-dependent automaton $A=\langle M,m^*,(\alpha_t,)_{t=1}^{\infty},(\beta_t)_{t=1}^{\infty}\rangle$ defines a strategy τ^A as follows. Set $m_1=m^*$ and $i_1=\alpha_1(m_1)$, and define inductively $m_{t+1}=\beta_t(m_t,i_t)$ and $\tau_t(i_1,j_1,\ldots,i_{t-1},j_{t-1})=\alpha_t(m_t)$. Note that τ^A is a pure strategy.

The set of all time-dependent automata of size n (as well as the set of all strategies induced by time-dependent automata of size n) is denoted $\mathcal{A}^*(n)$.

Obviously, $\mathcal{A}^*(n)\supset\mathcal{A}(n)$. Therefore, Theorem 1 implies that

$$\limsup_{k\to\infty}\min_{\tau\in\mathcal{A}^*(n_k)}\max_{\sigma\in\mathcal{A}^o(k)}\leq v(\theta)$$

whenever $\lim_{k\to\infty}\frac{\log n_k}{k}=\theta\geq 0$. In addition, the proof of Proposition 1 shows in fact that

$$\max_{\sigma\in\Delta(\mathcal{A}^o(k))}\min_{\tau\in\mathcal{A}^*(n)}E_{\sigma,\tau}\frac{1}{T}\sum_{t=1}^{T}g(i_t,j_t)\geq v(\frac{\log n}{k})$$

Therefore, if $\frac{\log n_k}{k}$ goes to $\theta \geq 0$ as k goes to infinity, then

$$\liminf_{k\to\infty} \max_{\sigma\in\Delta(\mathcal{A}^o(k))} \min_{\tau\in\mathcal{A}^*(n_k)} E_{\sigma,\tau} \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} g(i_t, j_t) \geq v(\theta)$$

and thus we conclude that

$$\lim_{k\to\infty} \max_{\sigma\in\Delta(\mathcal{A}^o(k))} \min_{\tau\in\mathcal{A}^*(n_k)} E_{\sigma,\tau} \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} g(i_t, j_t) = v(\theta)$$

Another finite-memory generalization of a finite-state automaton is a finite-state automaton with time-dependent mixed actions and mixed transitions. A time-dependent mixed actions and mixed transitions automaton of player 2 is a quadruple $A = \langle M, m^*, (\alpha_t,)_{t=1}^{\infty}, (\beta_t)_{t=1}^{\infty}\rangle$, where $\beta_t : M \times I \times J \to \Delta(M)$ and $\alpha_t : M \to \Delta(J)$.

A time-dependent automaton with mixed actions and mixed transitions $A = \langle M, m^*, (\alpha_t,)_{t=1}^{\infty}, (\beta_t)_{t=1}^{\infty}\rangle$ induces a mixed strategy τ^A as follows. Set $m_1 = m^*$. For every strategy σ of player 1 $P_{\sigma,\tau}(i_1 = i) = \alpha_1(m_1)[i]$, and define inductively $P_{\sigma,\tau}(m_{t+1} = m \mid m_1, i_1, j_1, \ldots, m_t, i_t, j_t) = \beta_t(m_t, i_t, j_t)[m]$ and $P_{\sigma,\tau}(i_t = i \mid m_1, i_1, j_1, \ldots, i_{t-1}, j_{t-1}) = \alpha_t(m_t)[i]$. Note that τ^A is a mixed strategy.

The set of all time-dependent automata with mixed actions and mixed transitions of size n (as well as the set of all strategies induced by time-dependent automata with mixed actions and mixed transitions of size n) is denoted $\mathcal{A}^*_{\Delta}(n)$.

After a finite history $(m_1, i_1, j_1, \ldots, m_d, i_d, j_d)$ of the play (and sequence of states of the automata of player 2) the strategy of player 2 in the subgame is one of n mixed strategies, depending on the automata state m_{d+1}. Therefore, Proposition 2 implies that there is a strategy $\sigma \in \Delta(\mathcal{A}^o(k)$ such that for every time-dependent automaton A with mixed actions and mixed transitions with n states and every $d \geq 1$ we have

$$E_{\sigma,\tau^A} \frac{1}{k}\sum_{t=1}^{k} g(i_{d+t}, j_{d+t}) \geq v(\frac{\log n}{k})$$

and for every $r \leq k$

$$E_{\sigma,\tau^A} \frac{1}{r}\sum_{t=1}^{r} g(i_t, j_t) \geq v(\frac{\log n}{k})$$

and therefore

$$\max_{\sigma\in\Delta(\mathcal{A}^o(k))}\min_{\tau\in\mathcal{A}^*_\Delta(n)} g(\sigma,\tau)\geq v(\frac{\log n}{k})$$

Therefore, the above result implies that if $\frac{\log n_k}{k}\to\theta\geq 0$ then

$$\lim_{k\to\infty}\max_{\sigma\in\Delta(\mathcal{A}^o(k))}\min_{\tau\in\mathcal{A}^*_\Delta(n_k)} g(\sigma,\tau)=v(\theta)$$

We conclude that a mixture of oblivious automata of size k can approximately secure the value of the stage game against any time-dependent automaton with mixed actions and mixed transitions and of size subexponential in k.

6 Equilibrium payoffs

Direct and indirect results about the values and equilibrium payoffs of the two-person repeated games $G(k,n)$ and $G^L(k,n)$ appeared in [11, 12, 13, 14, 15, 16, 17, 18]. The results of the present paper enable us to derive exact asymptotic results on the equilibrium payoffs of the repeated games $G[k,n]$ and $G^L[k,n]$, and to extend the asymptotic analysis of the values and equilibrium payoffs of the repeated games $G(k,n)$ and $G^L(k,n)$ in the case where the size of the larger automata is an exponential function of the size of the smaller automata.

Let $G=\langle I,J,g\rangle$, where $g=(g^1,g^2):I\times J\to\mathbb{R}^2$, be a two-person non-zero-sum game. The set of equilibrium payoffs of the infinitely repeated game $G[k,n]$ (respectively, $G(k,n)$) is denoted $E\,G[k,n]$ (respectively $E\,G(k,n)$). Let F denote the convex hull of the set of all vector payoffs $g(i,j)$ with $(i,j)\in I\times J$. For $\theta\geq 0$ define $v^1(\theta)=\max_{p\in\Delta(I)}\min_{Q\in\mathcal{Q}(p.\theta)}E_Q g^1$ and define $v^2(\theta)=\min_{p\in\Delta(I)}\max_{Q\in\mathcal{Q}(p,\theta)}E_Q g^2$.

An important ingredient in determining the equilibrium payoffs of the repeated games $G(k,n)$ and $G[k,n]$ is the characterization of the individual rational payoffs. The individual rational payoff of a player equals the value of a corresponding two-person zero-sum repeated game. Proposition 1 implies that the individual rational payoff of player i in the repeated game $G[k,n_k]$ converges to $v^i(\theta)$ as $k\to\infty$ and $\frac{\log n_k}{k}\to\theta$. Therefore, using the classical arguments used in the proof of the folk theorem, we have

$$\lim_{k\to\infty}E\,G[k,n_k]=\{x\in F:x_i\geq v^i(\theta)\}\text{ as }\frac{\log n_k}{k}\to\theta$$

By part (5) of Theorem 1 we deduce that the individual rational payoff of player i in the repeated game $G^{L_k}[k, n_k]$ converges to $v^i(\theta)$ as $k \to \infty$, $\frac{L_k}{k \log k} \to \infty$, and $\frac{\log n_k}{k} \to \theta$. Therefore,

$$\limsup_{k\to\infty} E\, G^{L_k}[k, n_k] \subset \{x \in F : x_i \geq v^i(\theta)\} \text{ as } (\frac{\log n_k}{k}, \frac{L_k}{k \log k}) \to (\theta, \infty)$$

The inclusion

$$\liminf_{k\to\infty} E\, G^{L_k}[k, n_k] \supset \{x \in F : x_i > v^i(\theta)\} \text{ as } (\frac{\log n_k}{k}, \frac{L_k}{k \log k}) \to (\theta, \infty)$$

follows, e.g., from simple arguments in [13].

The asymptotic behavior of the individual rational levels in the repeated games $G(k, n_k)$ as $k \to \infty$ and $\frac{\log n_k}{k} \to \theta > 0$ is still unknown. [12] raises the question whether or not the limit of the values of $G(k, n_k)$ (where G is a zero–sum game) exists as $k \to \infty$ and $\frac{\log n_k}{k} \to \theta > 0$, and seeks the characterization of the limit when it does exist.[6] As $\mathcal{A}^o(k) \subset \mathcal{A}(k)$, Proposition 1 provides a partial answer to this question by implying a lower bound:

$$\liminf_{k\to\infty} \text{Val } G(k, n_k) \geq v(\theta) \quad \text{as } \frac{\log n_k}{k} \to \theta \tag{13}$$

In addition, the inclusion $\mathcal{A}^o(k) \subset \mathcal{A}(k)$ implies that the individual rational level of player 1 (respectively, player 2) in $G(k, n)$ is at least (respectively, at most) his individual rational level in the game $G[k, n]$, which by Proposition 1 is at least $v^1(\frac{\log n}{k})$ (respectively, at most $v^2(\frac{\log n}{k} + o(1))$). Therefore (13) implies that when $n_k \geq k \to \infty$ and $\frac{\log n_k}{k} \to \theta$ we have

$$\liminf_{k\to\infty} EG(k, n_k) \supset \{x = (x_1, x_2) \in F : x_1 \geq v^1(0) \text{ and } x_2 \geq v^2(\theta)\}$$

and

$$\limsup_{k\to\infty} EG(k, n_k) \subset \{x = (x_1, x_2) \in F : x_1 \geq v^1(\theta) \text{ and } x_2 \geq v^2(0)\}$$

[6] [19] contains exact results on the corresponding questions in the repeated game model with bounded recall.

References

[1] Ben-Porath, E. (1993). Repeated games with finite automata, *Journal of Economic Theory*, **59**, 17–32. Based on Ben-Porath's Master's thesis (1986).

[2] Chandrasekaran, B. (1970). Finite-memory hypothesis testing: A critique, *IEEE Transactions Information Theory*, **IT-16**, 494–496.

[3] Chandrasekaran, B. (1971). Reply to 'Finite-memory hypothesis testing: Comments on a Critique', *IEEE Transactions Information Theory*, **IT-17**, 104–105.

[4] Cover, T. E. (1969). Hypothesis testing with finite statistics. *Ann. Math. Stat.*, **40(3)**, 828–835.

[5] Cover, T. M., and M. E. Hellman (1970). Finite-memory hypothesis testing: Comments on a critique, *IEEE Transactions Information Theory*, **IT-16**, 496–497.

[6] Cover, T. M., and J. A. Thomas (1991). *Elements of Information Theory.* John Wiley & Sons, Inc., New York.

[7] Gossner, O., P. Hernandez, and A. Neyman (2006). Optimal use of communication resources, *Econometrica* **76**, 1603–1636.

[8] Hellman, M. E. (1972). The effects of randomization on finite-memory decision schemes, IEEE *Tansactions*, **IT-18**, 499–502.

[9] Hellman, M. E., and T. M. Cover (1970). Learning with finite memory, *Ann. Math. Stat.*, **41** 765–782.

[10] Hellman, M. E., and T. M. Cover (1971). On memory saved by randomization, *Ann. Math. Stat.*, **42**, 1075–1078.

[11] Neyman, A. (1985). Bounded complexity justifies cooperation in the finitely repeated prisoner's dilemma, *Economics Letters*, **19**, 227–229.

[12] Neyman, A. (1997). Cooperation, repetition, and automata, S. Hart, A. Mas Colell, eds., *Cooperation: Game-Theoretic Approaches*, NATO ASI Series F, 155. Springer-Verlag, 233–255.

[13] Neyman, A. (1998). Finitely repeated games with finite automata, *Mathematics of Operations Research*, **23**, 513–552.

[14] Neyman, A. and D. Okada, (1999). Strategic entropy and complexity in repeated games, *Games and Economic Behavior*, Special Issue in Honor of David Blackwell, **29**, 191–223.

[15] Neyman, A. and D. Okada, (2000a). Repeated games with bounded entropy, *Games and Economic Behavior*, **30**, 228–247.

[16] Neyman, A. and D. Okada, (2000b). Two-person repeated games with finite automata, *International Journal of Game Theory*, **29**, 309–325.

[17] Neyman, A. and D. Okada (2005). Growth of strategy sets, entropy, and nonstationary bounded recall, DP , Center for Rationality, Hebrew University.

[18] Neyman, A. and J. Spencer (2006). Complexity and effective prediction, DP 435, Center for Rationality, Hebrew University.

[19] Peretz, R. (2007). The strategic value of recall, DP 470, Center for Rationality, Hebrew University.