Games and Economic Behavior 74 (2012) 332–351

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

The strategic value of recall

Ron Peretz [1]

Tel-Aviv University, Israel

ARTICLE INFO

Article history:
Received 15 January 2008
Available online 12 June 2011

JEL classification:
C72
C73

Keywords:
Bounded recall
Bounded memory
Bounded rationality
Repeated games
Entropy
de Bruijn sequences

ABSTRACT

This work studies the value of two-person zero-sum repeated games in which at least one of the players is restricted to (mixtures of) bounded recall strategies. A (pure) k-recall strategy is a strategy that relies only on the last k periods of history. This work improves previous results (Lehrer, 1988; Neyman and Okada, 2009) on repeated games with bounded recall. We provide an explicit formula for the asymptotic value of the repeated game as a function of the one-stage game, the duration of the repeated game, and the recall of the agents.

1. Introduction

Bounded recall is one of the alternatives proposed by Aumann (1981) to model bounded rationality in repeated games.[2] A fundamental problem in the study of repeated games with bounded recall is finding a characterization of the (asymptotic) set of equilibrium payoffs. In the classical theory of repeated games (with perfect recall) the folk theorem characterizes the set of equilibrium payoffs. Although Aumann's motivation was probably different, his paper was followed by a number of works devoted to finding analogies of the folk theorem in bounded recall models (and bounded complexity in general).

Lehrer (1988) studied infinitely repeated games where each player i is restricted to (mixtures of) deterministic k_i-recall strategies. He investigated the asymptotic structure of the set of equilibrium payoffs, as the k_i's tend to infinity. Lehrer noticed that the problem reduces to finding the individually rational level (IRL) of each player. In the two-player case the IRL of a player is the value of a zero-sum game formed by the payoff function of that player; hence the importance of the zero-sum case.

Lehrer's main result refers to a finite two-player zero-sum game in strategic normal form $G = \langle I, J, g\rangle$ that is repeated infinitely and each player i is restricted to deterministic k_i-recall strategies. We denote this game $G^\infty[k_1, k_2]$. In general, we denote by $G^T[k, m]$ the T-fold repeated game with recall capacities k and m. The game $G^\infty[k_1, k_2]$ is by itself a finite zero-sum game (since the set of deterministic k-recall strategies is finite), so it admits a value. Lehrer shows that if each k_i ($i = 1, 2$) is a sub-exponential function of the other, then the value of $G^\infty[k_1, k_2]$ converges to the value of G, as k_1, k_2 tend to infinity. Namely, if $\lim \frac{\log k_1}{k_2} = \lim \frac{\log k_2}{k_1} = 0$, then $\lim \operatorname{val} G^\infty[k_1, k_2] = \operatorname{val} G$.

E-mail address: ronprtz@gmail.com.

[1] Research done during Ph.D. studies in the Center for the Study of Rationality, Hebrew University of Jerusalem and in Tel-Aviv University. Research supported in part by Israel Science Foundation grants 263/03 and 212/09 and by the Google Inter-university center for Electronic Markets and Auctions.

[2] In Aumann's model each player observes the actions of the other players but not her own. In this paper each player observes the actions of any other player including her own. Note that the two models are not equivalent (in the presence of bounded recall).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2011.05.013

Lehrer also shows that his main result is tight in the sense that there exists a number L (that depends on $|I \times J|$, the number of joint actions in G) such that if $k_2 \geqslant \exp(Lk_1)$, then player two can beat player one. That is, the value of $G^\infty[k_1, k_2]$ converges to the lower value of G, $V_*(G) := \max_{i\in I} \min_{j\in J} g(i, j)$.

Neyman and Okada (2000, 2009) introduce various classes of bounded complexity strategies that contain the class of bounded recall strategies. They study settings in which one player is bounded while the other is fully rational. In their papers the game is repeated infinitely and the complexity of the bounded player grows as a function of time at a certain rate. In Neyman and Okada (2009) they examine specifically the case of bounded recall. Their bounded recall model can be simplified (without losing its mathematical essence) to a family of (T-fold) finitely repeated games where player one is restricted to k-recall strategies ($k = k(T)$), and player two is fully rational. We denote this game $G^T[k, T]$. Neyman and Okada show that there are numbers $0 < N < O < L$ (that depend on the stage game G) such that the value of $G^{\exp(Nk)}[k, \exp(Nk)]$ converges to the value of G and the value of $G^{\exp(Ok)}[k, \exp(Ok)]$ converges to $V_*(G)$.

Despite the fact that they refer to slightly different models, Neyman–Okada's result can be viewed as an improvement of Lehrer's constants. The current work provides an explicit formula[3] for the asymptotic value of $G^T[k, m]$ for any (large) k, m and T. In so doing, it "closes the gap" between the constants of Neyman–Okada, and shows that Neyman–Okada's constants were not tight. It also closes the gap between the model of Lehrer and the model of Neyman–Okada by showing that for large $k \leqslant m$ the value of $G^T[k, m]$ remains approximately the same for any $T \geqslant m$. The current work employs information-theoretic techniques, including an extremely potent criterion from Neyman and Okada (2009). It also relies on a theorem by de Bruijn (de Bruijn, 1946; Van Aardenne-Ehrenfest and de Bruijn, 1951) that enumerates the de Bruijn sequences.[4]

Section 2 contains essential definitions and the statement of the results. Section 3 provides a few examples to illustrate the results. Section 4 contains definitions and lemmas used in the proofs in Section 5. Section 6 contains extensions beyond the scope of this work. Finally, Appendices A and B contain additional technical details on the proofs of lemmas from Section 4.

2. Results

Let $G = \langle I, J, g\rangle$ be a two-person zero-sum game; where I (resp. J) is player one's (resp. two's) finite set of pure actions, and $g : I \times J \to \mathbb{R}$ is the payoff that player two pays player one. The bi-linear extension of g, defined on pairs of mixed actions $(\sigma, \tau) \in \Delta(I) \times \Delta(J)$, is given by

$$G(\sigma, \tau) := \sum_{i\in I} \sum_{j\in J} \sigma(i)\tau(j)g(i, j).$$

A pure strategy of player one (resp. two) in the repeated version of G is a function from $\bigcup_{n=0}^{\infty}(I \times J)^n$ to I (resp. J). A pair of (pure) strategies σ, τ of player one and two induces a play $a_1, a_2, \ldots \in (I \times J)^{\mathbb{N}}$, which is defined recursively by

$$a_1 = \big(\sigma(\emptyset), \tau(\emptyset)\big),$$
$$a_{n+1} = \big(\sigma(a_1, \ldots, a_n), \tau(a_1, \ldots, a_n)\big).$$

Definition. A k-recall strategy is a strategy that relies on the last k elements of a history. That is, σ is k-recall if and only if for every $n > k$ and every $h \in (I \times J)^n$

$$\sigma(h) = \sigma(h_{n-k+1}, \ldots, h_n).$$

Note that the set of k-recall strategies is finite.

Definition. The game $G^T[k, m]$ is the two-person zero-sum game in which

- the set of strategies for player one is the set of k-recall strategies of player one in the repeated version of G;
- the set of strategies for player two is the set of m-recall strategies of player two in the repeated version of G;
- the payoff function is a function of the induced play, given by

$$\begin{cases} \dfrac{1}{T}\displaystyle\sum_{n=1}^{T} g(a_n) & \text{if } T < \infty, \quad \text{or} \\ \displaystyle\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} g(a_n) & \text{if } T = \infty. \end{cases}$$

Note that the above limit exists since the play in this setting enters a loop at some stage.

[3] Except one singularity point that cannot be resolved in the exponential scale. See Section 6.1.

[4] We are aware of an elementary construction that does not rely on de Bruijn's theorem (see Peretz, 2010), yet we find it hard to resist mentioning de Bruijn's beautiful theorem.

Definition. An *oblivious* strategy in a repeated game is a strategy that ignores the history of actions of the other players. The game $G^T[k_{obl}, m]$ is derived from $G^T[k, m]$ by restricting player one to oblivious strategies only.

The entropy of a mixed action σ of player one, $H(\sigma) = -\sum_{i\in I}\sigma(i)\log(\sigma(i))$, is non-negative, bounded by $\log|I|$, and it equals zero if and only if σ is a pure action. For a real number h, the set of mixed actions of player one whose entropy is either zero or above h is compact; therefore

$$\nu(h) = \max_{\substack{\sigma\in\Delta(I):\\ H(\sigma)\geqslant h,\text{ or}\\ H(\sigma)=0}} \min_{\tau\in J} G(\sigma,\tau)$$

is well defined. The function ν is continuous at every point except perhaps $h = \log(|I|)$.

Let $k, m \in \mathbb{N}$, and let $T \in \mathbb{N}\cup\{\infty\}$. In words, k, m are positive integers and T is either a positive integer or "infinity." For ease of exposition we will assume w.l.o.g. that $k \leqslant m \leqslant T$.

The main result says that the asymptotic value of $G^T[k, m]$ is $\nu(\frac{\log m}{k})$, as long as $\frac{\log m}{k}$ is bounded away[5] from $\log|I|$.

Furthermore, player one can secure the same asymptotic payoff by restricting herself to oblivious strategies; therefore, the asymptotic value of $G^T[k_{obl}, m]$, which is obviously at most $\operatorname{val} G^T[k, m]$, is also $\nu(\frac{\log m}{k})$.

Theorem 2.1. *Let $G = \langle I, J, g\rangle$ be a two-player zero-sum game in strategic normal form. For every $\epsilon, \epsilon_0 > 0$, there is a positive integer k_0 such that for every $m \geqslant k \geqslant k_0$ satisfying $|\frac{\log m}{k} - \log|I|| \geqslant \epsilon_0$ and every $T \geqslant m$ (including $T = \infty$),*

$$\nu\left(\frac{\log m}{k}\right) - \epsilon < \operatorname{val} G^T[k_{obl}, m] \leqslant \operatorname{val} G^T[k, m] < \nu\left(\frac{\log m}{k}\right) + \epsilon. \tag{2.1}$$

In addition, if $\nu(h)$ is continuous at $h = \log|I|$, the theorem holds for $\epsilon_0 = 0$, as well.

An alternative definition for the function $\nu(h)$ uses an auxiliary one-shot game, a restriction of the game G. In the *restricted game of level h* player one is restricted to *mixed* strategies whose entropy is above h, and player two is unrestricted. Since the entropy function is continuous and concave the set $K(h) := \{\sigma \in \Delta(I)\colon\ H(\sigma) \geqslant h\}$ is convex and compact; therefore the restricted game admits a value. The *value of the restricted game* is denoted

$$\tilde{\nu}(h) := \max_{\sigma\in K(h)} \min_{\tau\in J} G(\sigma,\tau) = \min_{\tau\in\Delta(J)} \max_{\sigma\in K(h)} G(\sigma,\tau).$$

The function $\nu(h)$ is the maximum between the value of the restricted game of level h and the pure max–min value of G. With the notation

$$V_*(G) := \max_{j\in J}\min_{i\in I} g(i, j),$$
$$V^*(G) := \min_{j\in J}\max_{i\in I} g(i, j)$$

we have

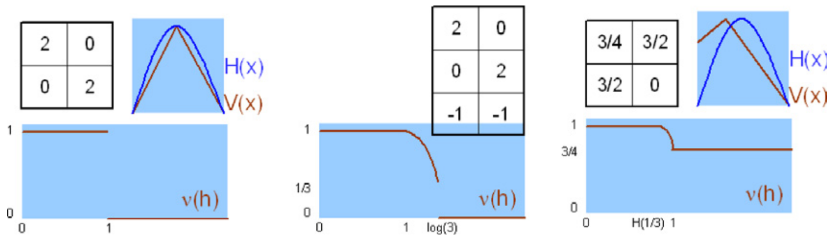
$$\nu(h) = \max\{\tilde{\nu}(h), V_*(G)\}.$$

The left-hand side of inequality (2.1) stems from a characterization of the sequences that can be implemented through k-recall strategies. We show that for every $p \in K(\frac{\log m}{k})$, there exists a random sequence implementable through k-recall strategies which cannot be distinguished from a sequence of i.i.d. random actions with distribution p by any m-recall strategy.

The right-hand side of inequality (2.1) relies on the observation that player two can force player one to obey the rules of the restricted game of level $\frac{\log m}{k}$ or else be punished with a best-response. For every mixed action, q, player two can implement a sequence which cannot be distinguished from a sequence of i.i.d. random actions with distribution q by any k-recall strategy. While doing so, player two can, also, punish player one with a best-response whenever the empirical distribution of the past k actions of player one is not in $K(\frac{\log m}{k})$. By taking q optimal in the restricted game of level $\frac{\log m}{k}$, a payoff of at most $\nu(\frac{\log m}{k})$ is ensured.

In addition to the main result, we study some special cases in which player two can secure her payoff with a pure strategy.

[5] See a discussion on the special point $h = \log|I|$ in Section 6.



(a) (b) (c)

Fig. 1. Examples (left to right): "matching pennies," "matching pennies+," and a game with a continuous ν.

Theorem 2.2. *Let $G = \langle I, J, g\rangle$ be a two-player zero-sum game in strategic normal form. For every $\epsilon, \epsilon_0 > 0$, there is a positive integer k_0 such that for every $m \geqslant k \geqslant k_0$ satisfying $|\frac{\log m}{k} - \log|I|| \geqslant \epsilon_0$ and every $T \geqslant m$ (including $T = \infty$),*

if $|I| = 2$ and $\tilde{\nu}(\frac{\log m}{k}) < \operatorname{val} G$, then

$$\left|V^*\left(G^T[k, m]\right) - \nu\left(\frac{\log m}{k}\right)\right| < \epsilon; \tag{2.2a}$$

if $T > k_0|I|^k$, then

$$\left|V^*\left(G^T[k_{obl}, m]\right) - \nu\left(\frac{\log m}{k}\right)\right| < \epsilon. \tag{2.2b}$$

In addition, if $\nu(h)$ is continuous at $h = \log|I|$, the theorem holds for $\epsilon_0 = 0$, as well.

Note that the condition "$\tilde{\nu}(\frac{\log m}{k}) < \operatorname{val} G$" is equivalent to demanding that every optimal strategy for player one in G has entropy less than $\frac{\log m}{k}$.

3. Examples

Consider the "matching pennies" game described in Fig. 1(a). Since the optimal strategy in this game is $(\frac{1}{2}, \frac{1}{2})$, which is also the one with maximal entropy, Theorem 2.1 says, roughly, that if $\frac{\log m}{k} < \log 2$, then the value of the repeated game is approximately the value of the one-stage game.

In the "matching pennies" game, ν is *not* continuous at $h = \log 2$. It can be shown that $\lim_{k\to\infty} \frac{\log m_k}{k} = h$ implies the convergence of the value of the repeated game $G^T[k, m_k]$ if and only if ν is continuous at h.

Fig. 1(c) describes a game in which ν *is* continuous at the point $h = \log 2$. This is because one of the pure strategies ensures the same payoff as the strategy with maximal entropy $(\frac{1}{2}, \frac{1}{2})$.

Finally, let us look at the "matching pennies+" game in Fig. 1(b). The third alternative of player one is strongly dominated in the one-shot game. Nevertheless, in the repeated game, player one can gain from occasionally playing the third alternative. An intuitive explanation is that by playing the third alternative rather than the first or second, player one can encode information about the history beyond her recall, and it turns out that memory is valuable in repeated games with bounded complexity.

4. Preliminaries

Let A be a finite set. The cardinality of A is denoted $|A|$. The set of probability measures on A equipped with the L_1 norm is denoted $\Delta(A)$. Let $s = (s_1, \dots, s_n)$ be a finite sequence of elements of A. The *empirical distribution* of s is an element of $\Delta(A)$ defined by

$$\operatorname{emp}(s)(a) = \frac{1}{n}\left|\{t\colon s_t = a\}\right|.$$

The empirical distribution of an infinite sequence is defined to be the limit of the empirical distribution of its finite prefixes[6]

$$\operatorname{emp}(s) = \lim_n \operatorname{emp}(s_1, \dots, s_n).$$

[6] The empirical distribution exists if and only if the limit exists.

4.1. Entropy

For completeness we provide a few standard notions in information theory. The reader may refer to Cover and Thomas (2006) for a more elaborated introduction of this topic.

Let I be a finite set and $p = (p_i)_{i \in I} \in \Delta(I)$. The *entropy* of p is defined by

$$H(p) = -\sum_{i \in I} p_i \log(p_i).$$

Sometimes it is convenient to consider random variables rather than probability measures. Let X be a random variable that assumes values in a finite set I. We define the entropy of X to be the entropy of its distribution. That is,

$$H(X) = -\sum_{i \in I} \Pr(X = i) \log\bigl(\Pr(X = i)\bigr).$$

Let X and Y be random variables that assume values in finite sets I and J, respectively. The *entropy of X conditioned on Y* is defined by

$$H(X|Y) := -\sum_{j \in J} \Pr(Y = j) \sum_{i \in I} \Pr(X = i|Y = j) \log\bigl(\Pr(X = i|Y = j)\bigr).$$

In other words: the distribution of X conditional on Y is a $\Delta(I)$-valued random variable, dependent on Y; the entropy of the latter random variable is, also, a random variable whose expectation is the $H(X|Y)$.

The entropy of the $(I \times J)$-valued random variable (X, Y) is denoted $H(X, Y)$. The *chain rule of entropy* is the following identity:

$$H(X|Y) = H(X, Y) - H(Y). \qquad \text{(chain rule)}$$

The *mutual information of X and Y* is defined by

$$I(X; Y) := H(X) - H(X|Y).$$

Applying the chain rule, one can easily verify that

$$I(X; Y) = H(X) + H(Y) - H(X, Y);$$

hence

$$I(X; Y) = I(Y; X).$$

Let $q, p \in \Delta(I)$. The *Kullback–Leibler divergence* of q from p is defined by

$$D(p\|q) = \sum_{i \in I} p_i \log \frac{p_i}{q_i},$$

with the conventions that $0 \log(0/q_i) = 0$, and $D(p\|q) = \infty$ if p is not absolutely continuous with respect to q.

By Jensen's inequality, $D(p\|q)$ is non-negative. Since it is a strictly convex function of p, $D(p\|q) = 0$ if and only if $p = q$.

Remark. As a function of p, $D(p\|q)$ is the difference between the affine function that supports the entropy function at q and the entropy function. To verify this fact note that $H(p) + D(p\|q)$ is linear in p.

Suppose that the distribution of the pair (X, Y) is Q, and the distributions of X and Y are Q_X and Q_Y, respectively. One can verify that

$$I(X; Y) = D(Q_X \otimes Q_Y \| Q), \tag{4.1}$$

which implies that $I(X; Y) \geqslant 0$, and $I(X; Y) = 0$ if and only if X and Y are independent. Equivalently, $H(X) \geqslant H(X|Y)$, and $H(X) = H(X|Y)$ if and only if X and Y are independent.

The Kullback–Leibler divergence is a non-negative lower semi-continuous function defined on a squared compact space ($\Delta(I) \times \Delta(I)$ in the above definition) and it assumes the value zero exactly on the diagonal; therefore it quantifies "closeness" between distributions in the sense that for every $\epsilon > 0$ there exists $\delta > 0$ such that if $D(p_n\|q_n) < \delta$ then $\|p_n - q_n\| < \epsilon$. In particular, (4.1) shows that the mutual information quantifies the dependence (or independence) between a pair of random variables.

The following proposition makes explicit the relation between the Kullback–Leibler divergence and the L_1 distance:

Proposition 4.1 *(Pinsker's inequality).* $\|p - q\|_1 \leqslant \sqrt{2D(p\|q)}$.

4.2. Neyman–Okada's criterion

An *oblivious* strategy in a repeated game is a strategy that ignores the history of actions of the other players. It can, therefore, be viewed as a sequence of actions. The next lemma and the following corollary provide a simple criterion to determine whether a given oblivious strategy secures the payoff that its empirical distribution secures in the one-shot game.

Lemma 4.2. *Let $G = \langle I, J, g\rangle$ be a two-person game. Let T be a positive integer. Let $\bar{x}$ and τ be (correlated) random variables that assume values in*

- *$\bar{x}$ – in I^T;*
- *τ – in the set of strategies of player two in the repeated version of G.*

Let $i_1, j_1, \ldots, i_T, j_T$ be the play induced by $\bar{x}$ and τ. Let x, and y be random variables whose joint distribution is $\mathbf{E}_{\bar{x},\tau}[\mathrm{emp}((i_1, j_1), \ldots, (i_T, j_T))]$. Then

$$I(x; y) \leqslant H(x) - \frac{1}{T}H(\bar{x}|\tau) \leqslant H(x) - \frac{1}{T}H(\bar{x}) + \frac{1}{T}H(\tau).$$

Remark. Lemma 4.2 has been implicitly proven by Neyman (2008, pp. 9, 15–16) while obtaining a lower bound for the value of repeated games with finite automata.

Proof. Given $\bar{x}$ and τ and their induced play $i_1, j_1, \ldots, i_T, j_T$, we denote by (x, y) the $(I \times J)$-valued random variable (i_t, j_t), where t is a random variable that assumes values in $\{1, \ldots, T\}$ uniformly and independently of $\bar{x}$ and τ. Since j_n is a function of $i_1, \ldots, i_{n-1}$ and τ, using the "chain rule" for entropy we have

$$H(x|y) \geqslant H(x|y,t) = \frac{1}{T}\sum_{n=1}^{T} H(i_n|j_n) \geqslant \frac{1}{T}\sum_{n=1}^{T} H(i_n|i_1, \ldots, i_{n-1}, \tau) = \frac{1}{T}H(\bar{x}|\tau) \geqslant \frac{1}{T}H(\bar{x}) - \frac{1}{T}H(\tau). \qquad \square$$

Corollary 4.3 *(Neyman–Okada's criterion). Let $G = \langle I, J, g\rangle$ be a finite two-player zero-sum game. Let $p \in \Delta(I)$. For every $\epsilon > 0$ there exists $\delta > 0$ such that for every positive integer T and all (correlated) random variables $\bar{x}$ and τ that assume values in*

- *$\bar{x}$ – in I^T;*
- *τ – in the set of strategies of player two in the repeated version of G,*

if

1. *$\frac{1}{T}H(\bar{x}) \geqslant H(p) - \delta$,*
2. *$\|\mathbf{E}[\mathrm{emp}(\bar{x})] - p\| < \delta$,*
3. *$\frac{1}{T}H(\tau) < \delta$,*

then

$$\mathbf{E}_{\bar{x},\tau}\left[\frac{1}{T}\sum_{t=1}^{T} g(i_t, j_t)\right] + \epsilon \geqslant G\big(p, \mathbf{E}\big[\mathrm{emp}(j_1, \ldots, j_T)\big]\big) \geqslant \min_{j\in J} G(p, j),$$

where $i_1, j_1, \ldots, i_T, j_T$ is the play induced by $\bar{x}$ and τ.

Remark. Neyman–Okada's criterion appears in Neyman and Okada (2009, pp. 421, 423).

Proof. The right-hand side inequality is obvious. It remains to show the left-hand side inequality. Given $\bar{x}$ and τ and their induced play $i_1, j_1, \ldots, i_T, j_T$ we denote by (x, y) the $(I \times J)$-valued random variable (i_t, j_t), where t is a random variable that assumes values in $\{1, \ldots, T\}$ uniformly and independently of $\bar{x}$ and τ. $\mathbf{E}_{\bar{x},\tau}\frac{1}{T}\sum_{t=1}^{T} g(i_t, j_t) = \mathbf{E}g(x, y)$. We will show that

$$I(x; y) \underset{\delta\to 0}{\longrightarrow} 0. \tag{4.2}$$

Indeed, by Lemma 4.2 and the continuity of the entropy function,

$$\begin{aligned} I(x;y) &\leqslant H(x) - \frac{1}{T}H(\bar{x}) + \frac{1}{T}H(\tau) \\ &\leqslant \big(H(x) - H(p)\big) + \left(H(p) - \frac{1}{T}H(\bar{x})\right) + \frac{1}{T}H(\tau) \\ &\leqslant \big|H(x) - H(p)\big| + \delta + \delta \xrightarrow[\delta\to 0]{} 0. \end{aligned}$$

Let Q be the joint distribution of the pair (x, y), and Q_x, Q_y the distributions of x and y, respectively. By (4.2) and Pinsker's inequality we have

$$\begin{aligned} G\big(p, \mathbf{E}\,\mathrm{emp}(j_1,\dots,j_T)\big) - \mathop{\mathbf{E}}_{\bar{x},\tau} \frac{1}{T}\sum_{t=1}^{T} g(i_t, j_t) &= \mathop{\mathbf{E}}_{(p\otimes Q_y)(i,j)} g(i,j) - \mathop{\mathbf{E}}_{Q\,(i,j)} g(i,j) \\ &\leqslant \max_{i,j}\big|g(i,j)\big|\big[\|p - Q_x\| + \|Q_x \otimes Q_y - Q\,\|\big] \\ &\leqslant \max_{i,j}\big|g(i,j)\big|\big[\delta + \sqrt{2I(x;y)}\big] \xrightarrow[\delta\to 0]{} 0. \qquad \square \end{aligned}$$

Let τ be a (pure or behavior) strategy in a repeated game. Let h be a history of actions in that game. The *induced strategy* $\tau_{|h}$, i.e., "τ given h," is defined by

$$\tau_{|h}(g) = \tau(hg).$$

Note that if τ is k-recall, so is $\tau_{|h}$.

We will apply Neyman–Okada's criterion in two situations. The first situation is when τ is fixed ($H(\tau) = 0$). The second situation is when τ assumes values in $\{\tau^*_{|h}\colon\ h \in \bigcup_{n=1}^{\infty}(I \times J)^n\}$, where τ^* is a fixed m-recall strategy and $m \ll T$. In each of the above situations the third condition of Neyman–Okada's criterion trivially holds. In the first case, $H(\tau) = 0$. In the second case, $H(\tau) \leqslant \log|\{\tau^*_{|h}\colon\ h \in \bigcup_{n=1}^{\infty}(I \times J)^n\}| \leqslant m \log|I \times J| \ll T$.

Definition. We call a sequence of random variables $\bar{x}$ that satisfies the first and second conditions of Corollary 4.3 a "δ-approximation of a sequence of T i.i.d. random variables with common distribution p," or simply an "approximated i.i.d. sequence," if the parameters are understood from the context. We also call the distribution of such random variables a "δ-approximation of p^n."

Example 4.4. Let $x_1, x_2, \dots$ be a sequence of i.i.d. random variables (that assume values in a finite alphabet) with common distribution p. Let A be an event of positive probability. The distribution of $x_1, \dots, x_n$ conditioned on A forms a $\delta(n)$-approximation of p^n, where $\delta(n) \to 0$ as $n \to \infty$. The reader may try to show this with the (weak) law of large numbers. Claim 4.5 bounds $\delta(n)$ by $O(\frac{1}{\sqrt{n}})$. The wider context of conditioning approximated i.i.d. sequences is discussed in Appendix B (see Lemma B.1).

Claim 4.5. *Let $x_1, x_2, \dots$ be a sequence of i.i.d. random variables (that assume values in a finite alphabet) with common distribution p in an arbitrary probability space. For every $\epsilon > 0$ there is a number $C = C(p, \epsilon) > 0$ such that for every*[7] *event A, if $\Pr(A) > \epsilon$ then for every $n \in \mathbb{N}$ the distribution of $x_1, \dots, x_n$ conditioned on A is a $\frac{C}{\sqrt{n}}$-approximation of p^n.*

Proof. See Lemma B.1. Apply it with μ equals the distribution of $x_1, \dots, x_n$, μ_1 equals the distribution of $x_1, \dots, x_n$ conditioned on A, and μ_2 equals the distribution of $x_1, \dots, x_n$ conditioned on A^c. $\square$

4.3. The method of types

Let A be a finite set. We will be interested in the set of sequences whose empirical distribution is a given distribution q.

Definition. Given an alphabet A, a positive integer l, $\epsilon > 0$, and a probability distribution $q \in \Delta(A)$, we define $T^q(l)$ and $T^q_\epsilon(l)$ by

$$\begin{aligned} T^q(l) &= \big\{s \in A^l\colon\ \mathrm{emp}(s) = q\big\}, \\ T^q_\epsilon(l) &= \big\{s \in A^l\colon\ \big\|\mathrm{emp}(s) - q\big\| < \epsilon\big\}. \end{aligned}$$

[7] The event A is not necessarily in the sigma algebra generated by $x_1, x_2, \dots$.

The cardinality of $T^q(l)$ is approximately $\exp(H(q)l)$.

Proposition 4.6. *If $T^q(l) \neq \emptyset$ then*

$$\frac{\exp(H(q)l)}{l^{|A|}} \leqslant \left|T^q(l)\right| \leqslant \exp\bigl(H(q)l\bigr).$$

The probability that a distribution p is obtained as the empirical distribution of l independent q distributed random variables is approximately $\exp(-D(p\|q)l)$.

Proposition 4.7 *(Large deviation). Let $x = x_1, \ldots, x_l$ be a sequence of i.i.d. random variables with common distribution $q \in \Delta(A)$. Let $p \in \Delta(A)$. If $T^p(l) \neq \emptyset$ then*

$$\frac{\exp(-D(p\|q)l)}{l^{|A|}} \leqslant \mathbf{P}\bigl(\mathrm{emp}(x) = p\bigr) \leqslant \exp\bigl(-D(p\|q)l\bigr).$$

Proofs. See Cover and Thomas (2006, pp. 350, 354–355). □

4.4. Oblivious strategies and sequences

It will be convenient to use special notation for sequences of actions that can be produced by oblivious k-recall strategies.

Definition. Let A be finite alphabet and k a positive integer. The set of *k-recall sequences* $B(k, A) \subset A^{\mathbb{Z}}$ is defined by

$$B(k, A) = \bigl\{(a_t) \in A^{\mathbb{Z}}\colon\ \forall t, t'(a_{t-k}, \ldots, a_{t-1}) = (a_{t'-k}, \ldots, a_{t'-1}) \rightarrow a_t = a_{t'}\bigr\}.$$

Since the set A^k is finite, every k-recall sequence must be periodic (note that k-recall sequences are $\mathbb{Z}$-indexed), and the period length must be at most $|A|^k$. A k-recall sequence whose period length is exactly $|A|^k$ is called a "de Bruijn sequence."

Definition. The set of *de Bruijn sequences* of order k over the alphabet A, is the set $DB(k, A)$ defined by

$$DB(k, A) = \bigl\{(a_t) \in B(k, A)\colon\ \forall (b_1, \ldots, b_k) \in A^k\ \exists t \in \mathbb{Z} \text{ s.t. } (a_{t+1}, \ldots, a_{t+k}) = (b_1, \ldots, b_k)\bigr\}.$$

Not only do de Bruijn sequences exist, but there are lots of them. The following result is due to de Bruijn and van Aardenne-Ehrenfest (de Bruijn, 1946; Van Aardenne-Ehrenfest and de Bruijn, 1951):

Theorem 4.8. $|DB(k, A)| = |A|!^{|A|^{k-1}}$.

It will be convenient to refer to finite prefixes of infinite sequences.

Definition. Let $T \in \mathbb{N}$. We define the T-prefix operator and its extension to sets of sequences by

$$\mathrm{pref}_T\, a = (a_1, \ldots, a_T),$$
$$\mathrm{pref}_T\, B = \{\mathrm{pref}_T\, b \mid b \in B\}$$

where $a \in A^{\mathbb{Z}}$ and $B \subset A^{\mathbb{Z}}$.

For every $\epsilon > 0$ consider the set of all k-recall sequences with empirical distribution within a distance of ϵ from p. Let us denote this set $C^p_\epsilon(k)$. Formally, we define

$$C^p_\epsilon(k) := \bigl\{x \in B(k, I)\colon\ \bigl\|\mathrm{emp}(x) - p\bigr\| < \epsilon\bigr\}.$$

Proposition 4.9. *$\forall p \in \Delta(I)\ \forall \epsilon > 0\ \exists k_0 \in \mathbb{N}$ such that $\forall k \geqslant k_0, \forall T \in \mathbb{N}$, if*

$$\frac{\log T}{k} < H(p) - \epsilon$$

then

$$\frac{\log|\mathrm{pref}_T\, C^p_\epsilon(k)|}{T} > H(p) - \epsilon.$$

Proof. The richness of k-recall sequences is treated in Appendix A. See Proposition A.2. □

5. Proofs

In this section we prove Theorems 2.1 and 2.2. Henceforth let $G = \langle I, J, g\rangle$ be a two-person zero-sum game.

Since $\nu(h)$ is non-increasing and continuous at any point $h \neq \log|I|$, we can conclude the proof with the following propositions:

Proposition 5.1. *For every $h \in \mathbb{R}$ and every $\epsilon, \epsilon_0 > 0$, there exists $k_0 \in \mathbb{N}$ such that for every $k_0 \leqslant k \leqslant m \leqslant T$, if*

$$\frac{\log m}{k} < h - \epsilon_0$$

then

$$\operatorname{val} G^T[k_{obl}, m] \geqslant \nu(h) - \epsilon.$$

Proposition 5.2. *For every $h \in \mathbb{R}$ and every $\epsilon, \epsilon_0 > 0$, there exists $k_0 \in \mathbb{N}$ such that for every $k_0 \leqslant k \leqslant m \leqslant T$, if*

$$\frac{\log m}{k} > h + \epsilon_0$$

then

$$\operatorname{val} G^T[k, m] \leqslant \nu(h) + \epsilon;$$

if $|I| = 2$ and $\nu(h) < \operatorname{val} G$, then

$$V^*\big(G^T[k, m]\big) \leqslant \nu(h) + \epsilon;$$

if $T/|I|^k > k_0$, then

$$V^*\big(G^T[k_{obl}, m]\big) \leqslant \nu(h) + \epsilon.$$

Proposition 5.1 says that player one can guarantee a payoff of at least $\nu(\frac{\log m}{k})$. Proposition 5.2 says that player two can keep the payoff at most $\nu(\frac{\log m}{k})$.

Claim. *Propositions* 5.1 *and* 5.2 *imply Theorems* 2.1 *and* 2.2.

For simplicity, we will only prove the main result, Theorem 2.1. Theorem 2.2 follows from replacing "val" by "V^*" in our proof.

Proof. Since the theorem says that player two can bring the payoff close to $V_*(G)$ by using $|I|^{2k}$-recall strategies and player one can always guarantee $V_*(G)$, it is sufficient to consider the case where $\frac{\log m}{k} \leqslant 2\log|I|$.

Let $\epsilon, \epsilon_0 > 0$. Take a finite set $\{h_l\}$ of real numbers such that for every $h \in [0, 2\log|I|]$ there exist some h_l such that $h_l - 2\epsilon_0 < h < h_l - \epsilon_0$. Proposition 5.1 provides for each h_l some $k_0(h_l)$. Let $k_0 = \max\{k_0(h_l)\}$. Let $k_0 \leqslant k \leqslant m \leqslant T$. There exist some h_l such that $h_l - 2\epsilon_0 < \frac{\log m}{k} < h_l - \epsilon_0$. Proposition 5.1 asserts that

$$\operatorname{val} G^T[k, m] \geqslant \nu(h_l) - \epsilon.$$

From the monotonicity of $\nu(h)$ we have the following lower bound:

$$\operatorname{val} G^T[k, m] \geqslant \nu\left(\frac{\log m}{k} + 2\epsilon_0\right) - \epsilon. \tag{5.1}$$

By similar considerations we deduce the following upper bound form Proposition 5.2:

$$\operatorname{val} G^T[k, m] \leqslant \nu\left(\frac{\log m}{k} - 2\epsilon_0\right) + \epsilon. \tag{5.2}$$

Combining (5.2) and (5.1) with the fact that $\nu(h)$ is uniformly continuous in any closed set that excludes the point $h = \log|I|$ concludes Theorem 2.1. □

In the next two sections we will provide proofs of Propositions 5.1 and 5.2.

5.1. $\operatorname{val} G^T[k_{obl}, m]$ is at least $\nu(\frac{\log m}{k})$

The proof of Proposition 5.1 relies on the richness of k-recall sequences among all sequences of length T, as expressed by Proposition 4.9.

Proof of Proposition 5.1. Since $\operatorname{val} G^T[k_{obl}, m] \geqslant V_*(G)$ is trivial, it is sufficient to show that $\operatorname{val} G^T[k_{obl}, m] \geqslant \tilde{\nu}(h) - \epsilon$.

Let $p = p(h) \in \arg\max\{\min_{j\in J} G(p, j)\colon\ H(p) \geqslant h\}$. Let $\delta = \delta(\epsilon)$ be given by Neyman–Okada's criterion (Corollary 4.3). We assume w.l.o.g. that $\delta < \frac{\epsilon_0}{2}$. Consider the set $C^p_\delta(k)$ of all k-recall sequences whose empirical distribution is within a distance of δ from p. Formally,

$$C^p_\delta(k) := \{x \in B(k, I)\colon\ \|\operatorname{emp}(x) - p\| < \delta\}.$$

For every k let the mixed strategy of player one $\sigma = \sigma_k$ be the uniform distribution over the set $C(k) = C^p_\delta(k)$.

Let τ^* be an arbitrary m-recall strategy for player two. Let $a_1, a_2, \ldots$ denote the play induced by σ and τ^*. We analyze the expected payoff by breaking the game's time line into consecutive time intervals of the form $[t_0 + 1, t_0 + T']$, where either $t_0 = 0$ and $T' \leqslant mk$, or $t_0 \geqslant 0$ and $T' = mk$. In each case we use Neyman–Okada's criterion to show that

$$\mathbf{E}\frac{1}{T'}\sum_{t=t_0+1}^{t_0+T'} g(a_t) + \epsilon \geqslant \min_{j\in J}\sum_{i\in I} p_i g(i, j) = \tilde{\nu}(h).$$

We would like to apply Neyman–Okada's criterion with $\bar{x} = a^1_{t_0+1}, \ldots, a^1_{t_0+T'}$ and $\tau = \tau^*_{|a_1,\ldots,a_{t_0}}$. We choose k_0 so large that, in addition to Proposition 4.9, it also satisfies $\frac{\log k}{k} < \frac{\epsilon_0}{2}$ and $\frac{\log|I\times J|}{k} < \delta$, for every $k \geqslant k_0$. Clearly,

$$\frac{\log T'}{k} \leqslant \frac{\log m}{k} + \frac{\log k}{k} \leqslant H(p) - \epsilon_0 + \frac{\epsilon_0}{2} < H(p) - \delta.$$

By Proposition 4.9, and since $C(k)$ is invariant to shifts (hence $a^1_1, a^1_2, \ldots$ is a stationary process), we have

$$\frac{1}{T'}H(a^1_{t_0+1}, \ldots, a^1_{t_0+T'}) = \frac{1}{T'}H(a^1_1, \ldots, a^1_{T'}) = \frac{\log|\operatorname{pref}_{T'} C^p_\delta(k)|}{T'} > H(p) - \delta.$$

We have just verified the first condition of Neyman–Okada's criterion. The second condition stems directly from the definition of $C^p_\delta(k)$. It remains to verify the third condition. If $t_0 = 0$, then $\tau = \tau^*$; therefore $H(\tau) = 0$. If $t_0 > 0$, then $T' = mk$. The random variable τ assumes values in the set $\{\tau^*_{|h}\colon\ h \in \bigcup_{n=1}^\infty (I\times J)^n\}$. The strategy τ^* is an m-recall strategy; therefore $|\{\tau^*_{|h}\colon\ h \in \bigcup_{n=1}^\infty (I\times J)^n\}| \leqslant |I\times J|^m$. Consequently,

$$\frac{1}{T'}H(\tau) \leqslant \frac{m\log|I\times J|}{mk} = \frac{\log|I\times J|}{k} < \delta. \qquad \square$$

5.2. $\operatorname{val} G^T[k, m]$ is at most $\nu(\frac{\log m}{k})$

In this section we prove the main part of Proposition 5.2, which relates to $\operatorname{val} G^T[k, m]$. The other parts, which relate to $V^*(G^T[k, m])$ and $V^*(G^T[k_{obl}m,])$, will be proved in Section 5.3.

Given $h \geqslant 0$ and an integer k we will construct a mixed strategy $\tau = \tau^{k,h}$ for player two in the repeated version of G. We begin with an informal description of this strategy.

Player two chooses an optimal strategy q in the restricted game of level h. He draws $x_1, x_2, \ldots, x_{k^2}$ i.i.d. with common distribution q, and plays this sequence periodically (i.e. $x_t = x_{t+k^2}$). By Neyman–Okada's criterion, $\{x_t\}$ "seems like" an i.i.d. sequence in the eyes of any k-recall player; therefore if the empirical distribution of the actions of player one had an entropy $\geqslant h$, a payoff of at most $\tilde{\nu}(h)$ would be secured. But what if it did not have? While acting according to $x_1, x_2, \ldots$, player two also looks for opportunities to predict player one's next action and best-reply to it. This can happen when the past k actions of history have already appeared sometime before. The strategy of player two is: *try to predict player one's next action and best-reply to it; otherwise, play according to* $x_1, x_2, \ldots$. Let us now analyze the play of the game. There are less than $k^{|I|}\exp(hk)$ k-tuples of actions of player one whose empirical distribution is less than h and there are only k^2 ways in which the sequence $x_1, x_2, \ldots$ can align with it; therefore, there are only $k^2\exp(hk)$ times when player one can "disobey" the rules of the restricted game and "get away with it" without being punished with a best-reply. Although this strategy seems plausible, we were not able to show that it actually "works." The strategy that we actually construct is a slight variation of the described strategy, for which we can bound the expected payoff from above.

In the first step, we prove that τ ensures the desired payoff in the case $T = m$. That is, for large k, if $\frac{\log m}{k} > h - \epsilon_0$, then $\tau^{k,h}$ ensures that the mean payoff in the first m steps against any k-recall strategy is at most $\nu(h) + \epsilon$. In the next step, we argue that, by "restarting" from the beginning after every m periods, player two can ensure the desired payoff for $T \gg m$. The remaining case $T = O(m)$ is reduced to the case $T \gg m$ by letting player two restrict herself to m'-recall strategies, where $m' \ll m$ and $\log m' \sim \log m$.

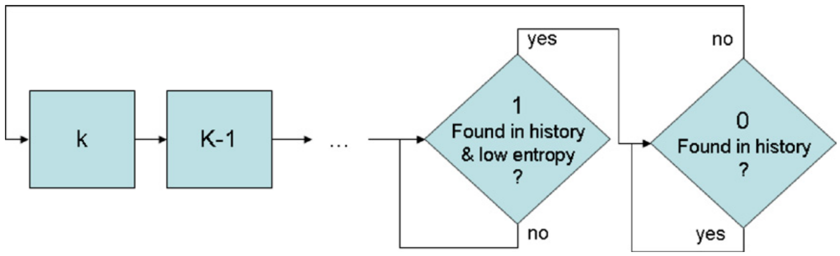


Fig. 2. An illustration of the function s as a flow chart.

Case I: $T = m$

Let k and h be given. The strategy τ will be a mixture of pure strategies of the form τ_x, where $x = x_1, x_2, \dots$ is a k^2-periodic sequence of actions of player two. We will first describe τ_x as a function of x and discuss its properties (for a special class of sequences). Then we will define τ by providing a probability measure over the possible values for x. The last part of the proof is the analysis of the payoff that τ secures.

We begin by defining a mapping s from the set of all finite histories to the set $\{0, \dots, k\}$ (Fig. 2). The set $\{0, \dots, k\}$ can be interpreted in this context as the "states of mind" of the strategy τ. Note that the mapping s does not depend on the sequence x (it does, however, depend on k and h).

$$s(\emptyset) = k$$

$$s(a_1, \dots, a_{t+1}) = \begin{cases} s(a_1, \dots, a_t) - 1 & \text{if } s(a_1, \dots, a_t) > 1; \\ 1 & \text{if } s(a_1, \dots, a_t) = 1, \text{ and} \\ & [\forall s < t\ ((a_{s-k+2}, \dots, a_{s+1}) \neq (a_{t-k+2}, \dots, a_{t+1}), \\ & \text{or } H(\text{emp}(a^1_{t-k+2}, \dots, a^1_{t+1})) \geqslant h)]; \\ 0 & \text{if } [s(a_1, \dots, a_t) = 1, \text{ and} \\ & \exists s < t \text{ s.t. } ((a_{s-k+2}, \dots, a_{s+1}) = (a_{t-k+2}, \dots, a_{t+1}), \\ & \text{and } H(\text{emp}(a^1_{t-k+2}, \dots, a^1_{t+1})) < h)] \\ & \text{or} \\ & [s(a_1, \dots, a_t) = 0, \text{ and} \\ & \exists s < t \text{ s.t. } (a_{s-k+2}, \dots, a_{s+1}) = (a_{t-k+2}, \dots, a_{t+1})]; \\ k & \text{otherwise.} \end{cases}$$

In words, the initial state is k. States 2 to k simply count down to 1. State 1 looks into the history for an occurrence of the past k actions as long as the actions of player one have *low* entropy. If one is found, state 1 goes to state 0; otherwise it stays in state 1. State 0 looks in the history for an occurrence of the past k actions. If one is found, state 0 stays in state 0; otherwise it jumps back to state k.

Given a play $a_1, a_2, \dots$ we denote $s_t = s(a_1, \dots, a_{t-1})$. Note that

- $s_{l+1}, \dots, s_{l+k} \neq 0$, iff $s_{l+k} = 1$;
- if player one is limited to k-recall, then whenever $s_t = 0$ player two can predict the next action of player one, namely, a^1_t.

Now we are ready to define τ_x. For every finite history $\mathfrak{h} = a_1, \dots, a_t$ ($t = 0$ indicates the empty history) define[8]

$$\tau_x(\mathfrak{h}) = \begin{cases} \gamma((a^1_{t+1})) & \text{if } s(\mathfrak{h}) = 0, \\ x_{t+1} & \text{otherwise,} \end{cases} \tag{5.3}$$

where $\gamma : I \to J$ is such that $\forall i \in I\ g(i, \gamma(i)) \leqslant V_*(G)$.

We now define τ by specifying a random choice of x. Let q be an optimal strategy for player two in the restricted game of level h. That is, $q \in \Delta(J)$ and for every $p \in K(h)$ $\sum_{i \in I, j \in J} p(i)q(j)g(i, j) \leqslant \tilde{\nu}(h)$. Let $x = x_1, \dots, x_{k^2}$ be i.i.d. random variables with common distribution q. Let $(x_t)_{t \in \mathbb{Z}}$ be the k^2-periodic extension of $x_1, \dots, x_{k^2}$.

For every realization of τ, τ_x, and every k-recall strategy for player one σ, consider the induced play $a_1, a_2, \dots$ and the sets

[8] If $s(\mathfrak{h}) = 0$, player two can predict the next action of player one, namely, a^1_{t+1}.

$$L^{(i)} = \{l \mid 0 \leqslant l \leqslant T,\ H(\mathrm{emp}(a^1_{l+1}, \dots, a^1_{l+k})) < h,\ s_{l+k} = 1,\ s_{l+k+1} = \dots = s_{l+k+i} = 0,\ s_{l+k+i+1} \neq 0\},$$

$$L = \bigcup_{i=0}^{\infty} L^{(i)},$$

$$A = \{(a_{l+1}, \dots, a_{l+k}) \mid l \in L\}.$$

From the definition of s we have for every $i > 0$,

$$L^{(i)} = \{l \mid s_{l+k} \neq 0,\ s_{l+k+1} = \dots = s_{l+k+i} = 0,\ s_{l+k+i+1} \neq 0\}. \tag{5.4}$$

Analogously we define

$$N^{(i)} = \{l + k + i + 1 \mid l \in L^{(i)}\} = \{n \mid s_{n-i-1} \neq 0,\ s_{n-i} = \dots = s_{n-1} = 0,\ s_n \neq 0\} \quad \text{for } i > 0,$$

and

$$N = \bigcup_{i=0}^{\infty} N^{(i)}.$$

For every $\bar{a} \in A$, let $L_{\bar{a}} = \{l \in L \mid (a_{l+1}, \dots, a_{l+k}) = \bar{a}\}$. Arrange the elements of $L_{\bar{a}}$ in increasing order $L_{\bar{a}} = \{l_1 < l_2 < \cdots\}$. At time l_i, $\bar{a}$ is followed by a sequence of $m(i)$ stages in which $s_{l_i+k+1} = \dots = s_{l_i+k+m(i)} = 0$ (the predictive state), $m(i) < m(i+1)$; therefore the map $l \mapsto (a_{l+1}, \dots, a_{l+k})$ is injective when restricted to each $L^{(i)}$. Hence

$$|L^{(i)}| = |N^{(i)}| \leqslant |A| \leqslant k^{2+|I|} \exp(hk).$$

The latter inequality follows by bounding by $k^{|I|} \exp(hk)$ the number of sequences of length k of actions of player one whose empirical distribution has entropy less than h, and by k^2 the number of possible reactions of player two. Hence $|\bigcup_{i<k^3} L^{(i)}| \leqslant k^{5+|I|} \exp(hk)$. By (5.4), $|\bigcup_{i \geqslant k^3} L^{(i)} \cap \{1, \dots, m\}| \leqslant m/k^3$. Consequently, for large k,

$$|(L \cup N) \cap \{1, \dots, m\}| < m/k^{2.99}.$$

Hence, for almost every $t \in \{0, \dots, m\}$, none of the numbers $t-k, t-k+1, \dots, t+k^2$ is in $L \cup N$; therefore, for almost every t, either

(i) $s_{t+1} = \dots = s_{t+k^2} = 0$, or
(ii) $s_{t+1}, \dots, s_{t+k^2} \neq 0$, and $\forall 0 \leqslant l < k^2\ H(\mathrm{emp}(a_{t+l+1}, \dots, a_{t+l+k})) \geqslant h$.

Claim. *There exists a k_0 such that, if $k_0 \leqslant k \leqslant m$, and $\frac{\log m}{k} > \epsilon_0$, then $\tau^{k,h}$ ensures an expected mean payoff of at most $\nu(h) + \epsilon$ in the first m steps of the game, against any k-recall strategy (of player one). Namely,*

$$\mathbf{E}\left[\frac{1}{m}\sum_{t=1}^{m} g(a_t)\right] < \nu(h) + \epsilon.$$

Proof. Fix $k \in \mathbb{N}$, and an arbitrary k-recall strategy for player one σ. The strategies σ and $\tau^{k,h}$ induce a play $a = a_1, a_2, \dots$ and a sequence of states $s_t = s(a_1, \dots, a_{t-1})$. Clearly, a and s are random variables (functions of $x = x_1, \dots, x_{k^2}$). Let t be another random variable independent of x, assuming values in $\{0, \dots, m\}$ with uniform distribution. Define a real-valued random variable W by

$$W = \frac{1}{k^2}\sum_{i=1}^{k^2} g(a_{t+i}).$$

The expectation of W approximates the actual expected payoff, namely, $|\mathbf{E}W - \mathbf{E}[g(a_t)|t \neq 0]| = O(\frac{k^2}{m})$; therefore the objective is to bound $\mathbf{E}W$ from above.

In case (i) W is at most $V_*(G)$. If the probability of case (ii) is greater than δ, then the actions of player two $x_{t+1}, \dots, x_{t+k^2}$ conditioned on (ii) form a $C(q,\delta)k^{-1}$-approximated i.i.d. sequence (by Claim 4.5) and therefore secure a payoff of $\tilde{\nu}(h) + \epsilon$ (by Neyman–Okada's criterion, taking $\bar{x} = (x_{t+1}, \dots, x_{t+k^2})$ and $\tau = \sigma_{|a_1,\dots,a_t}$). If the probability of case (ii) is less than δ, then for large k the probability of case (i) is greater than $1 - 2\delta$; hence[9] $\mathbf{E}W \leqslant V_*(G) + 2\delta\|G\|$.

In the case of $h > \log I$, the condition in (ii) cannot be satisfied and hence (i) occurs with probability close to one and $V_*(G) + \epsilon$ is secured. □

[9] Where $\|G\| = \max_{i,j} |g(i,j)|$.

Case II: $T \geqslant km$

We now turn to the case $T \gg m$. Concretely, let us assume that $T \geqslant km$. We would like to define a strategy that plays according to τ and "restarts" after every m steps:

$$\tau'(a_1, \dots, a_t) = \tau(a_{ml+1}, \dots, a_t),$$

where l is the integer such that $0 \leqslant t - ml < m$. Although τ' is not an m-recall strategy, since it depends also on t (modulo m), there are various alternative constructions that simulate τ' with an m-recall strategy. We present one alternative.

With our m-recall strategy τ'', ml is replaced by a sequence of stopping times $0 = T_0 < T_1 < \cdots$ denoted by $\mathcal{T} = \{T_i\}$. Our strategy τ'' and $\mathcal{T}$ satisfy the following properties:

(i) τ'' is an m-recall strategy;
(ii) $T_{i+1} - T_i \geqslant m - 3k$;
(iii) $a^2_{t+1} = \tau(a_{T_i+1}, \dots, a_t)$, if $T_i \leqslant t < T_i + m - 3k$;
(iv) $a^2_t = \gamma(a^1_t)$, if $T_i + m - 2k < t < T_{i+1} - k$,

where γ is the best-reply function mentioned in (5.3). Properties (ii)–(iv) and the assumption that $T \geqslant mk$, clearly, guarantee an average expected payoff of at most $\nu(h) + \epsilon + O(\frac{1}{k})$.

We will now construct τ'' and $\mathcal{T}$, and then claim their properties. Given the sequence $(x_t)_{t\in\mathbb{Z}}$, let $(y_t)_{t\in\mathbb{Z}}$ be a k-periodic sequence of actions of player two, such that $\forall t\ \forall s\ (x_{t+1}, \dots, x_{t+k}) \neq (y_{s+1}, \dots, y_{s+k})$.[10]

We define a set of stopping times $\mathcal{T}$, as a function of the play $a_1, a_2, \dots$, by

$$\tilde{\mathcal{T}} = \big\{t > 2k\colon \exists s\big(a^2_{t-2k}, \dots, a^2_t\big) = (y_{s-2k}, \dots, y_s),\ a^2_{t-k} \neq \gamma\big(a^1_{t-k}\big)\big\},$$
$$\mathcal{T} = \{t \in \tilde{\mathcal{T}}\colon t - l \notin \tilde{\mathcal{T}},\ l = 1, 2, \dots, k\}.$$

Let T_i to be the i-th element of $\mathcal{T}$. Namely,

$$T_0 = 0,$$
$$T_{i+1} = \min \mathcal{T} \setminus \{T_0, \dots, T_i\}.$$

At time t, let $n(t) = \max\{T_i\} \cap \{1, \dots, t\}$. We define τ'' by

$$\tau''(a_1, \dots, a_t) = \begin{cases} \tau(a_{n(t)+1}, \dots, a_t) & \text{if } t - n(t) < m - 3k, \\ \phi(a^2_{t-k+1}, \dots, a^2_t) & \text{otherwise.} \end{cases}$$

Let ϕ be the following oblivious k-recall strategy. Given $j_1, \dots, j_k \in J$, let the pair (i, s) of positive integers be minimal (in lexicographic order) with respect to the property

$$(j_i, \dots, j_k) = (y_{s+i}, \dots, y_{s+k}).$$

Defined $\phi(j_1, \dots, j_k)$ to be y_{s+k+1}. Note that the sequence that ϕ generates is equivalent to (y_t) up to a time shift.

Claim (i). *τ'' is an m-recall strategy.*

Proof. The truth of the proposition "$t \in \tilde{\mathcal{T}}$" depends only on $a_{t-2k}, \dots, a_t$; therefore "$t \in \mathcal{T}$" depends only on $a_{t-3k}, \dots, a_t$; hence "$\exists l$ s.t. $0 \leqslant l < m - 3k$ and $t - l \in \mathcal{T}$" depends only on $a_{t-m+1}, \dots, a_t$; therefore the condition "$t - n(t) < m - 3k$" can be verified with a recall of m, and if it holds then $t - n(t)$ can be observed. □

Claim (ii, iii). *$T_{i+1} - T_i \geqslant m - 3k$.*

Proof. Property (iii) stems directly from the definition of τ''. The definition of $\mathcal{T}$ implies that $T_{i+1} - T_i > k$. Assume, by way of contradiction, that $k < T_{i+1} - T_i < m - 3k$, for some i. By (iii), $a_{T_i+1}, \dots, a_{T_{i+1}}$ is the play induced by τ and $\sigma_{|a_1,\dots,a_{T_i}}$. Since $T_{i+1} \in \tilde{\mathcal{T}}$, $a^2_{T_{i+1}-k} \neq \gamma(a^2_{T_{i+1}-k})$. The definition of τ asserts that there must be some k consecutive steps that include $T_{i+1} - k$ in which τ has played according to (x_t). Namely, $\exists s$ such that $0 < T_{i+1} - T_i - k - s \leqslant k$ and $(a^2_{s+1}, a^2_{s+k}) = (x^2_{s+1}, x^2_{s+k})$. By the choice of (y_t), $(a^2_{s+1}, a^2_{s+k}) \neq (y^2_{t+1}, y^2_{t+k})$, for every $t \in \mathbb{Z}$; hence $T_{i+1} \notin \tilde{\mathcal{T}} \supset \mathcal{T}$. □

Claim (iv). *If $T_i + m - 2k < t < T_{i+1} - k$, then $a^2_t = \gamma(a^1_t)$.*

[10] Such a sequence exists as long as $|J|^k > k^3$.

Proof. Let t and i be positive integers satisfying $T_i+m-2k<t<T_{i+1}-k$. Assume, by way of contradiction, that $a_t^2 \neq \gamma(a_t^1)$. The proof of (ii) actually shows that $\{t \in \tilde{\mathcal{T}}:\ k<t-T_i<m-3k\}=\emptyset$; therefore $T_{i+1}=\min\{t \in \tilde{\mathcal{T}}:\ t>T_i+k\}$. Since $t+k<T_{i+1}$, $t+k \notin \tilde{\mathcal{T}}$. By the definition of τ'', $a^2_{T_i+m-3k+1},\dots,a^2_{T_{i+1}}$ is the play induced by $\phi_{|a^2_{T_i+m-2k+1},a^2_{T_i+m-3k}}$, which is equal to $y_{s+T_i+m-3k+1},\dots,y_{s+T_{i+1}}$, for some $s \in \mathbb{Z}$, so, by definition, $t+k \in \tilde{\mathcal{T}}$. $\square$

Case III: $m<T<km$

Finally, let us consider the remaining case $m<T<km$. Let $m'=[T/k]$. On the one hand $m'<m$; therefore $\operatorname{val} G^T[m',k] \geqslant \operatorname{val} G^T[m,k]$. On the other hand, $2m'>m/k$; therefore $\frac{\log m'}{k}>\frac{\log m}{k}-\frac{\log(2k)}{k}>h+\epsilon_0-\frac{\log(2k)}{k}$. For large k we have $\frac{\log m'}{k}>h+\frac{\epsilon_0}{2}$. $T \geqslant km'$, and in this case we have already proven, that there exists k_0 such that for $k \geqslant k_0$,

$$\nu(h)+\epsilon>\operatorname{val} G^T\big[m',k\big] \geqslant \operatorname{val} G^T[m,k].$$

5.3. The pure min–max value

In Section 5.2 player two secured a payoff of $\nu(h)+\epsilon$ by playing a *mixed* strategy. In this section we will see special cases where the same payoff can be secured by playing a *pure* m-recall strategy.

5.3.1. Special case I: $|I|=2$ and $\tilde{\nu}(h)<\operatorname{val} G$

Proposition 5.3. *If $|I|=2$, then for every $h \in \mathbb{R}$ such that $\tilde{\nu}(h)<\operatorname{val} G$ and every $\epsilon,\epsilon_0>0$, there exists $k_0 \in \mathbb{N}$ such that for every $k_0 \leqslant k \leqslant m \leqslant T$, if*

$$\frac{\log m}{k}>h+\epsilon_0,$$

then

$$V^*\big(G^T[k,m]\big) \leqslant \tilde{\nu}(h)+\epsilon.$$

Proof. Consider the strategy τ that was defined in the proof of Proposition 5.1. Recall that τ is a mixture of strategies of the form τ_x, where x is a sequence of i.i.d. random variables with distribution q and q is an optimal strategy for player two in the restricted game of level h. It is, therefore, sufficient to show that player two has a pure optimal strategy in the restricted game of level h.

We can represent a mixed action in $\Delta(I)$ by a real number $p \in [0,1]$. Denote the expected payoff that player one guarantees by playing p by $V(p)=\min_{j \in J} G(p,j)$. The function $V(p)$ is concave, and its maximum is $\operatorname{val} G$. Since $\tilde{\nu}(h)<\operatorname{val} G$, the maximum of $V(p)$ is not obtained in $K(h)$; hence $V(p)$ must be monotonic in $K(h)$. Let $p_0=\arg\max_{p \in K(h)} V(p)$. The function $V(p)$ is piecewise linear, and every linear segment corresponds to an action of player two. Let $j_0 \in J$ be an action that corresponds to a linear segment that includes p_0, so that the function $L(p)=G(p,j_0)$ coincides with $V(p)$ over some interval in $K(h)$. Clearly, $L(p)$ is increasing (resp. decreasing) iff $V(p)$ is increasing (resp. decreasing) in $K(h)$. Consequently,

$$\forall p \in K(h)\ G(p,j_0) \leqslant G(p_0,j_0)=\tilde{\nu}(h).$$

In other words, j_0 is a pure optimal strategy in the restricted game of level h. $\square$

5.3.2. Special case II: Player one is oblivious and $T \gg |I|^k$

Proposition 5.4. *For every $h \in \mathbb{R}_+$ and every $\epsilon,\epsilon_0>0$, there exists $k_0 \in \mathbb{N}$ such that for every $k_0 \leqslant k \leqslant m \leqslant T$, if*

$$\frac{\log m}{k}>h+\epsilon_0,$$
$$T>k_0|I|^k,$$

then

$$V^*\big(G^T[k_{obl},m]\big) \leqslant \nu(h)+\epsilon.$$

The next claim says that if the recall of player two is much greater than k, then player two can implement a pure oblivious strategy that does not correlate with any oblivious k-recall strategy of player one.

Claim 5.5. *$\forall q \in \Delta(J)$ $\forall \epsilon > 0$ $\exists k_0 \in \mathbb{N}$ $\forall k \geqslant k_0$ $\forall T \geqslant k_0|I|^k$ there exists a sequence $y \in B(k_0 \cdot k, J)$ such that for every $x \in B(k, I)$,*

$$\frac{1}{T}\sum_{t=1}^{T} g(x_t, y_y) \leqslant G\big(q, \operatorname{emp}(x_1, \dots, x_T)\big) + \epsilon.$$

Claim 5.5 implies Proposition 5.4. Under the conditions of Proposition 5.4, $m \geqslant k_0 k$ (for large k_0). Let $q \in \Delta(J)$ be an optimal action of player two in the restricted game of level h. Let $y \in B(m, J)$ be the sequence provided by Claim 5.5. We define the strategy τ of player two by

$$\tau(a_1, \dots, a_t) = \begin{cases} y_{t+1} & \text{if for every } s, t-m+k \leqslant s < t \\ & (a^1_{s-k+1}, \dots, a^1_s) \neq (a^1_{t-k+1}, \dots, a^1_t); \\ \gamma(a^1_{s+1}) & \text{if there exists } s, t-m+k \leqslant s < t \\ & \text{s.t. } (a^1_{s-k+1}, \dots, a^1_s) = (a^1_{t-k+1}, \dots, a^1_t). \end{cases}$$

Let x be an oblivious k-recall strategy (sequence) of player one. Note that $\{x_t\}_{t \geqslant |I|^k}$ must be periodic. Let n denote the length of its period. We argue that the entropy of the empirical distribution of x is at least $\log n$.

Proof. For simplicity of notation, we assume that $x \in B(k, I)$ (i.e., x is $\mathbb{Z}$ indexed and periodic). Let t be a random variable that assumes values in $\{1, \dots, n\}$ with uniform distribution, and let $z_m = x_{m+t}$. The random process $\{z_m\}$ is stationary; therefore $\frac{1}{k}\log n = \frac{1}{k}H(z_1, \dots, z_k) \leqslant H(z_1) = H(\operatorname{emp}(x))$.

If $n < m - k$, then τ best-replies to x at any time $t > |I|^k$. If $n \geqslant m - k$, then $H(\operatorname{emp}(x)) \geqslant \frac{\log(m-k)}{k} > h + \epsilon_0 + \frac{\log(1-\frac{k}{m})}{k} > h + \epsilon_0 + O(\frac{1}{k_0})$. Claim 5.5 together with the observation that $\|\operatorname{emp}(x_1, \dots, x_T) - \operatorname{emp}(x)\| = O(\frac{1}{k_0})$ concludes the proof. □

Proof of Claim 5.5. Let us assume, w.l.o.g., that $H(q) > 0$. For $\delta > 0$, take k_0 so large that

$$k_0 H(q) > 1 + \log|I|, \tag{5.5}$$

$$k_0 > \delta^{-1}\log|I|. \tag{5.6}$$

Consider the set $C = C^q_\delta(k_0 k) := \{y \in B(k_0 k, J)\colon\ \|\operatorname{emp}(y) - q\| < \delta\}$. By Proposition 4.9, if k_0 is large enough, then for every $S \leqslant 2k_0|I|^k$,

$$\frac{\log|\operatorname{pref}_S C^q_\delta(k_0 k)|}{S} > H(q) - \delta. \tag{5.7}$$

Let $f : C \to B(k, I)$ be an arbitrary function. Choose $y \in C$ at random with uniform distribution. Denote $x_n := f(y)_n$. For integers $t > 0$ and $k_0|I|^k \leqslant S \leqslant 2k_0|I|^k$ consider the joint sequence $((x_{t+1}, y_{t+1}), \dots, (x_{t+S}, y_{t+S}))$. Note that $H(f(y)) \leqslant \log|B(k, I)| \leqslant |I|^k \log|I| < \delta S$. Neyman–Okada's criterion asserts that, for δ small enough,

$$\mathop{\mathbf{E}}_{x,y} \frac{1}{S}\sum_{i=1}^{S} g(x_{t+i}, y_{t+i}) \leqslant G\Big(\mathop{\mathbf{E}}_{x,y} \operatorname{emp}(x_{t+1}, \dots, x_{t+S}), q\Big) + \epsilon.$$

Consequently (by writing $T = S_1 + \dots + S_n$, $k_0|I|^k \leqslant S_i \leqslant 2k_0|I|^k$)

$$\epsilon \geqslant \mathop{\mathbf{E}}_{x,y}\left[\frac{1}{T}\sum_{t=1}^{T} g(x_t, y_t) - G\big(\operatorname{emp}(x_1, \dots, x_T), q\big)\right] \geqslant \min_{y \in C}\left[\frac{1}{T}\sum_{t=1}^{T} g(x_t, y_t) - G\big(\operatorname{emp}(x_1, \dots, x_T), q\big)\right]. \tag{5.8}$$

Now, (5.8) is true for any function f. Taking an f that maximizes the right-hand side of (5.8) concludes the proof. □

Remark. Renault et al. (2008) study a problem similar to Claim 5.5. They study the pure min–max value of $G^\infty[k_{obl}, m_{obl}]$. They argue that if the periods of two sequences are relatively prime, then the empirical distribution of the joint sequence is a product distribution. Using this argument one can prove Claim 5.5 (and hence Proposition 5.4) for $T \gg |I|^{2k}$.

6. Extensions and remarks

6.1. The special point $h = \log|I|$

In this section we give an example of a game G for which the fact that $\frac{\log T_k}{k}$ converges does not imply that $\operatorname{val} G^{T_k}[k, T_k]$ converges.

Consider the game M, called the “matching pennies” game, which can be described in strategic normal form by the following matrix:

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

A close look at the proof of Theorem 2.1 shows that the argument that deals with $h > \log|I|$ is valid also for $T \geqslant k^6 2^k$; therefore, the value of $M^{k^6 2^k}[k, k^6 2^k]$ converges to $V_*(M) = 0$.

Proposition 6.1.

$$\liminf_{k\to\infty} \operatorname{val} M^{2^k}\left[k_{obl}, 2^k\right] \geqslant \frac{1}{2}. \tag{6.1}$$

Proof. Let $x = x^k = \{x_n^k\}_{n\in\mathbb{Z}}$ be a random sequence that takes values in $DB(k, \{0,1\})$ with uniform distribution. Let p_n be the (posterior) distribution of x_n given $x_1, \ldots, x_{n-1}$. Let t be a random variable that distributes uniformly on $\{1, \ldots, 2^k\}$ independently of x. Consider the random variable $H(p_t)$. By the chain rule of entropy and de Bruijn’s enumeration theorem (Theorem 4.8)

$$\mathbf{E}H(p_t) = \frac{1}{2^k} H(x) = \frac{1}{2^k} \log\left(\left|DB\left(k, \{0,1\}\right)\right|\right) = \frac{\log 2}{2}.$$

Let X be the event in which the $k-1$ actions right before time t occurred once more before time t. Formally,

$$X = \left\{\text{“}\exists k \leqslant t' < t \text{ such that } (x_{t'-k+1}, \ldots, x_{t'-1}) = (x_{t-k+1}, \ldots, x_{t-1})\text{”}\right\}.$$

The action at time t must be other than the action at time t'. Hence, $H(p_t)\mathbf{1}_X = 0$. Since every sequence of $k-1$ actions occurs exactly twice in a de Bruijn cycle, $\Pr(X) \to \frac{1}{2}$, as $k \to \infty$. The random variable $\mathbf{1}_{X^c} - \frac{1}{\log 2} H(p_t)$ is non-negative and its expectation converges to zero as k grows; therefore $\forall \epsilon > 0$, $\Pr(H(p_t) > \log 2 - \epsilon) \to \frac{1}{2}$; and therefore

$$\liminf_{k\to\infty} \min_{\tau^k} M^{2^k}\left(x^k, \tau^k\right) \geqslant \frac{1}{2}. \qquad \square$$

More generally, let x^k distribute uniformly in $DB(k, I)$. Let T^k be the recurrence time of $x_1^k, \ldots, x_{k-1}^k$. That is,

$$T^k = \min\left\{n \geqslant 1\colon \left(x_1^k, \ldots, x_{k-1}^k\right) = \left(x_{n+1}^k, \ldots, x_{n+k-1}^k\right)\right\}.$$

Let $f(k)$ be a positive real-valued function that converges to zero as k grows.

Conjecture 6.2.

$$\lim_{k\to\infty} \Pr\left(T^k < f(k)|I|^k\right) = 0. \tag{6.2}$$

Conjecture 6.2 can be used to prove the following conjecture:

Conjecture 6.3. *Let $G = \langle I, J, g\rangle$ be a two-person zero-sum game. If $f(k) \to 0$ as $k \to \infty$, then*

$$\lim_{k\to\infty} \operatorname{val} G^{f(k)|I|^k}\left[k_{obl}, f(k)|I|^k\right] = \nu\left(\log|I|\right). \tag{6.3}$$

The arguments in the proof of Proposition 6.1 actually show that for every $f(k)$, if (6.2) holds then (6.3) also holds. Regardless of the connection between these conjectures, finding the “largest” $f(k)$’s for which (6.2) and (6.3) hold are of independent interest.

The “best” $f(k)$ for which we are able to prove (6.3) is of the order of $1/k$. We sketch the outlines of the proof. By employing Lovász’s local lemma (Alon and Spencer, 2000, p. 65), one can show that the proportion of sequences in which no sub-sequence of length k occurs twice among all sequences of length $\frac{|I|^k}{12k}$ is a sub-exponential function of k. This argument can replace Proposition 4.9 in the proof of Proposition 5.1; therefore (6.3) holds for $f(k) = 1/12k$.

6.2. Finitely many players with bounded recall

A natural generalization of the two-player zero-sum model is the multiple-player non-zero-sum model. As an illustration consider the special case where there are n players all with the same recall bound. Lehrer's (1988) main result infers that the (asymptotic) set of equilibrium payoffs in the bounded recall model contains the set of equilibrium payoffs in the classical model. Furthermore, these two sets coincide if there are only two players ($n = 2$). Although it is beyond the scope of this work, recent unpublished results (Peretz, 2010), based on the methods developed here, show that for $n > 2$ there are games in which the IRL in the bounded recall model is strictly below the IRL in the unbounded model. Consequently, there is a proper containment of the (asymptotic) sets of equilibrium payoffs.

During the punishing phase of the folk theorem a team of players can perform "concealed correlation" against a boundedly rational player. That is, the expected punishment given the deviator's information is a correlated action profile. Bavly and Neyman (2003) study this possibility, in the presence of a public signal (or a significantly superior player whose actions are payoff irrelevant). Recently, it has been shown (Peretz, 2010) that the presence of a public signal is not necessary.

6.3. $V^*(G^T[k,m]) = ?$

In light of parts (2.2a) and (2.2b) of Theorem 2.2 one may ask what is the pure min–max value of $G^T[k,m]$ and $G^T[k_{obl}, m]$ in the general case. To the best of our knowledge this question has not been answered.

Acknowledgments

This work was done as part of the author's M.A. and Ph.D. studies at the Hebrew University. The author would like to thank his advisor Prof. Abraham Neyman for his close guidance and endless patience. The author is also grateful to two anonymous referees and the editor Ehud Lehrer without whom this work would never have resulted in a paper, and to Prof. Eilon Solan for his comments.

Appendix A. The richness of *k*-recall sequences

In this section we state and prove the following theorem:

Theorem A.1. *Let $\{T_k\}_{k=1}^{\infty}$ be a sequence of positive integers, $T_k \to \infty$. Let $x_1, x_2, \ldots$ be a sequence of i.i.d. random variables with common distribution p (over a finite set of values). Denote*

$$\phi_k = \Pr\bigl(\forall 0 \leqslant t < s < T_k\ (x_{t+1}, \ldots, x_{t+k}) \neq (x_{s+1}, \ldots, x_{s+k})\bigr).$$

If $\limsup \frac{\log T_k}{k} < H(p)$, then $\frac{\log \phi_k}{T_k} \to 0$, as $k \to \infty$.

Remark. The assumption of Theorem A.1 that $\limsup \frac{\log T_k}{k} < H(p)$ is necessary.[11] On the other hand, the theorem does not tell us how large ϕ_k is. Finding an explicit expression for (the asymptotic of) that probability is of interest.

We begin by reducing Theorem A.1 to the following combinatorial statement:

Proposition A.2. *Let $\{T_k\}_{k=1}^{\infty}$ be a sequence of positive integers, $T_k \to_{k\to\infty} \infty$. Let A be a finite alphabet and $p \in \Delta(A)$. If $\limsup \frac{\log T_k}{k} < H(p)$, then there exist sets $C(k) \subset B(k, A)$ satisfying*

$$\forall x \in C(k)\ \forall 0 \leqslant t < s < T_k\ (x_{t+1}, \ldots, x_{t+k}) \neq (x_{s+1}, \ldots, x_{s+k}),$$

$$\lim_k \max_{x \in C(k)} \bigl\|\mathrm{emp}(x) - p\bigr\| = 0,$$

$$\liminf_k \frac{\log |\mathrm{pref}_{T_k}\ C(k)|}{T_k} \geqslant H(p). \tag{A.1}$$

Claim. *Proposition* A.2 *implies Theorem* A.1.

Proof. First, since the number of different types $|\{T^q(n)\colon q \in \Delta(A)\}|$ is polynomial in n, there exist p_k's such that

[11] By considering only s and t congruent to 1 modulo k, one obtains an instance of the well-studied "birthday" problem. The birthday problem is the problem of computing the probability that n i.i.d. random variables with common distribution p will assume distinct values. See Feller (1958, pp. 31–32).

$$\liminf_k \frac{\log |T^{p_k}(T_k) \cap \mathrm{pref}_{T_k} C(k)|}{T_k} \geqslant H(p). \tag{A.2}$$

Take $x = x_1, \ldots, x_{T_k}$ i.i.d. random variables with distribution p.

$$\Pr\bigl(x \in T^{p_k}(T_k) \cap \mathrm{pref}_{T_k} C(k)\bigr) = \Pr\bigl(\mathrm{emp}(x) = p_k\bigr) \cdot \frac{|T^{p_k}(T_k) \cap \mathrm{pref}_{T_k} C(k)|}{|T^{p_k}(T_k)|}. \tag{A.3}$$

By Proposition 4.7,

$$\Pr_{x \sim p^n}\bigl(\mathrm{emp}(x) = q\bigr) \geqslant \frac{\exp(-D(q\|p)n)}{n^{|A|}}; \tag{A.4}$$

and, by Proposition 4.6,

$$\bigl|T^q(n)\bigr| \leqslant \exp\bigl(H(q)n\bigr). \tag{A.5}$$

Combining (A.2), (A.4), and (A.5) we get the required result. □

In the rest of this section we prove Proposition A.2.

For $k = 1, 2, \ldots,$ we construct sets $C(k)$ satisfying (A.1). We assume w.l.o.g. that $k < T_k$. Let $l = l(k)$ and $m = m(k)$ be integers with the following properties:

$$1 \ll l(k) \ll k, \tag{A.6}$$

$$m(k)l(k) + 2l(k) < k, \tag{A.7}$$

$$m(k)l(k) \sim k. \tag{A.8}$$

Let A, p, and T_k be as given in Proposition A.2. Let $q = q(k) \in \Delta(A)$ such that

$$q(k) \to_{k\to\infty} p,$$
$$\forall k \; T^q(l) \neq \emptyset.$$

Consider the alphabet $B = T^q(l)$. For every de Bruijn sequence $\hat{x} \in DB(B, m)$, we define a corresponding sequence x in $A^{\mathbb{Z}}$. Let α be a least probable element of A with respect to the probability mass q, and let $\beta \neq \alpha$ be another element[12] of A. Let $b = \alpha^l\beta$. That is, b is a word over the alphabet A that consists of l consecutive *alpha*s followed by one *beta*. The correspondence $\hat{x} \mapsto x$ is defined as follows:

$$x = \ldots b\hat{x}_1\hat{x}_2 \ldots \hat{x}_m b\hat{x}_{m+1}\hat{x}_{m+2} \ldots \hat{x}_{m+m} b \ldots.$$

That is, x is the concatenation of the elements of $\hat{x}$ separated by a b after every m-th element.

The set C $(= C(k))$ consists of the those corresponding sequences. That is,

$$C := \bigl\{x \in A^{\mathbb{Z}} \bigm| x \text{ corresponds to some } \hat{x} \in DB(B, m)\bigr\}.$$

The first two lines of (A.1) hold trivially. It remains to verify that the last line of (A.1) holds. Since $DB(B, m)$ is invariant to shifts, for every $T < |B|^m$ we have

$$\frac{\log |\{\hat{x}_1 \ldots \hat{x}_T \mid \hat{x} \in DB(B, m)\}|}{T} \geqslant \frac{\log |DB(B, m)|}{|B|^m} = \frac{\log(|B|!)}{|B|} \geqslant \log|B| - \log\log|B|. \tag{A.9}$$

The last inequality follows from the inequality of means $\frac{n}{\log n} \leqslant \frac{n}{1+\frac{1}{2}+\cdots+\frac{1}{n}} \leqslant \sqrt[n]{n!}$. Let T be an integer, $\frac{T_k}{l(k)} \leqslant T \leqslant B^{m(k)}$. Substituting in (A.9) we obtain

$$\frac{\log|\mathrm{pref}_{T_k} C(k)|}{T_k} \geqslant \frac{\log |\{\hat{x}_1 \ldots \hat{x}_T \mid \hat{x} \in DB(B, m)\}|}{(l(k))T} \geqslant \frac{\log|B| - \log\log|B|}{l(k)} \to H(p).$$

Finally, we have to verify that such a T exists. It is sufficient to show that $T_k \leqslant |B|^{m(k)}$, and indeed:

$$\frac{\log|B|^{m(k)}}{k} = \frac{l(k)m(k)}{k}\frac{\log|B|}{l(k)} \to H(p) > \limsup \frac{\log T_k}{k}. \qquad \square$$

[12] We assume w.l.o.g. that A includes more than one element.

Appendix B. Conditional probabilities of approximated i.i.d. sequences

The following lemma shows that an approximation of p^n conditioned on an event whose probability is bounded away from zero is also an approximation of p^n.

Lemma B.1. $\forall p \in \Delta(I)\ \forall \epsilon > 0\ \exists C > 0\ \forall 1 \geqslant \delta > 0\ \forall n \in \mathbb{N}\ \forall \mu \in \Delta(I^n)$.
If

- μ *is a* δ*-approximation of* p^n,
- μ *is a convex combination of* $\mu_1, \mu_2 \in \Delta(I^n)$, $\mu = \alpha_1\mu_1 + \alpha_2\mu_2$, *and*
- $\alpha_1 > \epsilon$,

then, μ_1 *is a* $(C\sqrt{\delta + H(\delta) + 1/n})$*-approximation of* p^n.

Proof. Given a measure $\mu \in \Delta(I)$, let us denote its empirical distribution by $\bar{\mu} \in \Delta(I)$. That is, $\bar{\mu}(A) = \frac{1}{n}\sum_{t=1}^{n} \mu(I^{t-1} \times A \times I^{n-t})$. We first assume that $\bar{\mu}$ is absolutely continuous[13] with respect to p (denoted $\bar{\mu} \ll p$). With this assumption we will prove the stronger statement that μ_1 is a $(C\sqrt{\max\{\delta, 1/n\}})$-approximation of p^n.

Denote $\lambda = \max\{\delta, 1/n\}$. Note that if $q \ll p$ then $L_p(q) := H(q) + D(q\|p) = -\sum_k \log(p_i)q_i$ is a linear function of q. Also, $L_{p^n}(\mu) = nL_p(\bar{\mu})$; therefore

$$\frac{1}{n}D\big(\mu\|p^n\big) + \left(\frac{1}{n}H(\mu) - H(p)\right) = D(\bar{\mu}\|p) + \big(H(\bar{\mu}) - H(p)\big) = L(\bar{\mu} - p). \tag{B.1}$$

With the assumptions of Lemma B.1, we have

$$\alpha_1 H(\mu_1) + \alpha_2 H(\mu_2) \geqslant H(\mu) - H(\alpha) \geqslant nH(p) - n\delta - H(\alpha). \tag{B.2}$$

Combining (B.1) and (B.2) we have

$$\sum_{i=1,2} \alpha_i \frac{1}{n} D\big(\mu_i\|p^n\big) = H(p) - \sum_{i=1,2} \alpha_i \frac{1}{n} H(\mu_i) + L(\bar{\mu} - p) \leqslant \delta + \frac{H(\alpha)}{n} + \|L\|\|\bar{\mu} - p\|_1 \leqslant \lambda\big(1 + H(\alpha) + \|L\|\big) \tag{B.3}$$

where $\|L\| = \max_{i \in I} \log \frac{1}{p_i}$. Pinsker's inequality, (B.1), (B.3), and the fact that the function $\frac{H(x)}{x}$ is decreasing yield

$$\frac{1}{2}\|\bar{\mu}_1 - p\|^2 \leqslant D(\bar{\mu}_1\|p) \leqslant \frac{1}{n}D\big(\mu_1\|p^n\big) \leqslant \lambda\frac{1 + \|L\| + H(\epsilon)}{\epsilon} = \lambda C_1. \tag{B.4}$$

By (B.1) and (B.4) we have

$$\left|\frac{1}{n}H(\mu_1) - H(p)\right| \leqslant \frac{1}{n}D\big(\mu_1\|p^n\big) + \|L\|\|\bar{\mu}_1 - p\| \leqslant \lambda C_1 + \|L\|\sqrt{2\lambda C_1} \leqslant \sqrt{\lambda}\big(C_1 + \|L\|\sqrt{2C_1}\big),$$

and

$$\|\bar{\mu}_1 - p\| \leqslant \sqrt{2C_1\lambda}.$$

Hence, there exists $C = C(\epsilon, p)$, such that μ_1 is a $(C\sqrt{\lambda})$-approximation of p^n.

Now, we no longer assume that $\bar{\mu} \ll p$. Let $\beta = \sum_{i \notin \operatorname{support} p} \bar{\mu}(i)$. We will show that there exists a $C = C(\epsilon, p)$ such that μ_1 is a $(C\sqrt{\delta + H(\beta) + 1/n})$-approximation of p^n. So far, the case $\beta = 0$ has been proven.

For a measure η on a (measurable) set A and a (measurable) function $\varphi : A \to B$, $\varphi \circ \eta$ denotes a measure on B, defined by $\varphi \circ \eta(X) = \eta(\varphi^{-1}(X))$.

Let $i_0 \in \operatorname{support} p$. Define $(f, g) : I \to I \times I$ as follows:

$$(f, g)(i) = \begin{cases} (i, i_0) & \text{if } i \notin \operatorname{support} p, \\ (i_0, i) & \text{if } i \in \operatorname{support} p. \end{cases}$$

The definition of f and g extends to I^n in the obvious way. Note that

- the map $\eta \mapsto g \circ \eta$ is a linear transformation; and
- $g \bar{\circ} \eta = g \circ \bar{\eta}$.

[13] We say that q is *absolutely continuous* with respect to p, if $p(A) = 0 \Rightarrow q(A) = 0$, for every A.

The mapping (f, g) is a one-to-one mapping; therefore $H((f, g) \circ \mu) = H(\mu)$; hence

$$H(g \circ \mu) \geqslant H(\mu) - H(f \circ \mu). \tag{B.5}$$

We can write $f \circ \bar{\mu} = \beta q + (1 - \beta) i_0$; where $q \in \Delta(I \setminus \text{support}(p))$; therefore

$$\frac{1}{n} H(f \circ \mu) \leqslant H(\overline{f \circ \mu}) = H(f \circ \bar{\mu}) \leqslant H(\beta) + \beta H(q) \leqslant H(\beta) + \beta \log |I| \leqslant \bigl(1 + \log |I|\bigr)\bigl(H(\beta) + \delta\bigr). \tag{B.6}$$

Also,

$$\|g \circ \bar{\mu} - p\| \leqslant \|\bar{\mu} - p\| < \delta. \tag{B.7}$$

By (B.5), (B.6), (B.7), and the fact that μ is a δ-approximation of p^n, $g \circ \mu$ is a $(2 + \log |I|)(H(\beta) + \delta)$-approximation of p^n. Since $\overline{g \circ \mu} \ll p$, we were able to show that there exists a constant $C = C(\epsilon, p)$, such that $g \circ \mu_1$ is a $C\sqrt{\delta + H(\beta) + 1/n}$-approximation of p^n.

$H(\mu_1) \geqslant H(g \circ \mu_1)$; therefore the lemma can be concluded by showing that $\|g \circ \bar{\mu}_1 - p\| \leqslant 2\|\bar{\mu}_1 - p\|$. Note that

$$\sum_{\substack{i \in I: \\ \bar{\mu}_1(i) < p(i)}} p(i) - \bar{\mu}_1(i) = \frac{1}{2} \|\bar{\mu}_1 - p\|,$$

and

$$\sum_{\substack{i \in I \setminus \{i_0\}: \\ \bar{\mu}_1(i) < p(i)}} p(i) - \bar{\mu}_1(i) = \frac{1}{2} \|g \circ \bar{\mu}_1 - p\|.$$

Since this is true for any choice of $i_0 \in \text{support}\, p$, assuming w.l.o.g. that $|\text{support}\, p| \geqslant 2$, there must be some $i_0 \in \text{support}\, p$ for which $\|g \circ \bar{\mu}_1 - p\| \leqslant 2\|\bar{\mu}_1 - p\|$. $\square$

References

Alon, Noga, Spencer, Joel H., 2000. The Probabilistic Method, second edn. John Wiley & Sons, New York.

Aumann, Robert J., 1981. Survey of repeated games. In: Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern. Bibliographisches Institut, Mannheim, pp. 11–42.

Bavly, Gilad, Neyman, Abraham, September 2003. Online concealed correlation by boundedly rational players. Discussion Paper 336, Center for the Study of Rationality, Hebrew University, Jerusalem.

Cover, Thomas M., Thomas, Joy A., 2006. Elements of Information Theory. Wiley Interscience, New York.

de Bruijn, Nicolaas Govert, 1946. A combinatorial problem. K. Ned. Akad. v. Wet. 49, 758–764.

Feller, William, 1958. An Introduction to Probability Theory and Its Applications, 2nd edn. Wiley Interscience, New York.

Lehrer, Ehud, 1988. Repeated games with stationary bounded recall strategies. J. Econ. Theory 46 (1), 130–144.

Neyman, Abraham, February 2008. Learning effectiveness and memory size. Discussion Paper 476, Center for the Study of Rationality, Hebrew University, Jerusalem.

Neyman, Abraham, Okada, Daijiro, 2000. Repeated games with bounded entropy. Games Econ. Behav. 30 (2), 228–247.

Neyman, Abraham, Okada, Daijiro, 2009. Growth of strategy sets, entropy, and nonstationary bounded recall. Games Econ. Behav. 66 (1), 404–425.

Peretz, Ron, 2010. Repeated games with bounded complexity. PhD thesis, Hebrew University, Jerusalem.

Renault, Jérôme, Scarsini, Marco, Tomala, Tristan, 2008. Playing off-line games with bounded rationality. Math. Soc. Sci. 56 (2), 207–223.

Van Aardenne-Ehrenfest, T., de Bruijn, Nicolaas Govert, 1951. Circuits and trees in oriented linear graphs. Simon Stevin 28, 203–217.