THE BIG MATCH WITH A CLOCK AND A BIT OF MEMORY

KRISTOFFER ARNSFELT HANSEN, RASMUS IBSEN-JENSEN,
AND ABRAHAM NEYMAN

Abstract. The Big Match is a multi-stage two-player game. In each stage Player 1 hides one or two pebbles in his hand, and his opponent has to guess that number; Player 1 loses a point if Player 2 is correct, and otherwise he wins a point. As soon as Player 1 hides one pebble, the players cannot change their choices in any future stage.

Blackwell and Ferguson (1968) give an ε-optimal strategy for Player 1 that hides, in each stage, one pebble with a probability that depends on the entire past history. Any strategy that depends just on the clock *or* on a finite memory is worthless. The long-standing natural open problem has been whether every strategy that depends just on the clock *and* a finite memory is worthless.

The present paper proves that there is such a strategy that is ε-optimal. In fact, we show that just two states of memory are sufficient.

Date: December 15, 2017.

1. Introduction

The game of Odd and Even (Latin: *Par Impar Ludere*, Greek: ἀρτιασμός) has been popular since ancient Greek and Roman times. It is played by two players, Player 1 and Player 2. Player 1 hides (e.g., in his hands) a number of pebbles or other items (e.g., beans, nuts, almonds, astragali, or coins), and his opponent, Player 2, has to guess whether the number of hidden items is odd or even. Player 1 then reveals the number. If Player 2 is right, Player 1 loses a point; otherwise, Player 1 wins a point (from Player 2).

Player 1 can guarantee that he gets (at least) zero points on average by hiding an odd or even number of items with equal probability. Player 2 can guarantee that Player 1 gets (at most) zero points on average by guessing odd or even with equal probability.

The repeated Odd and Even game is the same game repeated many times. The players can still guarantee getting zero points on average (per stage and hence also in total) by playing, independently in each stage as before.

The Big Match is also a multi-stage game. It is a variant of the repeated Odd and Even game. In each stage Player 1 hides one or two pebbles. In each stage, Player 1 wins or loses a point. As long as Player 1 hides two pebbles, Player 1 wins a point iff Player 2 guesses odd in that stage. The first stage in which Player 1 hides one pebble is called the stopping stage. In the stopping stage Player 1 wins a point iff Player 2 guesses even. In each subsequent stage, he wins a point iff he won a point in the stopping stage.

The Big Match was introduced by Gillette [3] and has been much studied, in part due to its arguably being the most basic game model that illustrates the difficulty of balancing the trade-off between short- and long-term strategic considerations.

In the Big Match, Player 2 can still guarantee that Player 1 gets zero points on average, independently of the number of stages, by guessing odd or even with equal probability and independently in each stage. Executing such a strategy does not require that Player 2 know past history, the number of stages, or the stage number. However, the situation of Player 1 is completely different! Henceforth, unless otherwise mentioned, a strategy refers to a strategy of Player 1.

If Player 1 knows the number of stages, n, in advance, he can guarantee that he gets (at least) zero points on average. To guarantee this, he *must* hide one pebble with probability $\frac{1}{k+1}$ when k stages remain. Thus, for example, in the last stage he hides one or two pebbles with equal probability, and in the first stage he hides one pebble with probability $\frac{1}{n+1}$. Executing such a strategy requires that Player 1 know the stage number and the number of stages, but it does not require that Player 1 know past history.

It follows from the above that if Player 1 does not know the number of stages n in advance, then he has no way of guaranteeing (at least) zero points (per stage) on average. This has led researchers to look for strategies that guarantee close to zero per stage on average in all sufficiently long Big Match games.

Any strategy in the Big Match has to decide on the stopping stage. A natural possibility is just to specify in advance the probability of each stage being the stopping stage. Such a strategy is called a Markov strategy. It has long been known, and it is easy to verify, that any Markov strategy in the Big Match is *worthless*; i.e., for any $\varepsilon > 0$ it does not guarantee more than $-1+\varepsilon$ points (per stage) on average in any sufficiently long Big Match game.

The principle of sunk cost seems to imply that optimizing from any point onwards should be independent of the past, and hence any optimization of the long-run average of the rewards can be achieved by a Markov strategy. Since any Markov strategy is worthless, one may erroneously conclude that any strategy is worthless.

However, this is not the case! Blackwell and Ferguson [2] introduced worthy (i.e., not worthless) strategies that prescribe the choice in each stage as a function of past history. Moreover, [2] introduced, for every $\varepsilon > 0$, a strategy that is *ε-optimal*; namely, it guarantees at least $-\varepsilon$ points (per stage) on average in all sufficiently long games.[1]

The question that arises is how much dependence on past history is needed for an ε-optimal strategy, or even a worthy one. This dependence is formalized using the following concept.

A *memory-based* strategy in the Big Match is a strategy in which the conditional probability of hiding one pebble depends on the current memory state and the clock (i.e, the stage number). The memory state is updated as a stochastic function of the current memory and of the guess of Player 2 in the previous stage, as well as of the clock.

The ε-optimal strategies in [2, Theorem 2] are memory-based, and those in [2, Theorem 1] are memory-based and clock-independent; i.e., the hiding and memory updating do not depend on the clock. The memory state is simply the difference between the number of odd and even guesses; hence, up to stage n it takes integer values in the interval $[-n, n]$.

The ε-optimal strategy in [4] is memory-based and clock-independent. The memory state can be encoded so that, with high probability, up to stage n it takes integer values in $[0, \ln^c n]$, for some constant c (and $n > 3$).

On the other hand, all memory-based strategies that have a finite set of memory states and either are clock-independent (see, e.g., [7]) or have a deterministic memory update function [4] are worthless in the Big Match.

It has been a long-standing natural open problem whether there exists a worthy memory-based strategy that has a finite set of memory states.

[1] Recall that Player 2 has a strategy that ensures 0 per stage on average.

The present paper answers this question positively. We show that, for every $\varepsilon > 0$, there is such a strategy that is ε-optimal. Moreover, it is a two-memory strategy; namely, it has a two-element memory set.

Our positive result applies to the infinite game as well. In the infinite game, the average win per stage need not be well defined, as the average number of wins over the first n stages need not converge. Nonetheless, our result is as strong as possible given that the strategy's ε-optimality in the infinite game is for the limit inferior.

2. The model and related results

2.1. **Stochastic games.** A *finite two-person zero-sum stochastic game* Γ, henceforth, a *stochastic game*, is defined by a tuple (Z, I, J, r, p), where Z is a finite state space, I and J are the finite actions sets of Players 1 and 2 respectively, $r : Z \times I \times J \to \mathbb{R}$ is a payoff function, and $p : Z \times I \times J \to \Delta(Z)$ is a transition function.

A state $z \in Z$ is called an *absorbing state* if $p(z, \cdot, \cdot) = \delta_z$, where δ_z is the Dirac measure on z. An absorbing game is a stochastic games with only one nonabsorbing state.

A *play* of the stochastic game is an infinite sequence $z_1, \ldots, z_t, i_t, j_t, \ldots$, where $(z_t, i_t, j_t) \in Z \times I \times J$. The set of all plays is denoted by H_∞. A play up to stage t is the finite sequence $h_t = (z_1, i_1, j_1, \ldots, z_t)$. The payoff r_t in stage t is $r(z_t, i_t, j_t)$ and the average of the payoffs in the first n stages, $\frac{1}{n}\sum_{t=1}^n r_t$, is denoted by $\bar{r}_n$.

The initial state of the multi-stage game is $z_1 \in Z$. In the t-th stage players simultaneously choose actions $i_t \in I$ and $j_t \in J$.

A behavioral strategy of Player 1, respectively Player 2, is a function σ, respectively τ, from the disjoint union $\dot\cup_{t=1}^\infty (Z \times I \times J)^{t-1} \times Z$ to $\Delta(I)$, respectively to $\Delta(J)$. The restriction of σ, respectively τ, to $(Z \times I \times J)^{t-1} \times Z$ is denoted by σ_t, respectively τ_t. In what follows, σ denotes a strategy of Player 1 and τ denotes a strategy of Player 2.

A strategy pair (σ, τ) defines a probability distribution $P_{\sigma,\tau}$ on the space of plays as follows. The conditional probability of $(i_t = i, j_t = j)$ given the play h_t up to stage t is the product of $\sigma(h_t)[i]$ and $\tau(h_t)[j]$. The conditional distribution of z_{t+1} given h_t, i_t, j_t is $p(z_t, i_t, j_t)$. The expectation w.r.t. $P_{\sigma,\tau}$ is denoted by $E_{\sigma,\tau}$

A stochastic game has a value $v = (v(z))_{z \in Z}$ if, for every $\varepsilon > 0$, there are strategies σ_ε and τ_ε such that for some positive integer n_ε

$$\varepsilon + E_{\sigma_\varepsilon,\tau}\bar{r}_n \geq v(z_1) \geq E_{\sigma,\tau_\varepsilon}\bar{r}_n - \varepsilon \quad \forall \sigma, \tau, n \geq n_\varepsilon, \tag{1}$$

and

$$\varepsilon + E_{\sigma_\varepsilon,\tau} \liminf_{n\to\infty} \bar{r}_n \geq v(z_1) \geq E_{\sigma,\tau_\varepsilon} \limsup_{n\to\infty} \bar{r}_n - \varepsilon \quad \forall \sigma, \tau. \tag{2}$$

It is known that all absorbing games [5] and, more generally, all stochastic games [6] have a value.

A strategy σ_ε that satisfies the left-hand inequality (1) is called *uniform ε-optimal.* A strategy σ_ε that satisfies the left-hand inequality (2) is called *limiting-average ε-optimal.*

A strategy σ_ε that satisfies both left-hand inequalities (1) and (2) is called *ε-optimal.*

2.2. Memory-based strategies.

A *memory-based strategy* σ generates a random sequence of memory states $m_1, \ldots, m_t, m_{t+1}, \ldots$, where the memory is updated stochastically in each stage, and selects its action i_t according to a distribution that depends on just the current time t, its current memory m_t, and the current state z_t. Explicitly, the conditional distribution of i_t, given $h_t^m := (z_1, m_1, i_1, j_1, \ldots, z_t, m_t)$, is a function σ_α of (t, z_t, m_t) and the conditional distribution of m_{t+1}, given $(h_t^m, i_t, j_t, z_{t+1})$, is a function σ_m of (t, z_t, m_t, i_t, j_t) (i.e., it depends on just the time t and the tuple (z_t, m_t, i_t, j_t)).

A memory-based strategy σ is *clock-independent* if the functions σ_α and σ_m are independent of t.

A *k-memory strategy* is a memory-based strategy in which the memory states m_t take values in a set with (at most) k elements. Note that a strategy is a Markov strategy if and only if it is a one-memory strategy, and a strategy is a stationary strategy if and only if it is a one-memory clock-independent strategy. A strategy *uses finite memory* if it is a k-memory strategy where k is finite. A strategy that uses finite memory is called a *finite-memory strategy.* The set of all k-memory strategies is denoted by $\mathcal{M}_k$.

The long-standing natural open problem that motivates the present paper is whether for every stochastic game, or even just the Big Match, there are ε-optimal strategies that use finite memory.

Another natural problem is the existence of memory-based strategies with infinite memory but where the number of distinct memory states used in the first n stages grows slowly with high probability. For this question we assume, w.l.o.g., that the memories take values in the set $\mathbb{N}$ of natural numbers. Let $f : \mathbb{N} \to \mathbb{R}_+$ be a nondecreasing function and $\gamma \geq 0$.

An *(f, γ)-memory strategy* is a memory-based strategy in which, with probability at least $1 - \gamma$, $\mathbb{N} \ni m_t \leq f(t)$ for all t and every strategy of the other player.

2.3. The Big Match.

The Big Match, introduced in [3], is a highly inspiring stochastic game. The state space Z is $\{-1, 0, 1\}$.

Each state $x \in \{-1, 1\}$ is absorbing and the payoff function (to Player 1) in an absorbing state x is $r(x, \cdot, \cdot) = x$.

The action sets I and J are $\{0, 1\}$, and the payoff function in the nonabsorbing state 0 is

$$r(0,i,j)=\begin{cases} 1 & \text{if } j\neq i,\\ -1 & \text{if } j=i.\end{cases}$$

The transition distributions from the nonabsorbing state 0 are given by

$$p(0,i,j)=\begin{cases} \delta_0 & \text{if } i=0\\ \delta_{\text{-}1} & \text{if } i=j=1\\ \delta_1 & \text{if } i=1\neq j.\end{cases}$$

[2] shows that the value of the Big Match is 0 by introducing, for every $\varepsilon > 0$, an ε-optimal strategy (which is, in addition, a clock-independent $(f, 0)$-memory strategy where $f(n) = 2n - 1$).

[4] introduces, for the Big Match (and also for any absorbing game), a clock-independent (f, γ)-memory strategy, where $f(n) = (\log n)^{O(1)}$ and any $\gamma > 0$, that is ε-optimal.

Fix $\varepsilon < 1$. It is known that there is neither a limiting-average nor a uniform ε-optimal strategy that is a finite-memory strategy that uses a deterministic memory updating function σ_m; see [4] for the limiting-average case. Moreover, there is no ε-optimal mixed strategy that is a mixture of finitely many finite-memory strategies that each use a deterministic memory updating function σ_m.

It is also known that there is neither a limiting-average nor a uniform ε-optimal strategy that is a clock-independent finite-memory strategy; see, e.g., [7] for the limiting-average case. Moreover, there is no mixed strategy that is a mixture of clock-independent finite-memory strategies that is ε-optimal [1].

3. The result

The main result of the present paper is that, in the Big Match, there is a finite-memory strategy that is ε-optimal and moreover that is a two-memory strategy.

Theorem 1. *For every $\varepsilon > 0$ there is a 2-memory strategy σ of Player 1 and n_ε such that for every strategy τ of Player 2,*

$$E_{\sigma,\tau} \liminf_{n\to\infty} \bar{r}_n \geq -\varepsilon, \tag{3}$$

and

$$E_{\sigma,\tau}\bar{r}_n \geq -\varepsilon \qquad \forall n \geq n_\varepsilon. \tag{4}$$

4. The proof

The set of stages $t = 1, 2, \ldots$ of the infinite game is partitioned into consecutive epochs, indexed by $i = 1, 2, \ldots$, where the number of stages of the i-th epoch is s_i.

The number of stages in the first n epochs equals $\sum_{i=1}^n s_i$ and is denoted by S_n. The payoff to Player 1 in the j-th round of epoch i is denoted by r^i_j. Note that the j-th round of epoch i is the $(S_{i-1} + j)$-th stage of the game. Therefore, $\sum_{t=1}^{S_n} r_t = \sum_{i=1}^n \sum_{j=1}^{s_i} r^i_j$.

If s_n/S_n goes to 0 as n goes to infinity, then in order to prove the theorem we have to define a strategy $\sigma \in \mathcal{M}_2$ of Player 1 and n_ε such that for every pure strategy x of Player 2 and $n \geq n_\varepsilon$, we have

$$E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n} \sum_{i=1}^{n} \sum_{j=1}^{s_i} r^i_j \geq -5\varepsilon \tag{5}$$

and

$$E_{\sigma,x} \frac{1}{S_n} \sum_{i=1}^{n} \sum_{j=1}^{s_i} r^i_j \geq -5\varepsilon. \tag{6}$$

The strategy σ consists of patching together strategies $\sigma_{s_i,\rho}$, which will be defined later, where $\sigma_{s_i,\rho}$ is a strategy in the i-th epoch. The strategy $\sigma_{s_i,\rho}$ in the i-th epoch depends on the number of stages s_i in the epoch and a fixed positive number ρ. If $s = 1$ then $\sigma_{s,\rho}$ plays the nonabsorbing action. We proceed with the definition of the strategy $\sigma_{s,\rho}$ for $s > 1$.

Let $\delta > 0$ be a sufficiently small number and let $s > 1$ be an integer.

We define a strategy $\sigma_{s,\delta}$ of Player 1 in the auxiliary game in three equivalent descriptions. The first description illustrates that $\sigma_{s,\delta}$ is in $\mathcal{M}_2$, namely, that it is a two-memory strategy.

The two states of memory of the strategy $\sigma_{s,\delta}$ are $\widehat{C}$ (for continuing throughout) and $\widehat{A}$ (for possible future absorption).

The initial state of memory, m_1, is $\widehat{A}$. On $m_t = A$ and $x_t = 1$ (i.e., in the case that $m_t = A$ and $x_t = 1$), the conditional probability that $m_{t+1} = C$ is $1 - e^{-2\delta}$, and in all other cases, $m_{t+1} = m_t$.

The strategy plays the action C if it is in memory state $\widehat{C}$. When it is in memory state $\widehat{A}$ it plays the absorbing action A with a conditional probability that depends on the round, and this conditional probability is given by $q_i/(1 - \sum_{k<j} q_k)$, where

$$q_j := (e^\delta - 1)e^{-\delta s}e^{\delta(j-1)}.$$

The second description, which is easily seen to be equivalent to the first one, is as follows. Select a positive integer ℓ, where for each $1 \leq j \leq s$, the probability that the selected ℓ equals j is q_j. Sample the action of Player 2 in each round k with probability $1 - e^{-2\delta}$ and let the sampling of the different rounds be independent. Play the absorbing action in round j iff $j = \ell \leq s$ and the payoff in each of the previously sampled rounds is -1.

The third description is useful in the proof, and in addition it is short. The strategy $\sigma_{s,\delta}$ of Player 1 plays the absorbing action in stage $1 \leq j \leq s$ with (unconditional) probability

$$p_j := (e^{\delta} - 1)e^{-\delta s}e^{-\delta \sum_{k<j} x_k},$$

where x_k is the payoff of Player 1 at stage k.

Obviously, one has to show that the third description indeed defines a strategy. This will follow from the sequel where we show that this description is equivalent to the previous ones, which clearly define a strategy. For completeness, we show that the third description defines a strategy without relying on the equivalence of the descriptions.

Note that p_j is a function of the stream of past actions and thus well defined. In addition, $p_j \geq 0$ and

$$\begin{aligned}\sum_{j=1}^{s} p_j &= \sum_{j=1}^{s}(e^{\delta} - 1)e^{-\delta s}e^{-\delta \sum_{k<j} x_k} \leq \sum_{j=1}^{s}(e^{\delta} - 1)e^{-\delta s}e^{\delta(j-1)} \\ &= 1 - e^{-\delta s},\end{aligned}$$

and therefore the third description defines a strategy.

Now we show that the third description is equivalent to the previous ones. The (unconditional) probability (of the second described strategy) of playing the absorbing action at stage j equals q_j times the probability of no sampling in an earlier round where Player 2 played the action 1. As the number of rounds before j where Player 2 played the action 1 equals $\sum_{k<j} \frac{x_k+1}{2}$ and the conditional probability of not sampling a round is $e^{-2\delta}$, the above-mentioned product equals

$$q_j e^{-2\delta \sum_{k<j} \frac{x_k+1}{2}} = (e^{\delta} - 1)e^{-\delta s}e^{\delta(j-1)}e^{-\delta(j-1)-\delta \sum_{k<j} x_k} = p_j.$$

Consider the auxiliary games with s+1 stages, where dynamics and stage payoffs follow the rules of the Big Match and the players are active only in the first s stages j, $j = 1, \dots, s$.

Let $\sigma = \sigma_{s,\delta}$. We study the distribution of the state in the last period, $s + 1$, as a function of the strategy σ of Player 1 and a pure strategy τ of Player 2.

Let τ be a pure strategy of Player 2. Labeling the left-column action of Player 2 by -1 and the right-column action of Player 2 by 1, the pure strategy τ of Player 2 is identified with the sequence of actions $x = x(\tau) = (x_1, \dots, x_s)$.

Define a function v on plays of the auxiliary (s+1)-stage game as follows. If the play is absorbed in the winning state for Player 1, then $v = 1$. If the play is absorbed in the losing state for Player 1, then $v = -e^{-\delta}$, and otherwise $v = 0$.

Lemma 1. *Let* $\alpha(x) = -\sum_{j=1}^{s} x_j/s$. *Then*

$$\begin{aligned} E_{\sigma,x}v &= e^{(\alpha(x)-1)s\delta} - e^{-s\delta} \qquad (7)\\ &\geq e^{(\varepsilon-1)s\delta}1_{\{\alpha(x)\geq\varepsilon\}} - e^{-s\delta} \quad \forall\varepsilon > 0. \qquad (8)\end{aligned}$$

Proof. For every integer c let J_c^+ be the set of all indices $1 \leq j \leq s$ such that $c = -\sum_{k<j} x_k$ and $x_j = -1$, and let J_c^- be the set of all indices $1 \leq j \leq s$ such that $c + 1 = -\sum_{k<j} x_k$ and $x_j = 1$. Obviously, for each integer c, the sets of indices J_c^+ and J_c^- are disjoint, and the set of integers is the disjoint union $\cup_c(J_c^+ \cup J_c^-)$.

Obviously, if $j \in J_c^+$ and $j' \in J_c^-$ then $e^{-\delta(c+1)} = e^{-\delta}e^{-\delta\sum_{k<j}x_k} = e^{-\delta\sum_{k<j'}x_k}$. Therefore, using the third description of the strategy, we have

$$\begin{aligned} E_{\sigma,x}v &= \sum_{j=1}^{s} 1_{\{x_j=-1\}}p_j - \sum_{j=1}^{s} 1_{\{x_j=1\}}p_j e^{-\delta}\\ &= \sum_c \sum_{j\in J_c^+} (e^\delta - 1)e^{-\delta s}e^{c\delta} - \sum_c \sum_{j\in J_c^-} (e^\delta - 1)e^{-\delta s}e^{c\delta}.\end{aligned}$$

Note that

$$|J_c^+| = \begin{cases} |J_c^-| & \text{if } \alpha > 0 \text{ and } \ c \notin \{0, 1, \ldots, \alpha s - 1\}\\ |J_c^-| + 1 & \text{if } \alpha > 0 \text{ and } \ c \in \{0, 1, \ldots, \alpha s - 1\}\\ |J_c^-| & \text{if } \alpha = 0\\ |J_c^-| & \text{if } \alpha < 0 \text{ and } \ -c \notin \{1, 2, \ldots, -\alpha s\}\\ |J_c^-| - 1 & \text{if } \alpha < 0 \text{ and } \ -c \in \{1, 2, \ldots, -\alpha s\}. \end{cases}$$

Therefore,

$$E_{\sigma,x}v = \begin{cases} \sum_{j=0}^{\alpha s-1}(e^\delta - 1)e^{-\delta s}e^{j\delta} = e^{(\alpha-1)s\delta} - e^{-s\delta} \text{ if } \alpha > 0,\\ \\ \sum_{j=1}^{-\alpha s}(e^\delta - 1)e^{-\delta s}e^{-j\delta} = e^{(\alpha-1)s\delta} - e^{-s\delta} \text{ if } \alpha \leq 0. \end{cases} \qquad (9)$$

This completes the proof of equality (7).

The function $\alpha \mapsto e^{(\alpha-1)s\delta}$ is nonnegative and monotonic increasing in α, and $e^{(\alpha-1)s\delta} \geq e^{(\varepsilon-1)s\delta}1_{\{\alpha\geq\varepsilon\}}$. Therefore, equality (7) implies inequality (8), which completes the proof of the lemma. □

Let $\varepsilon > 0$ be sufficiently small and let i_ε be a sufficiently large positive integer so that

$$\sum_{i=i_\varepsilon+1}^{\infty} \frac{e}{i^{(1+\varepsilon)}} < \min(\varepsilon, 1 - e^{-\delta}). \qquad (10)$$

Let $\delta > 0$ be sufficiently small so that $1 - e^{-\delta} < \varepsilon$. Let s_i be the largest integer such that $e^{s_i\delta} \leq i^{1+\varepsilon}$ if $i > i_\varepsilon$, and $s_i = 1$ if $i \leq i_\varepsilon$. Note that for $i > i_\varepsilon$ we have $e^{s_i\delta} \leq i^{1+\varepsilon} \leq e^{s_i\delta}e^\delta$.

Let v_i be the above-defined function v at the end of the i-th epoch; equivalently, at the start of the $(i+1)$-th epoch.

We proceed with the definition of the strategy σ of Player 1. The strategy σ plays in the i-th epoch the strategy $\sigma_{s_i,\delta}$.

Let τ be a pure strategy of Player 2 and let $x^i = (x^i_1, \ldots, x^i_{s_i})$ be the sequence of actions of Player 2 in epoch i assuming no absorption, and set

$$\alpha_i = \begin{cases} -\sum_{j=1}^{s_i} x^i_j/s_i \text{ if } \ v_{i-1} = 0 \text{ and } \ i > i_\varepsilon, \\ 0 \text{ otherwise.} \end{cases} \tag{11}$$

Inequality (8) along with the definition of α_i implies that for $i \geq i_\varepsilon$,

$$\begin{aligned} E_{\sigma,x}(v_i - v_{i-1} \mid \mathcal{H}_i) \ &\geq \ e^{(\varepsilon-1)s_i\delta} 1_{\{\alpha_i \geq \varepsilon\}} - e^{-s_i\delta} \\ &\geq \ i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}} - \frac{e}{i^{1+\varepsilon}}, \end{aligned} \tag{12}$$

where $\mathcal{H}_i$ is the history of play up to the start of the i-th epoch, and the definition of σ implies that for $1 \leq i < i_\varepsilon$, we have $v_i = v_{i-1}$.

Define

$$Y_i = v_i - \sum_{k > \max(i, i_\varepsilon)}^{\infty} \frac{e}{k^{(1+\varepsilon)}}.$$

Then, by the definition of v_i and (10), $|Y_i| < 1$, and, as $Y_i - Y_{i-1} = v_i - v_{i-1}$ for $i < i_\varepsilon$ and $Y_i - Y_{i-1} = v_i - v_{i-1} + \frac{e}{i^{1+\varepsilon}}$ for $i \geq i_\varepsilon$, inequality (12) implies that

$$E_{\sigma,x}(Y_i - Y_{i-1} \mid \mathcal{H}_i) \geq i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}}. \tag{13}$$

Therefore, $(Y_i)_{i > i_\varepsilon}$ is a bounded submartingale and therefore converges a.e. (namely, with probability 1) to a limit Y_∞. As $v_i - Y_i \to_{i\to\infty} 0$, v_i converges to Y_∞ as i goes to infinity.

Note that

$$\sum_{j=1}^{s_i} (r^i_j - v_{i-1}) \geq \begin{cases} -\varepsilon s_i \geq -\varepsilon s_i - s_i 1_{\{\alpha_i \geq \varepsilon\}} & \text{if } \ v_{i-1} = -e^{-\delta} \\ 0 \geq -\varepsilon s_i - s_i 1_{\{\alpha_i \geq \varepsilon\}} & \text{if } \ v_{i-1} = 1 \\ -\alpha_i s_i \geq -\varepsilon s_i - s_i 1_{\{\alpha_i \geq \varepsilon\}} & \text{if } \ v_i = v_{i-1} = 0 \\ -s_i = -s_i 1_{\{v_i \neq v_{i-1}\}} & \text{if } \ v_i \neq v_{i-1} = 0. \end{cases} \tag{14}$$

Therefore,

$$\sum_{j=1}^{s_i} (r^i_j - v_{i-1}) \ \geq \ -\varepsilon s_i - s_i 1_{\{\alpha_i \geq \varepsilon\}} - s_i 1_{\{v_{i-1} \neq v_i\}}. \tag{15}$$

Summing these inequalities over $1 \leq i \leq n$ we deduce that

$$\sum_{i=1}^{n} \sum_{j=1}^{s_i} r^i_j \geq \sum_{i=1}^{n} s_i(v_{i-1} - \varepsilon) - \sum_{i=1}^{n} s_i 1_{\{\alpha_i \geq \varepsilon\}} - \sum_{i=1}^{n} s_i 1_{\{v_{i-1} \neq v_i\}}. \tag{16}$$

Note that $-1 < Y_i < 1$. Therefore $Y_i - Y_j < 2$. Taking the expectations in inequality (13), we deduce that $E_{\sigma,x}(Y_i - Y_{i-1}) \geq E_{\sigma,x}\, i^{\varepsilon^1-1} 1_{\{\alpha_i \geq \varepsilon\}}$. Summing these inequalities over all i such that $1 \leq i \leq n$, we deduce that

$$2 > E_{\sigma,x}(Y_n - Y_0) \geq E_{\sigma,x} \sum_{i=1}^{n} i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}}. \tag{17}$$

Thus, $2 \geq E_{\sigma,x} \sum_{i=1}^{\infty} i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}}$. Hence, $\sum_{i=1}^{\infty} i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}}$ is finite a.e.

Note that $S_n \geq \frac{n}{2} \ln n$ for all sufficiently large n and, as $e^{\delta s_i} \leq i^{1+\varepsilon}$, $s_i \leq \frac{1+\varepsilon}{\delta} \ln i \leq \frac{1+\varepsilon}{\delta} \ln n$ for $i \leq n$. Therefore, as $\sum_{i=1}^{n} 1_{\{v_i \neq v_{i-1}\}} \leq 1$,

$$\frac{1}{S_n} \sum_{i=1}^{n} s_i 1_{\{v_i \neq v_{i-1}\}} \leq \frac{s_n}{S_n} \rightarrow_{n\to\infty} 0, \tag{18}$$

and for all sufficiently large n and $i \leq n$, we have $\frac{s_i}{S_n} \leq \frac{2(1+\varepsilon)}{\delta} n^{-\varepsilon^2} n^{\varepsilon^2-1} \leq \frac{2(1+\varepsilon)}{\delta} n^{-\varepsilon^2} i^{\varepsilon^2-1}$. Hence,

$$\frac{1}{S_n} \sum_{i=1}^{n} s_i 1_{\{\alpha_i \geq \varepsilon\}} \leq \frac{1+\varepsilon}{\delta} n^{-\varepsilon^2} \sum_{i=1}^{n} i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}} \rightarrow_{n\to\infty} 0 \text{ a.e.} \tag{19}$$

As $v_i - \varepsilon \rightarrow_{i\to\infty} Y_\infty - \varepsilon$, $\frac{s_i}{S_n} \rightarrow_{n\to\infty} 0$ (for each fixed i), and $S_n = \sum_{i=1}^{n} s_i$, we have

$$\frac{1}{S_n} \sum_{i=1}^{n} s_i (v_{i-1} - \varepsilon) \rightarrow_{n\to\infty} Y_\infty - \varepsilon \text{ a.e.} \tag{20}$$

Therefore, using (16), (18), (19), and (20), we deduce that

$$\liminf_{n\to\infty} \frac{1}{S_n} \sum_{i=1}^{n} \sum_{j=1}^{s_i} r_j^i \geq Y_\infty - \varepsilon \text{ a.e.} \tag{21}$$

As $E_{\sigma,x} Y_\infty \geq Y_0 \geq v_0 - \varepsilon$ we conclude that (5) holds, i.e.,

$$E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n} \sum_{i=1}^{n} \sum_{j=1}^{s_i} r_j^i \geq v_0 - 2\varepsilon. \tag{22}$$

We proceed to prove (6). Let n_ε be a sufficiently large integer so that $\frac{2(1+\varepsilon)}{\delta} n_\varepsilon^{-\varepsilon^2} < \frac{\varepsilon}{2}$. Hence, $\frac{s_i}{S_n} \leq i^{\varepsilon^2-1} \varepsilon/2$ for every $n \geq n_\varepsilon$ and $i \leq n$. Then, using inequality (17), we have

$$E_{\sigma,x} \frac{1}{S_n} \sum_{i=1}^{n} s_i 1_{\{\alpha_i \geq \varepsilon\}} \leq E_{\sigma,x} \sum_{i=1}^{n} i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}} \varepsilon/2 \leq \varepsilon \quad \forall n \geq n_\varepsilon. \tag{23}$$

As $v_i \geq Y_i$ and $E_{\sigma,x} Y_i \geq Y_0 \geq v_0 - \varepsilon$,

$$E_{\sigma,x} \frac{1}{S_n} \sum_{i=1}^{n} s_i v_{i-1} \geq v_0 - \varepsilon. \tag{24}$$

As $v_i = v_{i-1}$ whenever v_{i-1} equals either 1 or $-e^{-\delta}$, $\sum_{i=1}^n 1_{\{v_i \neq v_{i-1}\}} \leq 1$. Hence, $\frac{1}{S_n}\sum_{i=1}^n s_i 1_{\{v_i \neq v_{i-1}\}} \leq s_n/S_n \leq \varepsilon$ for every $n \geq n_\varepsilon$ and, in particular,

$$E_{\sigma,x}\frac{1}{S_n}\sum_{i=1}^n s_i 1_{\{v_i \neq v_{i-1}\}} \leq \varepsilon \;\; \forall n \geq n_\varepsilon. \tag{25}$$

Taking expectation in inequality (16), and using inequalities (25), (24), and (23), we deduce that for $n \geq n_\varepsilon$ we have

$$E_{\sigma,x}\frac{1}{S_n}\sum_{i=1}^n \sum_{j=1}^{s_i} r_j^i \quad \geq \quad v_0 - \varepsilon - \varepsilon - \varepsilon - \varepsilon \geq v_0 - 4\varepsilon,$$

which completes the proof of (6).

5. Open problems

The main open problem is whether or not in any stochastic game each player has a finite-memory strategy that is ε-optimal.

In the remainder of this section we introduce several additional open problems. These open problems are of independent interest and a few of them may turn out to be building blocks toward the solution of the main open problem.

5.1. **Private versus public memory states.**

The ε-optimal two-memory strategy in our proof uses private memory states (i.e., states that are not observable by Player 2).

We say that the memory states m_t are *public* if they are observed by all players. For example, the memory states of the Blackwell and Ferguson [2] strategy in the Big Match, which are the possible differences between the number of odd and even guesses of Player 2, are public. So are the memory states of the Mertens and Neyman [6] ε-optimal strategies in a stochastic game, and so are the memory states of any memory-based strategy in which the memory update functions are deterministic. The memory states of the ε-optimal strategies that are introduced in [4] are private.

All the above-mentioned ε-optimal strategies are memory-based strategies with an infinite set of memory states. A generalization of the proof of [4, Theorem 6] shows that in the Big Match any finite-memory strategy whose memory states are public is worthless.

A natural question that arises is what is the minimal size of a public memory (as a function of t) that is needed for an ε-optimal strategy in a stochastic game. In order to state this problem formally, we introduce the concept of a public (f,γ)-memory strategy, where $f : \mathbb{N} \to \mathbb{N}$ is a nondecreasing function and $\gamma > 0$.

A public (f,γ)-memory strategy is a memory-based strategy σ whose memory states are public and such that for every strategy τ of Player 2, with probability $P_{\sigma,\tau}$ at least $1-\gamma$, $\mathbb{N} \ni m_t \leq f(t)$ for all t.

Note that we distinguish between a public finite-memory strategy and a mixed strategy that is a mixture of such strategies. In fact, a general mixing principle implies that in any stochastic game, any k-memory strategy (even if all memory states are private) is equivalent to a mixed strategy that is a mixture of (uncountably many) public k-memory strategies.

This principle follows from the following construction of a mixture of public k-memory strategies. Let σ be a k-memory strategy with memory states m_t, action function σ_α, and memory updating function σ_m. For any sequence of permutations of $[k] := \{1, \dots, k\}$, $\pi = (\pi_t)_{t=1}^\infty$, we define the public k-memory $\pi\sigma$ strategy that follows the strategy σ and that makes public the memory states renamed according to π.

Formally, the memory states $\overline{m}_t$ of $\pi\sigma$ are functions of the the memory states of σ: $\overline{m}_t = \pi_t(m_t) \in [k]$); its action function $(\pi\sigma)_\alpha$ is given by $(\pi\sigma)_\alpha(t, z_t, \overline{m}_t) = \sigma_\alpha((t, z_t, m_t)$; and its memory updating function $(\pi\sigma)_m$ is given by $(\pi\sigma)_m(t, z_t, \overline{m}_t, i_t, j_t) = \sigma_m(t, z_t, m_t, i_t, j_t)$.

The mixture of $\pi\sigma$, where the sequence of random permutations π_t, $t = 1, 2 \dots$, is a sequence of i.i.d. permutations of $[k]$ and each π_t is uniformly distributed over all $k!$ permutations, is equivalent to the k-memory strategy σ.

5.2. **Recall-based strategies.** The definitions in this section apply to a general stochastic game. A few of the open problems in this section concern some specific stochastic game.

A *recall-based strategy* is a memory-based strategy in which the memory states m_t is simply $z_{t-k_t}, i_{t-k_t}, j_{t-k_t}, \dots, z_{t-1}, i_{t-1}, j_{t-1}, z_t$, where $k_t < t$. As it is a memory-based strategy it follows that $k_{t+1} \leq k_t + 1$. A *k-recall strategy* is a recall-based strategy where the recall size k_t equals k. A *finite-recall strategy* is a k-recall strategy for some fixed finite k.

In a recall-based strategy the memory states are public and the memory update function is deterministic. Therefore, it follows from [4, Theorem 6] that in the Big Match, Player 1 has no worthy strategy that is a finite-recall strategy.

A natural question that arises is what is the minimal recall (as a function of t) that is needed for an ε-optimal strategy in a stochastic game. In order to state this problem formally, we introduce the concept of f-recall strategies, where $f : \mathbb{N} \to \mathbb{N}$ is a nondecreasing function with $f(t) < t$ and $f(t+1) \leq f(t) + 1$.

An *f-recall strategy* is a memory-based strategy in which the memory state m_t is $z_{t-f(t)}, i_{t-f(t)}, j_{t-f(t)}, \dots, z_{t-1}, i_{t-1}, j_{t-1}, z_t$.

The question that arises is what are the functions f for which there is an f-recall strategy that is ε-optimal. The question applies to a general stochastic game as well as to the special case of the Big Match.

It is worthwhile to note that the ε-optimal strategy in the Big Match that is introduced in the present paper is an f-recall strategy with $f(t) \leq \frac{K \log t}{\varepsilon}$ for some positive constant K.

Acknowledgment. We are grateful to Elon Kohlberg for his helpful comments on an earlier draft.

References

[1] M. Amitai. Stochastic games with automata. Master's thesis, Hebrew University, Jerusalem, 1989. (in Hebrew).

[2] D. Blackwell and T. S. Ferguson. The big match. *The Annals of Mathematical Statistics*, 39(1):159–163, 1968.

[3] D. Gillette. Stochastic games with zero-stop probabilities. In *Contributions to the Theory of Games III*, volume 39 of *Ann. Math. Studies*, pages 179–187. Princeton University Press, 1957.

[4] K. A. Hansen, R. Ibsen-Jensen, and M. Koucký. The big match in small space (extended abstract). In M. Gairing and R. Savani, editors, *Proceedings of 9th International Symposium on Algorithmic Game Theory, SAGT 2016*, volume 9928 of *Lecture Notes in Computer Science*, pages 64–76. Springer, 2016.

[5] E. Kohlberg. Repeated games with absorbing states. *The Annals of Statistics*, 2(4):724–738, 1974.

[6] J.-F. Mertens and A. Neyman. Stochastic games. *Int. J. of Game Theory*, 10(2):53–66, 1981.

[7] S. Sorin. *A First Course on Zero-Sum Repeated Games.* Springer, 2002.

Kristoffer Arnsfelt Hansen, Aarhus University
E-mail address: `arnsfelt@cs.au.dk`

Rasmus Ibsen-Jensen, IST Austria
E-mail address: `ribsen@ist.ac.at`

Abraham Neyman, Hebrew University
E-mail address: `aneyman@huji.ac.il`