Adaptive game playing using multiplicative weights

Yoav Freund Robert E. Schapire
AT&T Labs
180 Park Avenue
Florham Park, NJ 07932-0971
{yoav, schapire}@research.att.com
http://www.research.att.com/~{yoav, schapire}

November 20, 1997

1 Introduction

We study the problem of learning to play a repeated game. Let M be a matrix. On each of a series of rounds, one player chooses a row i and the other chooses a column j. The selected entry $\mathrm{M}(i, j)$ is the loss suffered by the row player. We study play of the game from the row player's perspective, and therefore leave the column player's loss or utility unspecified.

A simple goal for the row player is to suffer loss which is no worse than the value of the game M (if viewed as a zero-sum game). Such a goal may be appropriate when it is expected that the opposing column player's goal is to maximize the loss of the row player (so that the game is in fact zero-sum). In this case, the row player can do no better than to play using a minmax mixed strategy which can be computed using linear programming, provided that the entire matrix M is known ahead of time, and provided that the matrix is not too large. This approach has a number of potential drawbacks. For instance,

- M may be unknown;
- M may be so large that computing a minmax strategy using linear programming is infeasible; or
- the column player may not be truly adversarial and may behave in a manner that admits loss significantly smaller than the game value.

Overcoming these difficulties in the one-shot game is hopeless. In repeated play, however, one can hope to learn to play well against the particular opponent that is being faced.

Algorithms of this type were first proposed by Hannan [20] and Blackwell [3], and later algorithms were proposed by Foster and Vohra [15, 13, 12]. These algorithms have the property that the loss of the row player in repeated play is guaranteed to come close to the minimum loss achievable with respect to the sequence of plays taken by the column player.

In this paper, we present a simple algorithm for solving this problem, and give a simple analysis of the algorithm. The bounds we obtain are *not* asymptotic and hold for any finite number of rounds. The algorithm and its analysis are based directly on the "on-line prediction" methods of Littlestone and Warmuth [24].

The analysis of this algorithm yields a new (as far as we know) and simple proof of von Neumann's minmax theorem, as well as a provable method of approximately solving a game. We also give more refined variants of the algorithm for this purpose, and we show that one of these is optimal in a very strong sense.

The paper is organized as follows. In Section 2 we define the mathematical setup and notation. In Section 3 we introduce the basic multiplicative weights algorithm whose average performance is guaranteed to be almost as good as that of the best fixed mixed strategy. In Section 4 we outline the relationship between our work and some of the extensive existing work on the use of multiplicative weights algorithms for on-line prediction. In Section 5 we show how the algorithm can be used to give a simple proof of Von-Neumann's min-max theorem. In Section 6 we give a version of the algorithm whose distributions are guaranteed to converge to an optimal mixed strategy. We note the possible application of this algorithm to solving linear programming problems and reference other work that have used multiplicative weights to this end. Finally, in Section 7 we show that the convergence rate of the second version of the algorithm is asymptotically optimal.

2 Playing repeated games

We consider non-collaborative two-person games in normal form. The game is defined by a matrix M with n rows and m columns. There are two players called the row player and column player. To play the game, the row player chooses a row i, and, simultaneously, the column player chooses a column j. The selected entry $\mathrm{M}(i, j)$ is the *loss* suffered by the row player. The column player's loss or utility is unspecified.

For the sake of simplicity, throughout this paper, we assume that all the entries of the matrix M are in the range $[0, 1]$. Simple scaling can be used to get similar results for general bounded ranges. Also, we restrict ourselves to the case where the number of choices available to each player is finite. However, most of the results translate with very mild additional assumptions to cases in which the number of choices is infinite. For a discussion of infinite matrix games see, for instance, Chapter 2 in Ferguson [11].

Following standard terminology, we refer to the choice of a specific row or column as a *pure strategy* and to a distribution over rows or columns as a *mixed strategy*. We use $\mathbf{P}$ to denote a mixed strategy of the row player, and $\mathbf{Q}$ to denote a mixed strategy of the column player. We use $\mathbf{P}(i)$ to denote the probability that $\mathbf{P}$ associates with the row i, and we write $\mathrm{M}(\mathbf{P}, \mathbf{Q}) = \mathbf{P}^{\mathrm{T}}\mathrm{M}\mathbf{Q}$ to denote the expected loss (of the row player) when the two mixed strategies are used. In addition, we write $\mathrm{M}(\mathbf{P}, j)$ and $\mathrm{M}(i, \mathbf{Q})$ to denote the expected loss when one side uses a pure strategy and the other a mixed strategy. Although these quantities denote *expected* losses, we will usually refer to them simply as losses.

If we assume that the loss of the row player is the gain of the column player, we can think about the game as a zero-sum game. Under such an interpretation we use $\mathbf{P}^*$ and $\mathbf{Q}^*$ to denote optimal mixed strategies for M, and $v = \mathrm{M}(\mathbf{P}^*, \mathbf{Q}^*)$ to denote the value of the game.

The main subject of this paper is an algorithm for adaptively selecting mixed strategies. The algorithm is used to choose a mixed strategy for one of the players in the context of *repeated play*. We usually associate the algorithm with the row player. To emphasize the roles of the two players in our context, we sometimes refer to the row and column players as the *learner* and the *environment*, respectively. An instance of repeated play is a sequence of *rounds* of interactions between the learner and the environment. The game matrix M used in the interactions is fixed but is unknown to the learner. The learner only knows the number of choices that it has, i.e., the number of rows. On round $t = 1, \ldots, T$:

1. the learner chooses mixed strategy $\mathbf{P}_t$;

2. the environment chooses mixed strategy $\mathbf{Q}_t$ (which may be chosen with knowledge of $\mathbf{P}_t$)

3. the learner is permitted to observe the loss $\mathrm{M}(i, \mathbf{Q}_t)$ for each row i; this is the loss it would have suffered had it played using pure strategy i;

4. the learner suffers loss $\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)$.

The basic goal of the learner is to minimize its total loss $\sum_{t=1}^{T} \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)$. If the environment is maximally adversarial then a related goal is to approximate the optimal mixed row strategy $\mathbf{P}^*$. However, in more benign environments, the goal may be to suffer the minimum loss possible, which may be much better than the value of the game.

Finally, in what follows, we find it useful to measure the distance between two distributions $\mathbf{P}_1$ and $\mathbf{P}_2$ using the *Kullback-Leibler divergence*, also called the *relative entropy*, which is defined to be

$$\mathrm{RE}\left(\mathbf{P}_1 \;\|\; \mathbf{P}_2\right) \doteq \sum_{i=1}^{n} \mathbf{P}_1(i) \ln\left(\frac{\mathbf{P}_1(i)}{\mathbf{P}_2(i)}\right).$$

As is well known, the relative entropy is a measure of discrepancy between distributions in that it is non-negative and is equal to zero if and only if $\mathbf{P}_1 = \mathbf{P}_2$. For real numbers $p_1, p_2 \in [0, 1]$, we use the shorthand $\mathrm{RE}\left(p_1 \;\|\; p_2\right)$ to denote the relative entropy between Bernoulli distributions with parameters p_1 and p_2, i.e.,

$$\mathrm{RE}\left(p_1 \;\|\; p_2\right) \doteq p_1 \ln\left(\frac{p_1}{p_2}\right) + (1 - p_1) \ln\left(\frac{1 - p_1}{1 - p_2}\right).$$

3 The basic algorithm

We now describe our basic algorithm for repeated play, which we call MW for "multiplicative weights." This algorithm is a direct generalization of Littlestone and Warmuth's "weighted majority algorithm" [24], which was discovered independently by Fudenberg and Levine [17].

The learning algorithm MW starts with some initial mixed strategy $\mathbf{P}_1$ which it uses for the first round of the game. After each round t, the learner computes a new mixed strategy $\mathbf{P}_{t+1}$ by a simple multiplicative rule:

$$\mathbf{P}_{t+1}(i) = \mathbf{P}_t(i) \frac{\beta^{\mathbf{M}(i, \mathbf{Q}_t)}}{Z_t}$$

where Z_t is a normalization factor:

$$Z_t = \sum_{i=1}^{n} \mathbf{P}_t(i) \beta^{\mathbf{M}(i, \mathbf{Q}_t)},$$

and $\beta \in [0, 1)$ is a parameter of the algorithm.

The main theorem concerning this algorithm is the following:

Theorem 1 *For any matrix $\mathbf{M}$ with n rows and entries in $[0, 1]$, and for any sequence of mixed strategies $\mathbf{Q}_1, \ldots, \mathbf{Q}_T$ played by the environment, the sequence of mixed strategies $\mathbf{P}_1, \ldots, \mathbf{P}_T$ produced by algorithm* MW *satisfies:*

$$\sum_{t=1}^{T} \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \leq \min_{\mathbf{P}} \left[a_\beta \sum_{t=1}^{T} \mathbf{M}(\mathbf{P}, \mathbf{Q}_t) + c_\beta \mathrm{RE}\left(\mathbf{P} \;\|\; \mathbf{P}_1\right) \right]$$

where

$$a_\beta = \frac{\ln(1/\beta)}{1 - \beta} \qquad c_\beta = \frac{1}{1 - \beta}.$$

Our proof uses a kind of "amortized analysis" in which relative entropy is used as a "potential" function. This method of analysis for on-line learning algorithms is due to Kivinen and Warmuth [23]. The heart of the proof is in the following lemma, which bounds the change in potential before and after a single round.

Lemma 2 *For any iteration t where* MW *is used with parameter β, and for any mixed strategy $\tilde{\mathbf{P}}$,*

$$\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_t\right) \leq \left(\ln\frac{1}{\beta}\right)\mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) + \ln\left(1 - (1-\beta)\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right).$$

Proof: The proof of the lemma can be summarized by the following sequence of inequalities:

$$\begin{aligned}
&\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_t\right) \\
&= \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\tilde{\mathbf{P}}(i)}{\mathbf{P}_{t+1}(i)} - \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\tilde{\mathbf{P}}(i)}{\mathbf{P}_t(i)} &(1)\\
&= \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\mathbf{P}_t(i)}{\mathbf{P}_{t+1}(i)} &(2)\\
&= \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{Z_t}{\beta^{\mathbf{M}(i,\mathbf{Q}_t)}} &(3)\\
&= \left(\ln\frac{1}{\beta}\right) \sum_{i=1}^{n} \tilde{\mathbf{P}}(i)\mathbf{M}(i, \mathbf{Q}_t) + \ln Z_t &(4)\\
&\leq \left(\ln\frac{1}{\beta}\right) \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) + \ln\left[\sum_{i=1}^{n} \mathbf{P}_t(i)\left(1 - (1-\beta)\mathbf{M}(i, \mathbf{Q}_t)\right)\right] &(5)\\
&= \left(\ln\frac{1}{\beta}\right)\mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) + \ln\left(1 - (1-\beta)\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right).
\end{aligned}$$

Line (1) follows from the definition of relative entropy. Line (3) follows from the update rule of MWand line (4) follows by simple algebra. Finally, line (5) follows from the definition of Z_t combined with the fact that, by convexity, $\beta^x \leq 1 - (1-\beta)x$ for $\beta \geq 0$ and $x \in [0, 1]$. ■

Proof of Theorem 1: Let $\tilde{\mathbf{P}}$ be any mixed row strategy. We first simplify the last term in the inequality of Lemma 2 by using the fact that $\ln(1-x) \leq -x$ for any $x < 1$ which implies that

$$\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_t\right) \leq \left(\ln\frac{1}{\beta}\right)\mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) - (1-\beta)\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)$$

Summing this inequality over $t = 1, \ldots, T$ we get

$$\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{T+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_1\right) \leq \left(\ln\frac{1}{\beta}\right) \sum_{t=1}^{T} \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) - (1-\beta) \sum_{t=1}^{T} \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t).$$

Noting that $\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{T+1}\right) \geq 0$, rearranging the inequality and noting that $\tilde{\mathbf{P}}$ was chosen arbitrarily gives the statement of the theorem. ■.

In order to use MW, we need to choose the initial distribution $\mathbf{P}_1$ and the parameter β. We start with the choice of $\mathbf{P}_1$. In general, the closer $\mathbf{P}_1$ is to a good mixed strategy $\tilde{\mathbf{P}}$, the better the bound on the total loss MW. However, even if we have no prior knowledge about the good mixed strategies, we can achieve reasonable performance by using the uniform distribution over the rows as the initial strategy. This gives us a performance bound that holds uniformly for all games with n rows:

Corollary 3 *If* MW *is used with $\mathbf{P}_1$ set to the uniform distribution then its total loss is bounded by*

$$\sum_{t=1}^{T} \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \leq a_\beta \min_{\mathbf{P}} \sum_{t=1}^{T} \mathbf{M}(\mathbf{P}, \mathbf{Q}_t) + c_\beta \ln n$$

where a_β and c_β are as defined in Theorem 1.

Proof: If $\mathbf{P}_1(i) = 1/n$ for all i then RE $(\mathbf{P} \;\|\; \mathbf{P}_1) \leq \ln n$ for all $\mathbf{P}$. ■

Next we discuss the choice of the parameter β. As β approaches 1, a_β approaches 1 from above while c_β increases to infinity. On the other hand, if we fix β and let the number of rounds T increase, the second term $c_\beta \ln n$ becomes negligible (since it is fixed) relative to T. Thus, by choosing β as a function of T which approaches 1 for $T \to \infty$, the learner can ensure that its average per-trial loss will not be much worse than the loss of the best strategy. This is formalized in the following corollary:

Corollary 4 *Under the conditions of Theorem 1 and with β set to*

$$\frac{1}{1+\sqrt{\frac{2\ln n}{T}}},$$

the average per-trial loss suffered by the learner is

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}_t,\mathbf{Q}_t) \leq \min_{\mathbf{P}}\frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P},\mathbf{Q}_t) + \Delta_{T,n}$$

where

$$\Delta_{T,n} = \sqrt{\frac{2\ln n}{T}} + \frac{\ln n}{T} = O\left(\sqrt{\frac{\ln n}{T}}\right).$$

Proof: It can be shown that $-\ln\beta \leq (1-\beta^2)/(2\beta)$ for $\beta \in (0,1]$. Applying this approximation and the given choice of β yields the result. ■

Since $\Delta_{T,n} \to 0$ as $T \to \infty$, we see that the amount by which the average per-trial loss of the learner exceeds that of the best mixed strategy can be made arbitrarily small for large T.

Note that in the analysis we made no assumption about the strategy used by the environment. Theorem 1 guarantees that its cumulative loss is not much larger than that of *any* fixed mixed strategy. As shown below, this implies that the loss cannot be much larger than the game value. However, if the environment is non-adversarial, there might be a better row strategy, in which case the algorithm is guaranteed to be almost as good as this better strategy.

Corollary 5 *Under the conditions of Corollary 4,*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}_t,\mathbf{Q}_t) \leq v + \Delta_{T,n}$$

where v is the value of the game **M**.

Proof: Let $\mathbf{P}^*$ be a minmax strategy for M so that for all column strategies $\mathbf{Q}$, $\mathrm{M}(\mathbf{P}^*,\mathbf{Q}) \leq v$. Then, by Corollary 4,

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{M}(\mathbf{P}_t,\mathbf{Q}_t) \leq \frac{1}{T}\sum_{t=1}^{T}\mathrm{M}(\mathbf{P}^*,\mathbf{Q}_t) + \Delta_{T,n} \leq v + \Delta_{T,n}.$$

■

3.1 Convergence with probability one

Suppose that the mixed strategies that are generated by MW are used to select one of the rows at each iteration. From Theorem 1 and Corollary 4 we know that the expected per-iteration loss of MW approaches the optimal achievable value for any fixed strategy as $T \to \infty$. However, we might want a stronger assurance of the performance of MW; for example, we would like to know that the *actual* per-iteration loss is, with high probability, close to the expected value. As the following lemma shows, the per-trial loss of any algorithm for the repeated game is, with high probability, at most $O(1/\sqrt{T})$ away from the expected value. The only required game property is that the game matrix elements are all in [0, 1].

Lemma 6 *Let the players of a matrix game use any pair of methods for choosing their mixed strategies on iteration t based on past game events. Let $\mathbf{P}_t$ and $\mathbf{Q}_t$ denote the mixed strategies used by the players on iteration t and let $\mathbf{M}(i_t, j_t)$ denote the actual game outcome on iteration t that is chosen at random according to $\mathbf{P}_t$ and $\mathbf{Q}_t$. Then, for every $\epsilon > 0$,*

$$\Pr\left[\frac{1}{T}\left|\sum_{t=1}^{T}(\mathbf{M}(i_t, j_t) - \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t))\right| > \epsilon\right] \leq 2\exp\left(-\frac{1}{2}T\epsilon^2\right) ,$$

where probability is taken with respect to the random choice of rows $i_1, \ldots, i_T$ and columns $j_1, \ldots, j_T$.

Proof: The proof follows directly from a theorem proved by Hoeffding [22] about the convergence of a sum of bounded-step martingales which is commonly called "Azuma's lemma." The sequence of random variables $Y_t = \mathrm{M}(i_t, j_t) - \mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)$ is a martingale difference sequence. As the entries of M are bounded in [0, 1] we have that $|Y_t| \leq 1$. Thus we can directly apply Azuma's Lemma and get that, for any $a > 0$

$$\Pr\left[\left|\sum_{t=1}^{T} Y_t\right| > a\right] \leq 2\exp\left(-\frac{a^2}{2T}\right) .$$

Substituting $a = \epsilon T$ we get the statement of the lemma. ■

If we want to have an algorithm whose performance will converge to the optimal performance we need the value of β to approach 1 as the length of the sequence increases. One way of doing this, which we describe here, is to have the row player divide the time sequence into "epochs." In each epoch, the row player restarts the algorithm MW (resetting all the row distribution to the uniform distribution) and uses a different value of β which is tuned according to the length of the epoch. We show that such a procedure can guarantee, almost surely, that the long term per-iteration loss is at most the expected loss of any fixed mixed strategy.

We denote the length of the kth epoch by T_k and the value of β used for that epoch by β_k. One choice of epochs that gives convergence with probability one is the following:

$$T_k = k^2, \quad \beta_k = \frac{1}{1 + \sqrt{\frac{2\ln n}{k^2}}} . \tag{6}$$

The convergence properties of this strategy are given in the following theorem:

Theorem 7 *Suppose the repeated game is continued for an unbounded number of rounds. Let $\mathbf{P}_t$ be chosen according to the method of epochs with the parameters described in Equation (6), and let i_t be chosen at random according to $\mathbf{P}_t$. Let the environment choose j_t as an arbitrary stochastic function of past plays. Then, for every $\epsilon > 0$, with probability one with respect to the randomization used by both players, the following inequality holds for all but a finite number of values of T:*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(i_t, j_t) \leq \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}, j_t) + \epsilon.$$

Proof: For each epoch k we select the accuracy parameter $\epsilon_k = 2\sqrt{\ln k}/k$. We denote the sequence of iterations that constitute the k'th epoch by S_k. We call the kth epoch "good" if the average per trial loss for that epoch is within ϵ_k from its expected value, i.e., if

$$\sum_{t \in S_k} \mathbf{M}(i_t, j_t) \leq \sum_{t \in S_k} \mathbf{M}(\mathbf{P}_t, j_t) + T_k \epsilon_k. \tag{7}$$

From Lemma 6 (where we define Q_t to be the mixed strategy which gives probability one to j_t), we get that the probability that the kth epoch is bad is bounded by

$$2 \exp\left(-\frac{1}{2} T_k \epsilon_k^2\right) = \frac{2}{k^2} .$$

The sum of this bound over all k from 1 to ∞ is finite. Thus, by the Borel-Cantelli lemma, we know that with probability one all but a finite number of epochs are good. Thus for the sake of computing the average loss for $T \to \infty$ we can ignore the influence of the bad epochs.

We now use Corollary 4 to bound the expected total loss. We apply this corollary in the case that $\mathbf{Q}_t$ is again defined to be the mixed strategy which gives probability one to j_t. We have from the corollary:

$$\sum_{t \in S_k} \mathbf{M}(\mathbf{P}_t, j_t) \leq \min_{\mathbf{P}} \sum_{t \in S_k} \mathbf{M}(\mathbf{P}, j_t) + \sqrt{2T_k \ln n} + \ln n . \tag{8}$$

Combining Equations (7) and (8) we find that if the kth epoch is good then, for any distribution $\tilde{\mathbf{P}}$ over the actions of the algorithm

$$\begin{aligned}
\sum_{t \in S_k} \mathbf{M}(i_t, j_t) &\leq \sum_{t \in S_k} \mathbf{M}(\tilde{\mathbf{P}}, j_t) + \sqrt{2T_k \ln n} + \ln n + T_k \epsilon_k \\
&\leq \sum_{t \in S_k} \mathbf{M}(\tilde{\mathbf{P}}, j_t) + k\sqrt{2 \ln n} + \ln n + 2k\sqrt{\ln k} .
\end{aligned}$$

Thus the total loss over the first m epochs (ignoring the finite number of bad iterations whose influence is negligible) is bounded by

$$\begin{aligned}
\sum_{t \in S_1 \cup \ldots \cup S_m} \mathbf{M}(i_t, j_t) &\leq \sum_{t \in S_1 \cup \ldots \cup S_m} \mathbf{M}(\tilde{\mathbf{P}}, j_t) + \sum_{k=1}^{m} \left[k\sqrt{2 \ln n} + \ln n + 2k\sqrt{\ln k}\right] \\
&\leq \sum_{t \in S_1 \cup \ldots \cup S_m} \mathbf{M}(\tilde{\mathbf{P}}, j_t) + m^2\sqrt{\ln m}\left[\sqrt{2 \ln n} + \ln n + 2\right] .
\end{aligned}$$

As the total number of rounds in the first m epochs is $\sum_{k=1}^{m} k^2 = O(m^3)$ we find that, after dividing both sides by the number of rounds, the error term decreases to zero. ■

4 Relation to on-line learning

One interesting use of game theory is in the context of predictive decision making (see, for instance, Blackwell and Girshick [4] or Ferguson [11]). On-Line decision making can be viewed as a repeated game between a decision maker and nature. The entry $\mathbf{M}(i, t)$ represents the loss of (or negative utility for) the prediction algorithm if it chooses action i at time t. The goal of the algorithm is to adaptively generate distributions over actions so that its expected cumulative loss will not be much worse than the cumulative loss it would have incurred had it been able to choose a *single fixed distribution* with prior knowledge of the whole sequence of columns.

This is a non-standard framework for analyzing on-line decision algorithms in that one makes no statistical assumptions regarding the relationship between actions and their losses. The only assumption is that there exists some fixed mixed strategy (distribution over actions) whose expected performance is nontrivial. This approach was previously described in one of our earlier papers [16]; the current paper expands and refines the results given there.

The algorithm MW was originally suggested by Littlestone and Warmuth [24] and (in a somewhat more sophisticated form) by Vovk [29] in the context of on-line prediction. The algorithm was also discovered independently by Fudenberg and Levine [17]. Research on the use of the multiplicative weights algorithm for on-line prediction is extensive and on-going, and it is out of the scope of this paper to give a complete review of it. However, we try to sketch some of the main connections between the work described in this paper and this expanding line of research.

The on-line prediction framework is a refinement of the decision theoretic framework described above. Here the prediction algorithm generates distributions over *predictions*, nature chooses an *outcome* and the loss incurred by the prediction algorithm is a known *loss function* which maps action/outcome pairs to real values. This framework restricts the choices that can be made by nature because once the predictions have been fixed, the only loss columns that are possible are those that correspond to possible outcomes. This is the reason that for various loss functions one can prove better bounds than in the less structured context of on-line decision making. The approach is closely related to work by Dawid [9], Foster [14] and Vovk [29].

One loss function that has received particular attention is the log loss function. Here the prediction is assumed to be a distribution P over some domain X, the outcome is an element from the domain $x \in X$, and the loss is $-\log P(x)$. This loss has several important interpretations which connect it to likelihood analysis and to coding theory. Note that as the probability of an element can be arbitrarily small, the loss can be arbitrarily high. On-Line algorithms for making predictions in this case have been extensively studied in information theory under the name *universal compression of individual sequences* [31, 27]. In particular, a well-known result is that the multiplicative weights algorithm, with β set to $1/e$ is a near-optimal algorithm in this context. It is also interesting to note that this version of the multiplicative weights algorithm is equivalent to the Bayes prediction rule, where the generated distributions over the rows are equal to the Bayesian posterior distributions. On the other hand, this equivalence holds *only* for the log-loss; for other loss functions there is no simple relationship between the multiplicative weights algorithm and the Bayesian algorithm.

Cover and Ordentlich [7, 6] and later Helmbold et al. [21] extended the log-loss analysis to the design of algorithms for "universal portfolios." There is an extensive literature on on-line prediction with other specific loss functions. For example, for work on prediction loss, see Feder, Merhav and Gutman [10], Cesa-Bianchi et al. [5] and for work on more general families of loss functions see Vovk [28] and Kivinen and Warmuth [23].

Another extension of the on-line decision problem that is worth mentioning here is making decisions when the feedback given is a single entry of the game matrix. In other words, we assume that after the row player has chosen a distribution over the rows, a single row is chosen at random according to the distribution. The row player suffers the loss associated with the selected row and the column chosen by its opponent, and the game repeats. The goal of the row player is the same as before—to minimize its expected average loss over a sequence of repeated games. Clearly, the goal is much harder here since only a single entry of the matrix is revealed on each round. Auer et al. [2] study this model in detail and show that a variant of the multiplicative weights algorithm converges to the performance of the best row distribution in repeated play.

5 Proof of the minmax theorem

Corollary 5 shows that the loss of MW can never exceed the value of the game M by more than $\Delta_{T,n}$. More interestingly, Corollary 4 can be used to derive a very simple proof of von Neumann's minmax theorem. To prove this theorem, we need to show that

$$\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}, \mathbf{Q}) \leq \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \mathbf{Q}). \tag{9}$$

(Proving that $\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}, \mathbf{Q}) \geq \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \mathbf{Q})$ is relatively straightforward and so is omitted.)

Suppose that we run algorithm MW against a maximally adversarial environment which always chooses strategies which maximize the learner's loss. That is, on each round t, the environment chooses

$$\mathbf{Q}_t = \arg\max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}_t, \mathbf{Q}). \tag{10}$$

Let $\overline{\mathbf{P}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t$ and $\overline{\mathbf{Q}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{Q}_t$. Clearly, $\overline{\mathbf{P}}$ and $\overline{\mathbf{Q}}$ are probability distributions.

Then we have:

$$\begin{aligned}
\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q} &\leq \max_{\mathbf{Q}} \overline{\mathbf{P}}^{\mathrm{T}}\mathbf{M}\mathbf{Q} \\
&= \max_{\mathbf{Q}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q} && \text{by definition of } \overline{\mathbf{P}} \\
&\leq \frac{1}{T}\sum_{t=1}^{T} \max_{\mathbf{Q}} \mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q} \\
&= \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q}_t && \text{by definition of } \mathbf{Q}_t \\
&\leq \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q}_t + \Delta_{T,n} && \text{by Corollary 4} \\
&= \min_{\mathbf{P}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\overline{\mathbf{Q}} + \Delta_{T,n} && \text{by definition of } \overline{\mathbf{Q}} \\
&\leq \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q} + \Delta_{T,n}.
\end{aligned}$$

Since $\Delta_{T,n}$ can be made arbitrarily close to zero, this proves Eq. (9) and the minmax theorem.

6 Approximately solving a game

Aside from yielding a proof for a famous theorem that by now has many proofs, the preceding derivation shows that algorithm MW can be used to find an approximate minmax or maxmin strategy. Finding these "optimal" strategies is called *solving* the game M.

We give three methods for solving a game using exponential weights. In Section 6.1 we show how one can use the average of the generated row distributions over T iterations as an approximate solution for the game. This method sets T and β as a function of the desired accuracy before starting the iterative process.

In Section 6.2 we show that if an upper bound u on the value of the game is known ahead of time then one can use a variant of MW that generates a sequence of row distributions such that the expected loss of the tth distribution approaches u. Finally, in Section 6.3 we describe a related adaptive method that generates a sparse approximate solution for the column distribution. At the end of the paper, in Section 7, we show that the convergence rate of the two last methods is asymptotically optimal.

6.1 Using the average of the row distributions

Skipping the first inequality of the sequence of equalities and inequalities at the end of Section 5, we see that

$$\max_{\mathbf{Q}} \mathbf{M}(\overline{\mathbf{P}}, \mathbf{Q}) \quad \leq \quad \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \mathbf{Q}) + \Delta_{T,n} = v + \Delta_{T,n}.$$

Thus, the vector $\overline{\mathbf{P}}$ is an approximate minmax strategy in the sense that for all column strategies $\mathbf{Q}$, $\mathrm{M}(\overline{\mathbf{P}}, \mathbf{Q})$ does not exceed the game value v by more than $\Delta_{T,n}$. Since $\Delta_{T,n}$ can be made arbitrarily small, this approximation can be made arbitrarily tight.

Similarly, ignoring the last inequality of this derivation, we have that

$$\min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \overline{\mathbf{Q}}) \geq v - \Delta_{T,n}$$

so $\overline{\mathbf{Q}}$ also is an approximate maxmin strategy. Furthermore, it can be shown that a column strategy $\mathbf{Q}_t$ satisfying Eq. (10) can always be chosen to be a pure strategy (i.e., a mixed strategy concentrated on a single column of M). Therefore, the approximate maxmin strategy $\overline{\mathbf{Q}}$ has the additional favorable property of being *sparse* in the sense that at most T of its entries will be nonzero.

6.2 Using the final row distribution

In the analysis presented so far we have shown that the *average* of the strategies used by MW converges to an optimal strategy. Now we show that if the row player knows an upper bound u on the value of the game v then it can use a variant of MW to generate a sequence of mixed strategies that approach a strategy which achieves loss u.[1] To do that we have the algorithm select a different value of β for each round of the game. If the expected loss on the tth iteration $\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)$ is less than u, then the row player does not change the mixed strategy, because, in a sense, it is "good enough." However, if $\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t) \geq u$ then the row player uses MW with parameter

$$\beta_t = \frac{u(1 - \mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t))}{(1 - u)\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)} .$$

We call this algorithm vMW (the "v" stands for "variable"). For this algorithm, as the following theorem shows, the distance between $\mathbf{P}_t$ and any mixed strategy that achieves u decreases by an amount that is a function of the divergence between $\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)$ and u.

Theorem 8 *Let $\tilde{\mathbf{P}}$ be any mixed strategy for the rows such that $\max_{\mathbf{Q}} \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}) \leq u$. Then on any iteration of algorithm* vMW *in which $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \geq u$ the relative entropy between $\tilde{\mathbf{P}}$ and $\mathbf{P}_{t+1}$ satisfies*

$$\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) \leq \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_t\right) - \mathrm{RE}\left(u \;\|\; \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) .$$

Proof: Note that when $u \leq \mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)$ we get that $\beta_t \leq 1$. Combining this observation with the definition of $\tilde{\mathbf{P}}$ and the statement of Lemma 2 we get that

$$\begin{aligned} &\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_t\right) \\ &\leq \quad \mathrm{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) \ln(1/\beta_t) + \ln\left(1 - (1 - \beta_t)\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) \qquad (11) \\ &\leq \quad u \ln(1/\beta_t) + \ln\left(1 - (1 - \beta_t)\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) . \end{aligned}$$

[1]If no such upper bound is known, one can use the standard trick of solving the larger game matrix

$$\left(\begin{array}{cc} M & \mathbf{0} \\ \mathbf{0} & -M^{\mathrm{T}} \end{array} \right) ,$$

whose value is always zero.

The choice of β_t was chosen to minimize the last expression. Plugging the given choice of β_t into this last expression we get the statement of the theorem. ■

Suppose $\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t) \geq u$ for all t. Then the main inequality of this theorem can be applied repeatedly yielding the bound

$$\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{T+1}\right) \leq \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_1\right) - \sum_{t=1}^{T} \mathrm{RE}\left(u \;\|\; \mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)\right).$$

Since relative entropy is nonnegative, and since the inequality holds for all T, we have

$$\sum_{t=1}^{\infty} \mathrm{RE}\left(u \;\|\; \mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) \leq \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_1\right). \tag{12}$$

Assuming that $\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_1\right)$ is finite (as it will be, for example, if $\mathbf{P}_1$ is uniform), this inequality implies, for instance, that $\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)$ can exceed $u + \epsilon$ at most finitely often for any $\epsilon > 0$. More specifically, we can prove the following:

Corollary 9 *Suppose that* vMW *is used to play a game* $\mathbf{M}$ *whose value is known to be at most* u*. Suppose also that we choose* $\mathbf{P}_1$ *to be the uniform distribution. Then for any sequence of column strategies* $\mathbf{Q}_1, \mathbf{Q}_2, \ldots$*, the number of rounds on which the loss* $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \geq u + \epsilon$ *is at most*

$$\frac{\ln n}{\mathrm{RE}\left(u \;\|\; u + \epsilon\right)}.$$

Proof: Since rounds on which $\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t) < u$ are effectively ignored by vMW, we assume without loss of generality that $\mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t) \geq u$ for all rounds t. Let $S = \{t : \mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t) \geq u + \epsilon\}$ be the set of rounds for which the loss is at least $u + \epsilon$, and let $\mathbf{P}^*$ be a minmax strategy. By Eq. (12), we have that

$$\begin{aligned}
\sum_{t \in S} \mathrm{RE}\left(u \;\|\; u + \epsilon\right) &\leq \sum_{t \in S} \mathrm{RE}\left(u \;\|\; \mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) \\
&\leq \sum_{t=1}^{\infty} \mathrm{RE}\left(u \;\|\; \mathrm{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) \\
&\leq \mathrm{RE}\left(\mathbf{P}^* \;\|\; \mathbf{P}_1\right) \leq \ln n.
\end{aligned}$$

Therefore,

$$|S| \leq \frac{\ln n}{\mathrm{RE}\left(u \;\|\; u + \epsilon\right)}.$$

■

In Section 7, we show that this dependence on n, u and ϵ cannot be improved by any constant factor.

6.3 Convergence of a column distribution

When β is fixed, we showed in Section 6.1 that the average $\overline{\mathbf{Q}}$ of the $\mathbf{Q}_t$'s is an approximate solution of the game, i.e., that there are no rows i for which $\mathrm{M}(i, \overline{\mathbf{Q}})$ is less than $v - \Delta_{T,n}$. For the algorithm described above in which β_t varies, we can derive a more refined bound of this kind for a weighted mixture of the $\mathbf{Q}_t$'s.

Theorem 10 *Assume that on every iteration of algorithm* vMW, *we have that* $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \geq u$. *Let*

$$\hat{\mathbf{Q}} = \frac{\sum_{t=1}^{T} \mathbf{Q}_t \ln(1/\beta_t)}{\sum_{t=1}^{T} \ln(1/\beta_t)}.$$

Then

$$\sum_{i:\mathbf{M}(i,\hat{\mathbf{Q}}) \leq u} \mathbf{P}_1(i) \leq \exp\left(-\sum_{t=1}^{T} \mathrm{RE}\left(u \;\|\; \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right)\right).$$

Proof: If $\mathbf{M}(\tilde{\mathbf{P}}, \hat{\mathbf{Q}}) \leq u$, then, combining Eq. 11 for $t = 1, \ldots, T$, we have

$$\begin{aligned}
\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{T+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_1\right) &\leq \sum_{t=1}^{T} \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) \ln(1/\beta_t) + \sum_{t=1}^{T} \ln\left(1 - (1-\beta_t)\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) \\
&= \mathbf{M}(\tilde{\mathbf{P}}, \hat{\mathbf{Q}}) \sum_{t=1}^{T} \ln(1/\beta_t) + \sum_{t=1}^{T} \ln\left(1 - (1-\beta_t)\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) \\
&\leq u \cdot \sum_{t=1}^{T} \ln(1/\beta_t) + \sum_{t=1}^{T} \ln\left(1 - (1-\beta_t)\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) \\
&= -\sum_{t=1}^{T} \mathrm{RE}\left(u \;\|\; \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right)
\end{aligned}$$

for our choice of β_t. In particular, if i is a row for which $\mathbf{M}(i, \hat{\mathbf{Q}}) \leq u$, then, setting $\tilde{\mathbf{P}}$ to the associated pure strategy, we get

$$\ln\left(\frac{\mathbf{P}_1(i)}{\mathbf{P}_{T+1}(i)}\right) \leq -\sum_{t=1}^{T} \mathrm{RE}\left(u \;\|\; \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right)$$

so

$$\begin{aligned}
\sum_{i:\mathbf{M}(i,\hat{\mathbf{Q}}) \leq u} \mathbf{P}_1(i) &\leq \sum_{i:\mathbf{M}(i,\hat{\mathbf{Q}}) \leq u} \mathbf{P}_{T+1}(i) \exp\left(-\sum_{t=1}^{T} \mathrm{RE}\left(u \;\|\; \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right)\right) \\
&\leq \exp\left(-\sum_{t=1}^{T} \mathrm{RE}\left(u \;\|\; \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right)\right)
\end{aligned}$$

since $\mathbf{P}_{T+1}$ is a distribution. ■

Thus, if $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)$ is bounded away from u, the fraction of rows i (as measured by $\mathbf{P}_1$) for which $\mathbf{M}(i, \hat{\mathbf{Q}}) \leq u$ drops to zero exponentially fast. This will be the case, for instance, if Eq. (10) holds and $u \leq v - \epsilon$ for some $\epsilon > 0$ where v is the value of $\mathbf{M}$.

Thus a single application of the exponential weights algorithm yields approximate solutions for both the column and row players. The solution for the row player consists of the multiplicative weights, while the solution for the column player consists of the distribution on the observed columns as described in Theorem 10.

Given a game matrix $\mathbf{M}$, we have a choice of whether to solve $\mathbf{M}$ or $-\mathbf{M}^{\mathrm{T}}$. One natural choice would be to choose the orientation which minimizes the number of rows. In a related paper [16], we studied the relationship between solving $\mathbf{M}$ or $-\mathbf{M}^{\mathrm{T}}$ using the multiplicative weights algorithm in the context of machine learning. In that context, the solution for game matrix $\mathbf{M}$ is related to the on-line prediction problem described in Section 4, while the "dual" solution for $-\mathbf{M}^{\mathrm{T}}$ corresponds to a method of learning called "boosting."

6.4 Application to linear programming

It is well known that any linear programming problem can be reduced to the problem of solving a game (see, for instance, Owen [25, Theorem III.2.6]). Thus, the algorithms we have presented for approximately solving a game can be applied more generally for approximate linear programming.

Similar and closely related methods of approximately solving linear programming problems have previously appeared, for instance, in the work of Young [30], Grigoriadis and Khachiyan [18, 19] and Plotkin, Shmoys and Tardos [26].

Although, in principle, our algorithms are applicable to general linear programming problems, they are best suited to problems of a particular form. Specifically, they may be most appropriate for the setting we have described of approximately solving a game when an oracle is available for choosing columns of the matrix on every round. When such an oracle is available, our algorithm can be applied even when the number of columns of the matrix is very large or even infinite, a setting that is clearly infeasible for some of the other, more traditional linear programming algorithms. Solving linear programming problems in the presence of such an oracle was also studied by Young [30] and Plotkin, Shmoys and Tardos [26]. See also our earlier paper [16] for detailed examples of problems arising naturally in the field of machine learning with exactly these characteristics.

7 Optimality of the convergence rate

In Corollary 9, we showed that using the algorithm vMW starting from the uniform distribution over the rows guarantees that the number of times that $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)$ can exceed $u+\epsilon$ is bounded by $(\ln n)/\mathrm{RE}\left(u \;\|\; u+\epsilon\right)$ where u is a known upper bound on the value of the game $\mathbf{M}$. In this section, we show that this dependence of the rate of convergence on n, u and ϵ is optimal in the sense that no adaptive game-playing algorithm can beat this bound even by a constant factor.

This result is formalized by the following theorem:

Theorem 11 *Let $0 < u < u+\epsilon < 1$, and let n be a sufficiently large integer. Then for any adaptive game-playing algorithm A, there exists a game matrix $\mathbf{M}$ of n rows and a sequence of column strategies such that:*

1. *the value of game $\mathbf{M}$ is at most u; and*
2. *the loss $\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)$ suffered by A on each round $t = 1, \ldots, T$ is at least $u+\epsilon$, where*

$$T = \left\lfloor \frac{\ln n - 5\ln\ln n}{\mathrm{RE}\left(u \;\|\; u+\epsilon\right)} \right\rfloor \geq \frac{(1-o(1))\ln n}{\mathrm{RE}\left(u \;\|\; u+\epsilon\right)}.$$

Proof: The proof uses a probabilistic argument to show that for any algorithm, there exists a matrix (and sequence of column strategies) with the properties stated in the theorem. That is, for the purposes of the proof, we imagine choosing the matrix $\mathbf{M}$ at random according to an appropriate distribution, and we show that the stated properties hold with strictly positive probability, implying that there must exist at least one matrix for which they hold.

Let $r = u+\epsilon$. The random matrix $\mathbf{M}$ has n rows and T columns, and is chosen by selecting each entry $\mathbf{M}(i,j)$ independently to be 1 with probability r, and 0 with probability $1-r$. On round t, the row player (algorithm A) chooses a row distribution $\mathbf{P}_t$, and, for the purposes of our construction, we assume that the column player responds with column t. That is, the column strategy $\mathbf{Q}_t$ chosen on round t is concentrated on column t.

Given this random construction, we need to show that properties 1 and 2 hold with positive probability for n sufficiently large.

We begin with property 2. On round t, the row player chooses a distribution $\mathbf{P}_t$, and the column player responds with column t. We require that the loss $\mathbf{M}(\mathbf{P}_t, t)$ be at least $r = u + \epsilon$. Since the matrix $\mathbf{M}$ is chosen at random, we need a lower bound on the probability that $\mathbf{M}(\mathbf{P}_t, t) \geq r$. Moreover, because the row player has sole control over the choice of $\mathbf{P}_t$, we need a lower bound on this probability which is independent of $\mathbf{P}_t$. To this end, we prove the following lemma:

Lemma 12 *For every $r \in (0, 1)$, there exists a number $B_r > 0$ with the following property: Let n be any positive integer, and let $\alpha_1, \ldots, \alpha_n$ be nonnegative numbers such that $\sum_{i=1}^n \alpha_i = 1$. Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with* $\Pr[X_i = 1] = r$ *and* $\Pr[X_i = 0] = 1 - r$. *Then*

$$\Pr\left[\sum_{i=1}^n \alpha_i X_i \geq r\right] \geq B_r > 0.$$

Proof: See appendix. ■

To apply the lemma, let $\alpha_i = \mathbf{P}_t(i)$ and let $X_i = \mathbf{M}(i, t)$. Then the lemma implies that

$$\Pr\left[\mathbf{M}(\mathbf{P}_t, t) \geq r\right] \geq B_r$$

where B_r is a positive number which depends on r but which is independent of n and $\mathbf{P}_t$. It follows that

$$\Pr\left[\forall t : \mathbf{M}(\mathbf{P}_t, t) \geq r\right] \geq B_r^T.$$

In other words, property 2 holds with probability at least B_r^T.

We next show that property 1 fails to hold with probability strictly smaller than B_r^T so that both properties must hold simultaneously with positive probability.

Define the *weight* of row i, denoted $W(i)$, to be the fraction of 1's in the row: $W(i) = \sum_{j=1}^T \mathbf{M}(i, j)/T$. We say that a row is *light* if $W(i) \leq u - 1/T$. Let $\mathbf{P}'$ be a row distribution which is uniform over the light rows and zero on the heavy rows. We will show that, with high probability, $\max_j \mathbf{M}(\mathbf{P}', j) \leq u$, implying an upper bound of u on the value of game $\mathbf{M}$.

Let λ denote the probability that a given row i is light; this will be the same probability for all rows. Let n' be the number of light rows.

We show first that $n' \geq \lambda n/2$ with high probability. The expected value of n' is λn. Using a form of Chernoff bounds proved by Angluin and Valiant [1], we have that

$$\Pr\left[n' < \lambda n/2\right] \leq \exp(-\lambda n/8). \tag{13}$$

We next upper bound the probability that $\mathbf{M}(\mathbf{P}', j)$ exceeds u for any column j. Conditional on i being a light row, the probability that $\mathbf{M}(i, j) = 1$ is at most $u - 1/T$. Moreover, if i_1 and i_2 are distinct rows, then $\mathbf{M}(i_1, j)$ and $\mathbf{M}(i_2, j)$ are independent, even if we condition on both being light rows. Therefore, applying Hoeffding's inequality [22] to column j and the n' light rows, we have that, for all j,

$$\Pr\left[\mathbf{M}(\mathbf{P}', j) > u \mid n'\right] \leq e^{-2n'/T^2}.$$

Thus,

$$\Pr\left[\max_j \mathbf{M}(\mathbf{P}', j) > u \mid n'\right] \leq T e^{-2n'/T^2}$$

and so

$$\Pr\left[\max_j \mathbf{M}(\mathbf{P}', j) > u \;\mid\; n' \geq \lambda n/2\right] \leq Te^{-\lambda n/T^2}.$$

Combined with Eq. (13), this implies that

$$\Pr\left[\max_j \mathbf{M}(\mathbf{P}', j) > u\right] \leq e^{-\lambda n/2} + Te^{-\lambda n/T^2} \leq (T+1)e^{-\lambda n/T^2}$$

for $T \geq 3$.

Therefore, the probability that either of properties 1 or 2 fails to hold is at most

$$(T+1)e^{-\lambda n/T^2} + 1 - B_r^T.$$

If this quantity is strictly less than 1, then there must exist at least one matrix $\mathbf{M}$ for which both properties 1 and 2 hold. This will be the case if and only if

$$\lambda > \frac{T^2}{n}(T\ln(1/B_r) + \ln(T+1)). \tag{14}$$

Therefore, to complete the proof, we need only prove Eq. (14) by lower bounding λ.

We have that

$$\begin{aligned}
\lambda &= \Pr\left[W(i) \cdot T \leq Tu - 1\right] \\
&\geq \Pr\left[W(i) \cdot T = \lfloor Tu - 1 \rfloor\right] \\
&\geq \frac{1}{T+1} \exp\left(-T \cdot \mathrm{RE}\left(\lfloor Tu - 1 \rfloor / T \;\|\; u + \epsilon\right)\right) \\
&\geq \frac{1}{T+1} \exp\left(-T \cdot \mathrm{RE}\left(u - 2/T \;\|\; u + \epsilon\right)\right).
\end{aligned}$$

The second inequality follows from Cover and Thomas [8, Theorem 12.1.4].

By straightforward algebra,

$$\begin{aligned}
T \cdot \mathrm{RE}\left(u - 2/T \;\|\; u + \epsilon\right) &= T \cdot \left(\mathrm{RE}\left(u \;\|\; u + \epsilon\right) - \mathrm{RE}\left(u \;\|\; u - 2/T\right)\right) \\
&\quad + 2\ln\left(\frac{1 - u + 2/T}{1 - u - \epsilon} \cdot \frac{u + \epsilon}{u - 2/T}\right) \\
&\leq T \cdot \mathrm{RE}\left(u \;\|\; u + \epsilon\right) + C
\end{aligned}$$

for T sufficiently large, where C is the constant

$$C = 2\ln\left(\frac{1 - u/2}{1 - u - \epsilon} \cdot \frac{u + \epsilon}{u/2}\right).$$

Thus,

$$\lambda \geq \frac{e^{-C}}{T+1} \exp\left(-T \cdot \mathrm{RE}\left(u \;\|\; u + \epsilon\right)\right)$$

and therefore, Eq. (14) holds if

$$T \cdot \mathrm{RE}\left(u \;\|\; u + \epsilon\right) < \ln n - C - \ln\left(T^2(T+1)(T\ln(1/B_r) + \ln(T+1))\right).$$

By our choice of T, we have that the left hand side of this inequality is at most $\ln n - 5\ln\ln n$, and the right hand side is $\ln n - (4 + o(1))\ln\ln n$. Therefore, the inequality holds for n sufficiently large. ∎

Acknowledgments

We are especially grateful to Neal Young for many helpful discussions, and for bringing much of the relevant literature to our attention. Dean Foster and Rakesh Vohra also helped us to locate relevant literature. Thanks finally to Colin Mallows and Joel Spencer for their help in proving Lemma 12.

References

[1] Dana Angluin and Leslie G. Valiant. Fast probabilistic algorithms for Hamiltonian circuits and matchings. *Journal of Computer and System Sciences*, 18(2):155–193, April 1979.

[2] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *36th Annual Symposium on Foundations of Computer Science*, pages 322–331, 1995.

[3] David Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, Spring 1956.

[4] David Blackwell and M.A. Girshick. *Theory of games and statistical decisions*. dover, 1954.

[5] Nicolò Cesa-Bianchi, Yoav Freund, David P. Helmbold, David Haussler, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on the Theory of Computing*, pages 382–391, 1993. To appear, *Journal of the Association for Computing Machinery*.

[6] T. M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, March 1996.

[7] Thomas M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, January 1991.

[8] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[9] A. P. Dawid. Statistical theory: The prequential approach. *Journal of the Royal Statistical Society, Series A*, 147:278–292, 1984.

[10] M. Feder, N. Merhav, and M. Gutman. Universal prediction of individual sequences. *IEEE Transactions on Information Theory*, 38:1258–1270, 1992.

[11] Thomas S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, 1967.

[12] Dean Foster and Rakesh Vohra. Regret in on-line decision making. unpublished manuscript, 1996.

[13] Dean Foster and Rakesh V. Vohra. Asymptotic calibration. unpublished manuscript, 1995.

[14] Dean P. Foster. Prediction in the worst case. *The Annals of Statistics*, 19(2):1084–1090, 1991.

[15] Dean P. Foster and Rakesh V. Vohra. A randomization rule for selecting forecasts. *Operations Research*, 41(4):704–709, July–August 1993.

[16] Yoav Freund and Robert E. Schapire. Game theory, on-line prediction and boosting. In *Proceedings of the Ninth Annual Conference on Computational Learning Theory*, pages 325–332, 1996.

[17] Drew Fudenberg and David K. Levine. Consistency and cautious fictitious play. *Journal of Economic Dynamics and Control*, 19:1065–1089, 1995.

[18] Michael D. Grigoriadis and Leonid G. Khachiyan. Approximate solution of matrix games in parallel. Technical Report 91-73, DIMACS, July 1991.

[19] Michael D. Grigoriadis and Leonid G. Khachiyan. A sublinear-time randomized approximation algorithm for matrix games. *Operations Research Letters*, 18(2):53–58, Sep 1995.

[20] James Hannan. Approximation to Bayes risk in repeated play. In M. Dresher, A. W. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*, volume III, pages 97–139. Princeton University Press, 1957.

[21] David P. Helmbold, Robert E. Schapire, Yoram Singer, and Manfred K. Warmuth. On-line portfolio selection using multiplicative updates. In *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 243–251, 1996.

[22] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[23] Jyrki Kivinen and Manfred K. Warmuth. Additive versus exponentiated gradient updates for linear prediction. In *stoc95*, pages 209–218, 1995. See also technical report UCSC-CRL-94-16, University of California, Santa Cruz, Computer Research Laboratory.

[24] Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

[25] Guillermo Owen. *Game Theory*. Academic Press, second edition, 1982.

[26] Serge A. Plotkin, David B. Shmoys, and Éva Tardos. Fast approximation algorithms for fractional packing and covering problems. *Mathematics of Operations Research*, 20(2):257–301, May 1995.

[27] Y. M. Shtar'kov. Universal sequential coding of single messages. *Problems of information Transmission (translated from Russian)*, 23:175–186, July-September 1987.

[28] V. G. Vovk. A game of prediction with expert advice. In *Proceedings of the Eighth Annual Conference on Computational Learning Theory*, 1995.

[29] Volodimir G. Vovk. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, pages 371–383, 1990.

[30] Neal Young. Randomized rounding without solving the linear program. In *Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 170–178, 1995.

[31] Jacob Ziv. Coding theorems for individual sequences. *IEEE Transactions on Information Theory*, 24(4):405–412, July 1978.

A Proof of Lemma 12

Let

$$Y = \frac{\sum_{i=1}^{n} \alpha_i X_i - r}{\sqrt{\sum_{i=1}^{n} \alpha_i^2}}.$$

Our goal is to derive a lower bound on $\Pr[Y \geq 0]$. Let $s = r(1-r)$. It can be easily verified that $\mathrm{E}Y = 0$ and $\mathrm{Var}\, Y = s$. In addition, by Hoeffding's inequality [22], it can be shown that, for all $\epsilon > 0$,

$$\Pr[Y \geq \epsilon] \leq e^{-2\epsilon^2} \tag{15}$$

and

$$\Pr[Y \leq -\epsilon] \leq e^{-2\epsilon^2}.$$

For $x \in \mathbb{R}$, let $D(x) = \Pr[Y = x]$. Throughout this proof, we use $\sum_x$ to denote summation over a finite set of x's which includes all x for which $D(x) > 0$. Restricted summations (such as $\sum_{x>0}$) are defined analogously.

Let $d > 0$ be any number. We define the following quantities:

$$\begin{aligned} G &= \sum_{0<x<d} D(x) \\ R &= -\sum_{-d<x<0} xD(x) \\ E_1 &= \sum_{x \geq d} xD(x) \\ E_2 &= \sum_{x \leq -d} x^2 D(x) \\ E_3 &= \sum_{x \geq d} x^2 D(x). \end{aligned}$$

We prove the lemma by deriving a lower bound on $G \leq \Pr[Y \geq 0]$.

The expected value of Y is:

$$\begin{aligned} 0 = \mathrm{E}Y &= \sum_x xD(x) \\ &= \sum_{x \leq -d} xD(x) + \sum_{-d<x<0} xD(x) + \sum_{0<x<d} xD(x) + \sum_{x \geq d} xD(x) \\ &\leq 0 - R + dG + E_1. \end{aligned}$$

Thus,

$$R \leq dG + E_1. \tag{16}$$

Next, we have that

$$\begin{aligned} s = \mathrm{Var}\, Y &= \sum_x x^2 D(x) \\ &= \sum_{x \leq -d} x^2 D(x) + \sum_{-d<x<0} x^2 D(x) + \sum_{0<x<d} x^2 D(x) + \sum_{x \geq d} x^2 D(x) \\ &\leq E_2 + dR + d^2 G + E_3. \end{aligned}$$

Combined with Eq. (16), it follows that

$$s \leq 2d^2 G + dE_1 + E_2 + E_3. \tag{17}$$

We next upper bound E_1, E_2 and E_3. This will allow us to immediately lower bound G using Eq. (17). To bound E_1, note that

$$dE_1 = d\sum_{x \geq d} xD(x) \leq \sum_{x \geq d} x^2 D(x) = E_3. \tag{18}$$

To bound E_3, let $d = y_0 < y_1 < \cdots < y_m$ be a sequence of numbers such that if $D(x) > 0$ and $x \geq d$ then $x = y_i$ for some i. In other words, every $x \geq d$ with positive probability is represented by some y_i. Let $S(y) = \sum_{x \geq y} D(x)$. By Eq. (15), $S(y) \leq e^{-2y^2}$ for $y > 0$. We can compute E_3 as follows:

$$\begin{aligned}
E_3 = \sum_{x \geq d} x^2 D(x) &= \sum_{i=0}^{m} y_i^2 D(y_i) \\
&= y_0^2 \sum_{j=0}^{m} D(y_j) + \sum_{i=0}^{m-1} \left[(y_{i+1}^2 - y_i^2) \sum_{j=i+1}^{m} D(y_j) \right] \\
&= y_0^2 S(y_0) + \sum_{i=0}^{m-1} (y_{i+1}^2 - y_i^2) S(y_{i+1}) \\
&\leq d^2 e^{-2d^2} + \sum_{i=0}^{m-1} (y_{i+1}^2 - y_i^2) e^{-2y_{i+1}^2}.
\end{aligned}$$

To bound the summation, note that

$$\begin{aligned}
\sum_{i=0}^{m-1} (y_{i+1}^2 - y_i^2) e^{-2y_{i+1}^2} &= \sum_{i=0}^{m-1} \int_{y_i}^{y_{i+1}} 2x e^{-2y_{i+1}^2} \, dx \\
&\leq \sum_{i=0}^{m-1} \int_{y_i}^{y_{i+1}} 2x e^{-2x^2} \, dx \\
&= \int_{y_0}^{y_m} 2x e^{-2x^2} \, dx \\
&= \tfrac{1}{2}(e^{-2y_0^2} - e^{-2y_m^2}) \leq \tfrac{1}{2} e^{-2d^2}.
\end{aligned}$$

Thus, $E_3 \leq (d^2 + 1/2)e^{-2d^2}$. A bound on E_2 follows by symmetry.

Combining with Eqs. (17) and (18), we have

$$s \leq 2d^2 G + 3(d^2 + 1/2)e^{-2d^2}$$

and so

$$\Pr[Y \geq 0] \geq G \geq \frac{s - 3(d^2 + 1/2)e^{-2d^2}}{2d^2}.$$

Since this holds for all d, we have that $\Pr[Y \geq 0] \geq B_r$ where

$$B_r = \sup_{d > 0} \frac{s - 3(d^2 + 1/2)e^{-2d^2}}{2d^2}$$

and $s = r(1-r)$. This number is clearly positive since the numerator of the inside expression can be made positive by choosing d sufficiently large. (For instance, it can be shown that this expression is positive when we set $d = \sqrt{1/s}$.) ■