Asymptotic Convergence in Online Learning with Unbounded Delays

Scott Garrabrant and **Nate Soares** and **Jessica Taylor**
Machine Intelligence Research Institute
{scott,nate,jessica}@intelligence.org

Abstract

We study the problem of predicting the results of computations that are too expensive to run, via the observation of the results of smaller computations. We model this as an online learning problem with delayed feedback, where the length of the delay is unbounded, which we study mainly in a stochastic setting. We show that in this setting, consistency is not possible in general, and that optimal forecasters might not have average regret going to zero. However, it is still possible to give algorithms that converge asymptotically to Bayes-optimal predictions, by evaluating forecasters on specific sparse independent subsequences of their predictions. We give an algorithm that does this, which converges asymptotically on good behavior, and give very weak bounds on how long it takes to converge. We then relate our results back to the problem of predicting large computations in a deterministic setting.

1 Introduction

We study the problem of predicting the results of computations that are too large to evaluate, given observation of the results of running many smaller computations. For example, we might have a physics simulator and want to predict the final location of a ball in a large environment, after observing many simulated runs of small environments.

When predicting the outputs of computations so large that they cannot be evaluated, generating training data requires a bit of creativity. Intuitively, one potential solution is this: Given enough computing resources to evaluate "medium-sized" computations, we could train a learner by showing it many runs of small computations, and having it learn to predict the medium-sized ones, in a way that generalizes well. Then we could feed it runs of many medium-sized computations and have it predict large ones. This is an online learning problem, where the learner observes the results of more and more expensive computations, and predicts the behavior of computations that are much more difficult to evaluate than anything it has observed so far.

The standard online learning setting, in which the learner predicts an outcome in a sequence after observing all previous outcomes, does not capture this problem, because delays between prediction and observation are the key feature. [1], [2], and others have studied online learning with delayed feedback, but they assume that delays are bounded, whereas in our setting the delays necessarily grow ever-larger. In this paper, we propose an algorithm `EvOp` for online learning with unbounded delays. `EvOp` not a practical algorithm; it is only a first step towards modeling the problem of predicting large computations as an online learning problem.

Predicting a sequence generated by arbitrary computations is intractable in general. Consider, for instance, the bitstring that tells which Turing machines

Research supported by the Machine Intelligence Research Institute (intelligence.org).

halt. However, the problem is not hopeless, either: Consider the bitstring where the nth digit is a 1 if and only if the 10^nth digit in the decimal expansion of π is a 7. This is an online learning problem with ever-growing delays where a learner should be able to perform quite well. A learner that attempts to predict the behavior of computations in full generality will encounter some subsequences that it cannot predict, but it will encounter others that are highly regular, and it should be able to identify those and predict them well.

Consider, for instance, the bitstring that interleaves information about which Turing machines halt with the 10^nth digits of π. Intuitively, a good predictor should identify the second subsequence, and assign extreme probabilities whenever it has the computing resources to compute the digit, and roughly 10% probability otherwise, in lieu of other information about the digit. However, it's not clear how to formalize this intuition: What does it mean for a forecaster to have no relevant information about a digit of π that it knows how to compute? What are the "correct" probabilities a bounded reasoner should assign to deterministic facts that it lacks the resources to compute?

In this paper, we sidestep those questions, by analyzing the problem in a stochastic setting. This lets us study the problem of picking out patterns in subsequences in the face of unbounded delays, in a setting where the "correct" probabilities that a predictor should be assigning are well-defined. In Section 5 we relate our findings back to the deterministic setting, making use of "algorithmic randomness" as described by, e.g., [3].

We propose an algorithm `EvOp` with the property that, on any subsequence for which an expert that it consults predicts the true probabilities, it converges to optimal behavior on that subsequence. We show that regret and average regret are poor measures of performance in this setting, by demonstrating that in environments with unbounded delays between prediction and feedback, optimal predictors can fail to have average regret going to zero. `EvOp` works around these difficulties by comparing forecasters on sparse subsequences of their predictions; this means that, while we can put bounds on how long it takes `EvOp` to converge, the bounds are very, very weak. Furthermore, `EvOp` is only guaranteed to converge to good behavior on subsequences when it has access to optimal experts; we leave it to future work to give a variant that can match the behavior of the best available expert even if it is non-optimal.

In Section 2 we define the problem of online learning with unbounded delays. In Section 3 we show that consistency is impossible and discuss other difficulties. In Section 4 we define `EvOp`, prove that it converges to Bayes-optimal behavior on any subsequence for which some expert makes Bayes-optimal predictions, and provide very weak bounds on how long convergence takes. In Section 5 we relate these results back to the deterministic setting. Section 6 concludes.

1.1 Related Work

An early example of online sequence learning using expert advice is [4]; much work has been done since then to understand how to perform well relative to a given set of forecasters [5, 6, 7]. [8] improve performance of online learning algorithms assuming some structure in the environment, while maintaining worst-case guarantees. [9] study the case with a potentially unbounded number of experts.

Most work in online learning has focused on the case where feedback is immediate. [10] study online prediction with less rigid feedback schemes, proving only weak performance bounds. [11] show that running experts on sub-sampled sequences can give better bounds, for the case with bounded feedback delay. In the widely studied bandit setting [12], some attention has been given to learning with bounded delays [13, 1]. There have been some attempts to work with unbounded feedback delays [14, 15, 16], with either strong assumptions on the target function or with weak performance bounds. A review, and a very general framework for online learning with arbitrary (but bounded) feedback delay is given by [2].

Online learning with delayed feedback has applications in domains such as webpage prefetching, since the prediction algorithm has to make some prefetching deci-

sions before learning whether a previously fetched page ended up being requested by the user [17]. The idea of learning from computations with delay has seen some use in parallel computation, e.g., distributed stochastic optimization where computations of gradients may take longer in some nodes [18, 19].

Outside the field of online learning, our work has interesting parallels in the field of mathematical logic. [20] and [21] study the problem of assigning probabilities to sentences in logic while respecting certain relationships between them, a practice that dates back to [22]. Because sentences in mathematical logic are expressive enough to make claims about the behavior of computations (such as "this computation will use less memory than that one"), their work can be seen as a different approach to the problems we discuss in this paper.

2 The Unbounded Delay Model

Let $\mathcal{X}$ be a set of possible outcomes and $\mathcal{Y}$ be a set of possible predictions, where $\mathcal{Y}$ is a convex subset of $\mathbb{R}^n$ for some n. Let $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a loss function measuring the difference between them, which is strongly convex (with strong convexity constant ρ) and Lipschitz (with Lipschitz constant κ). Roughly speaking, the environment will stochastically produce an infinite sequence of outcomes x_i, and an infinite sequence of observations o_i, where each o_i contains information about finitely many x_n. Formally, for each $i = 1, 2, \ldots$, let $o_i : \mathbb{N} \to \mathcal{X}$ be a finite-domain partial function from indices to outcomes; in other words, o_i is a set of (n, x) "feedback" pairs such that each n appears in at most one pair. We write $o_i(n)$ for the value of x associated with n, which is feedback about the outcome x_n, and which may be undefined. If $o_i(n)$ is defined, we say that o_i reveals x_n.

Formally, we write $\mathbf{X}_i$ for the random variable representing the ith output and $\mathbf{O}_i$ for the random variable representing the ith observation. We define the *true environment* P to be a joint distribution over the $\mathbf{X}_i$ and the $\mathbf{O}_i$, such that if $o_i(n) = x_n$ then $P(\mathbf{O}_i = o_i \wedge \mathbf{X}_n \neq x_n) = 0$, which means that all $o_i(n)$ which are defined agree on the value of x_n. We omit the random variables if we can do so unambiguously, writing, e.g., $P(x_n \mid o_i)$.

Note that there may exist n such that $o_i(n)$ is not defined for any i, in which case the forecaster will never observe x_n. We write $o_{\prec i}$ for the list of observations up to time i, and $o_{\prec i}(n)$ for the value of x_n if any observation in $o_{\prec i}$ reveals it.

We consider learning algorithms that make use of some set $\mathcal{F}$ of forecasters.

Definition 1. *A **forecaster** is a partial function f which takes as input n observations $o_{\prec n}$ and might produce a prediction $y_n \in \mathcal{Y}$, interpreted as a prediction of x_n.*

Because some outcomes may never be observed, and because forecasters are partial (and so may abstain from making predictions on certain subsequences of the outcomes), we will compare forecasters only on subsequences on which both are defined.

Definition 2. *A **subsequence** s of the outcomes is a monotonic strictly increasing list of natural numbers $s_1 s_2 \ldots$. We write $|s|$ for the length of s, which may be ∞. A forecaster f is **defined on** s if it outputs a prediction for all elements s_i of s, i.e., if, for all $i \leq |s|$, $y_{s_i} := f(o_{\prec s_i})$ is defined.*

We assume that at least one $f \in \mathcal{F}$ is defined everywhere. It may seem prohibitively expensive to evaluate $f(o_{\prec s_i})$ if s_i is large. For example, consider the subsequence $s = 1, 10, 100, \ldots$; f only predicts $x_{10^{10}}$ after making 10^{10} observations, despite the fact that $x_{10^{10}}$ is the eleventh element in the subsequence. However, there is no requirement that observations contain lots of feedback: $o_{\prec s_i}$ might not reveal very much, even if s_i is large.

The goal of a forecaster is to minimize its loss $\sum_{i=1}^{n} \mathcal{L}(x_{s_i}, y_{s_i})$, for $n \geq 1$. Two forecasters can be compared by comparing their total loss.

Definition 3. *Given a forecaster f defined on a subsequence s of length at least n, let*

$$\mathcal{F}_s \coloneqq \{f' \in \mathcal{F} \mid f' \text{ is defined on } s\}. \tag{1}$$

*Then the **regret** of f (on s, through n) is*

$$R_s^n(f) \coloneqq \max_{f' \in \mathcal{F}_s} \sum_{t=i}^{n} \mathcal{L}\left(x_{s_i}, f(o_{\prec s_i})\right) - \sum_{i=1}^{n} \mathcal{L}\left(x_{s_i}, f'(o_{\prec s_i})\right). \tag{2}$$

*f is **consistent** (with respect to $\mathcal{F}_s$) if its average expected regret goes to zero, that is, if*

$$\lim_{n\to\infty} \mathbb{E}[R_s^n(f)]/n = 0. \tag{3}$$

In our setting, consistency is too strong a guarantee to ask for, as we will see in Section 3. Instead, we present an algorithm `EvOp` with the property that, whenever there is a forecaster $f \in \mathcal{F}$ that is Bayes-optimal on some subsequence, `EvOp` eventually learns to predict optimally on that subsequence.

Definition 4. *A forecaster f is **Bayes-optimal** (for the true environment, in its domain) if:*

1. *Everything f predicts is almost surely eventually revealed. That is, if $f(o_{\prec n})$ is defined, then with probability 1 there is some N such that $o_N(n)$ is defined.*
2. *f minimizes expected loss against the true environment whenever it makes a prediction. That is, if $y_n \coloneqq f(o_{\prec n})$ is defined, then $y_n = \arg\min_y \mathbb{E}[\mathcal{L}(x_n, y) \mid o_{\prec n}]$.*

We will occasionally refer to a Bayes-optimal f as simply "optimal".

The main result of our paper is this: Whenever there is an optimal forecaster $f \in F$ defined on s, our algorithm `EvOp` converges to optimal behavior on s.

Theorem 1. *For any Bayes-optimal $f^s \in \mathcal{F}$ defined on s,*

$$\lim_{n\to\infty} \left|\mathcal{L}\left(x_{s_n}, \texttt{EvOp}(o_{\prec s_n})\right) - \mathcal{L}\left(x_{s_n}, f^s(o_{\prec s_n})\right)\right| = 0. \tag{4}$$

We call algorithms with this property *eventually optimal*. We will define `EvOp` in Section 4, and prove Theorem 1 in Section 4.1. Weak bounds on how long it takes `EvOp` to converge to Bayes-optimal behavior on any individual subsequence are given in Section 4.2.

Eventual optimality is a very strong condition, and only yields guarantees if $\mathcal{F}$ contains Bayes-optimal forecasters. In this paper we focus on showing that an eventually optimal predictor exists, and providing weak bounds on how long it takes it to converge to optimal behavior on a subsequence (and how much loss can be accumulated in the meantime). As we will see in Section 3, this is non-trivial. We leave the problem of converging on the best available forecaster of a subsequence (even if it is not optimal) to future research.

3 Difficulties in this Setting

Total regret and average regret are poor measures of forecaster performance in this setting, and consistency (as defined by Definition 3) is impossible in general. To show this, we will describe an environment $P^{\#}$ which exploits the long delays to make learning difficult.

$P^{\#}$ generates outcomes as follows. It flips a fair coin and reveals it once, and then flips another and reveals it ten times, then flips a third and reveals it one hundred times, and so on, always revealing the kth coin 10^{k-1} times. The forecasters spend one timestep predicting the first coin, ten timesteps predicting the second coin, one hundred timesteps predicting the third coin, and so on. The observations are set

up such that they contain no information about the coin currently being predicted: The forecasters must predict the kth coin all 10^{k-1} times before it is revealed.

Formally, let $\mathcal{X} := \{\text{H}, \text{T}\}$ corresponding to "heads" and "tails" respectively. Let $\mathcal{Y}$ be the set of probability distributions over $\mathcal{X}$, which can be represented as real number $p \in [0,1]$. $P^\#$ is a Markov chain, where each x_{i+1} is conditionally independent from all other outcomes given x_i. $P^\#(\mathbf{X}_1 = \text{H}) = 0.5$. For $i = 2, 12, 112, 1112, \ldots$, x_i "reveals a new coin" and is independent of x_{i-1}: $P^\#(\mathbf{X}_i = \text{H} \mid \mathbf{X}_{i-1} = \cdot\,) = 0.5$. For all other i, x_i "reveals the same coin again:" $x_i = x_{i-1}$. Each $\mathbf{O}_n$ is a deterministic function of $\mathbf{X}_1 \ldots \mathbf{X}_n$ which reveals the first $\lceil \log_{10}(n \cdot 9/10) \rceil$ outcomes. Let $\mathcal{L}$ be squared error; that is, let $\mathcal{L}(\text{H}, p) = (1-p)^2$ and $\mathcal{L}(\text{T}, p) = p^2$.

Clearly, the best prediction of x_n that a forecaster can make given $o_{\prec n}$ is 0.5, because $o_{\prec n}$ does not contain any information about the coin revealed by x_n, which is fair. Thus, the simple forecaster $f^*(o_{\prec n}) = 0.5$ is Bayes-optimal. However, the regret of f^* may be very high! To see this, consider a forecaster f^1, the "gambler," defined $f^1(o_{\prec n}) = 1$. In expectation, f^1 will receive higher total loss on any subsequence of the true outcomes. However, f^1 will spend about half the time with a lower total loss than f^*, because each time a new coin begins being predicted, it has the opportunity to recoup all its losses.

f^* accumulates loss at a rate of $1/4$ units per prediction, which means that, after the kth coin has been predicted all 10^{k-1} times, its aggregate loss is $1/4 \cdot \sum_{i=1}^{k} 10^{i-1}$. f^1 accumulates either 0 or 1 unit of loss in each step according to whether the coin comes up heads or tails, so in the worst case, it will have $\sum_{i=1}^{k} 10^{i-1}$ total loss after the kth coin. If the $k+1$ coin comes up heads, then f^* gains an additional $1/4 10^k$ loss while f^1's loss remains unchanged. 10^k accounts for more than nine tenths of $\sum_{i=1}^{k} 10^i$, so if the coin came up heads then f^1's total loss is at most a tenth of $\sum_{i=1}^{k} 10^i$, whereas f^*'s total loss is a quarter of $\sum_{i=1}^{k} 10^i$. In fact, any predictor that assigns average probability ≤ 0.5 across all 10^{k-1} reveals of the kth coin will have at least 15% more loss than f^1 after the $\sum_{i=1}^{k}$th step, if that coin comes up heads.

By a similar logic, whenever the kth coin comes up tails, f^1's loss shoots up above that of f^*, no matter how lucky it was previously. Thus we see that if $f^1 \in \mathcal{F}$, the regret of f^* will swing wildly back and forth. Any predictor which is maintaining a mixture of forecasters and weighting them according to their regret will have trouble singling out f^*.

Indeed, if the environment is $P^\#$, and if $\mathcal{F}$ contains both f^1 and the opposite gambler f^0 defined as $f^0(o_{\prec n}) = 0$, then it is impossible for a forecaster to be consistent in the sense of Definition 3. If the average probability a forecaster assigns to the kth coin is ≤ 0.5 and the coin comes up heads, it gets very high regret relative to f^1, whereas if it's ≥ 0.5 and the coin comes up tails, it gets very high regret relative to f^0. The only way for a forecaster to avoid high regret against both gamblers is for it to place higher probability on the true result of the coin every single time. With probability 1 it must slip up infinitely often (because the coins are fair), so each forecaster's regret will be high infinitely often. And the amount of regret—at least 15% of all possible loss—is proportional to n, so $\lim_{n\to\infty} \mathbb{E}[R^n(f)]/n$ cannot go to zero.

Lest this seem like a peculiarity of the stochastic setting, observe that a similar problem could easily occur in the deterministic setting, when a learner is predicting the behavior of large computations. For example, imagine that the "coins" are chaotic subsystems inside a physics simulation, such that large environments have many correlated subsystems. In this case, some experts might start "gambling" by making extreme predictions about those subsystems, and it may become difficult to distinguish the accurate forecasters from the gamblers, while looking at total or average regret.

The first fix that comes to mind is to design a predictor with a learning rate that decays over time. For example, if the learner weights the loss on x_n by $1/10n$ then it will assign each cluster of 10^{k-1} predictions roughly equal weight, thereby neutralizing the gamblers. However, this fix is highly unsatisfactory: It runs into

exactly the failures described above on the environment $P_2^\#$ which reveals the kth coin 10^{10^k} times instead. It might be the case that for each specific environment one could tailor a learning rate to that environment that allows a predictor to successfully distinguish the optimal forecasters from the gamblers using regret, but this would be an ad-hockery tantamount to hardcoding the optimal forecaster in from the beginning. This motivates the study of how a predictor can successfully identify optimal experts at all in this setting.

4 The EvOp Algorithm

Section 3 showed that in this setting, it is possible for gamblers to take advantage of correlated outputs and unbounded delays to achieve drastic swings in their total loss, which makes total and average regret bad measures of a forecaster. We can address this problem by comparing forecasters only on *independent* subsequences of outcomes on which they are both defined.

Intuitively, the gamblers are abusing the fact that they can correlate many predictions before any feedback on those predictions is received, so we can foil the gamblers by assessing them only on a subsequence of predictions where each prediction in the subsequence was made only after receiving feedback on the previous prediction in the subsequence. EvOp is an algorithm which makes use of this intuition, and Theorem 1 shows that it is sufficient to allow EvOp to zero in on Bayes-optimal predictors regardless of what strategies other forecasters in $\mathcal{F}$ use.

Definition 5. *A sequence s is **independent** if, for all $i > 1$, $o_{s_i}(s_{i-1})$ is defined.*

Algorithm 1: EvOp, an eventually optimal predictor. $\|\cdot\|$ is the l^2 norm, and $1/0 = \infty$.

Input: $o_{\prec n}$, the first n observations
Data: ε, an arbitrary constant < 1

```
// Computes an independent subsequence on which f_i and f_j
   disagree.
def testseq_n(i, j, m):
    t ← 0
    waiting ← false
    for k in 1, 2, ..., n:
        if waiting and t ∈ dom(o_k):
            output(k)
            waiting ← false
        elif y^i_k := f_i(o_≺k) and y^j_k := f_j(o_≺k) are defined, and ||y^i_k − y^j_k|| > 1/m
        :
            t ← k
            waiting ← true
// Computes the difference between the scores of f_i and f_j on an
   independent subsequence on which they disagree.
def relscore_n(i, j, m):
    s ← testseq_n(i, j, m)
    return Σ_{k=1}^{|s|} ( L(x_{s_k}, y^i_{s_k}) − L(x_{s_k}, y^j_{s_k}) − ρε/(2m^2) )
def maxscore_n(i):
    return max_{j∈N≥1, m∈N≥0} i − j − m + relscore_n(i, j, m)
f ← min_{i∈N≥1 such that o_≺n∈dom(f_i)} maxscore_n(i)
return f(o_≺n)
```

EvOp works as follows. Fix an enumeration $f_1, f_2, \ldots$ of $\mathcal{F}$, which must be countable but need not be finite; we can assume without loss of generality that this

enumeration is countably infinite. `EvOp` compares f_i to f_j by giving it a relative score, which is dependent on the difference between their loss measured only on an independent subsequence of predictions on which they are both defined, constructed greedily. Lower scores are better for f_i. The score is also dependent on ρ, the strong convexity constant for $\mathcal{L}$, and an arbitrary positive $\varepsilon < 1$, which we use to ensure that if f_i and f_j make different predictions infinitely often then their scores actually diverge. `EvOp` follows the prediction of the f_i chosen by minimaxing this score, i.e., it copies the f_i that has the smallest worst-case score relative to any other f_j. Pseudocode for `EvOp` is given by Algorithm 1.

To see that the max step terminates, note that it can be computed by checking only finitely many j and m: $\texttt{relscore}_n(i, j, m)$ is bounded above by $\sum_{k=1}^{|s|} \mathcal{L}(x_k, y_k^i)$, so all (j, m) pairs such that $j + m$ is greater than this value may be discarded. To see that the min step terminates, note that it can be computed by checking only finitely many i (assuming that at least one f is defined on $o_{\prec n}$), because when $m = 0$, $\texttt{testseq}_n(i, j, m)$ is empty; thus when $j = 1$ and $m = 0$, $\texttt{maxscore}_n(i, j, m)$ is at least $i - 1$. Therefore, after finding the smallest k such that f_k is defined on $o_{\prec n}$, the min step need only continue searching up through $i = \texttt{maxscore}_n(k) + 1$.

`EvOp` gets around the problems of Section 3 by comparing forecasters only on greedily-constructed independent subsequences of the outcomes. Note that if the delay between prediction and feedback grows quickly, these subsequences might be very sparse. For example, in the environment $P^{\#}$ of Section 3, the independent subsequence will have at least 10^i timesteps between the $i - 1$st element in the subsequence and the next. This technique allows `EvOp` to converge on Bayes-optimal behavior, but it also means that it may do so very slowly (if the subsequence is very sparse). Under certain assumptions about the speed with which delays grow and the frequency with which forecasters disagree, it is possible to put bounds on how quickly `EvOp` converges on Bayes-optimal behavior, as discussed in Section 4.2. However, these bounds are quite weak.

4.1 Proof of Theorem 1

To prove Theorem 1 we need two lemmas, which, roughly speaking, say that (1) if f_z is Bayes-optimal then $\texttt{maxscore}_n(z)$ is bounded; and (2) if f_j is not Bayes-optimal and some $f_z \in \mathcal{F}$ is, then $\texttt{maxscore}_n(j)$ goes to infinity. From there, the proof is easy.

In what follows, let f_z be a Bayes-optimal forecaster (as per Definition 4) that makes infinitely many predictions all of which are almost surely eventually revealed—that is, such that $f_z(o_{\prec n})$ is almost surely defined infinitely often, and whenever it is defined, $o_i(n)$ is almost surely defined for some i. Let z be the index of f_z in the enumeration over $\mathcal{F}$. In general, we will write y_n^i for $f_i(o_{\prec n})$ when it is defined.

Lemma 1. *If f_z is Bayes-optimal and makes infinitely many predictions all of which are almost surely eventually revealed, then with probability 1,* $\texttt{maxscore}_n(z)$ *is bounded.*

Proof. For all j, $\texttt{relscore}_n(z, j, 0) = 0$, because $\texttt{testseq}_n(z, j, 0)$ never outputs. Thus, $\texttt{maxscore}_n(z)$ is bounded below by $z - 1$ (consider the case where $j = 1$ and $m = 0$) and bounded above by $z - j - m + \texttt{relscore}_n(z, j, m)$. When $m = 0$ this is bounded above by $z - j$, so it suffices to show that there is almost surely some bound B such that $\texttt{relscore}_n(z, j, m) - j - m$ is bounded above by B for every j and $m \geq 1$.

Intuitively, in expectation, $\texttt{relscore}_n(z, j, m)$ should either be finite or diverge to $-\infty$, because f_z is Bayes-optimal and is only being compared to other forecasters on independent subsequences. We will prove not only that it's bounded above in expectation, but that it is bounded above with probability 1. To do this we use Lemma 2 in Appendix A, which (roughly speaking) says that something which is zero in expectation, and which has "not too much" variance in expectation, can't get too far from zero in fact.

Fix j, $m \geq 1$, and λ; we will bound the probability that $\texttt{relscore}_n(z,\, j,\, m) \geq \lambda$. Let $s = s_1 s_2 \ldots$ be the outputs of $\texttt{testseq}_\infty(x,\, j,\, m)$, that is, the entire greedily-generated sparse independent subsequence of outputs on which both f_z and f_j make predictions that differ by at least $1/m$ (which could be generated by running $\texttt{testseq}_n(x,j,m)$ on larger and larger n). s may or may not be finite.

Because $\mathcal{L}$ is strongly convex,

$$\mathcal{L}(x_k, y_k^j) \geq \mathcal{L}(x_k, y_k^z) + \underset{y}{\nabla} \mathcal{L}(x_k, y_k^z) \cdot (y_k^j - y_k^z) + \frac{\rho}{2}\|y_k^j - y_k^z\|^2, \tag{5}$$

where ∇_y takes the gradient of $\mathcal{L}$ with respect to the prediction, ρ is the strong convexity constant of $\mathcal{L}$, and $\|\cdot\|$ is the l^2 norm. In other words, the loss of f_j in any given round is at least that of f_z plus a linear term (which, note, is related to the Lipschitz constant of $\mathcal{L}$) plus a quadratic term. Rearranging this inequality,

$$\mathcal{L}(x_k, y_k^z) - \mathcal{L}(x_k, y_k^j) \leq -\underset{y}{\nabla} \mathcal{L}(x_k, y_k^z) \cdot (y_k^j - y_k^z) - \frac{\rho}{2}\|y_k^j - y_k^z\|^2. \tag{6}$$

We will show that the sum of the right-hand side for $k = 1, 2, \ldots, n$ is bounded, using Lemma 2.

Lemma 2 requires a sequence of random variables $G_1 H_1 G_2 H_2 \ldots$ that form a Markov chain, and two real-valued functions v and r defined on the G_i and the H_i respectively, such that $\mathbb{E}[r(H_i) \mid v(G_i)] = 0$, and $|r(H_i)| \leq a\sqrt{v(G_i)}$ for some constant a. Intuitively, these constraints say that r is zero in expectation, and that its absolute value is bounded by v. Lemma 2 then gives us a bound on the probability that $\sum_{i=1}^{n} r(H_i) - v(G_i) \geq \lambda$. We use it with r as the first term on the right-hand side of equation (6), and v as the negative of the second. Roughly, r can be thought of as a first-order approximation to the amount by which f_j did better than expected (a "residual"), and v as a bound on how wildly r can swing (a "variance").

Let G_i be $o_{\prec s_i}$ and H_i be $o_{\prec k}$ where k is the least time after s_i such that $s_i \in \mathrm{dom}(o_k)$. k exists, because f_z only makes predictions that, with probability 1, are eventually revealed. Intuitively, our Markov chain alternates between elements of s and the times when those elements were revealed. For $i > |s|$, let $G_i = H_i = o_{\prec\infty}$, the (infinite) combination of all observations.

Define r to be the function $r(H_i) = -\nabla_y \mathcal{L}(x_{s_i}, y_{s_i}^z) \cdot (y_{s_i}^j - y_{s_i}^z)$ when $i \leq |s|$, and 0 otherwise. Observe that this value can be calculated from H_i, f_z, and f_j, because $H_i = o_{\prec k}$, with $k > s_i$ and $x_{s_i} \coloneqq o_{\prec k}(s_i)$ defined.

Define v to be the function $v(G_i) = \frac{\rho}{2}\|y_{s_i}^j - y_{s_i}^z\|^2$ when $i \leq |s|$, and $\rho/2m^2$ otherwise, which can be calculated from G_i, f_z, and f_j, because G_i is just $o_{\prec s_i}$. Note that $\mathbb{E}[r(H_k) \mid G_k] = 0$, because f_z is a Bayes-optimal predictor, which means it minimizes expected loss, making the gradient in $r(H_i)$ zero in expectation for all i. Note also that because $\mathcal{L}$ is Lipschitz, $|r(H_k)| \leq \kappa\|y_n^j - y_n^z\|$ where κ is the Lipschitz constant of $\mathcal{L}$. Thus, with $a = \frac{\kappa\sqrt{2}}{\sqrt{\rho}}$, $|r(H_k)| \leq a\sqrt{v(G_k)}$. Therefore, r and v meet the conditions of Lemma 2, so for all M,

$$\mathbb{P}\left(\sum_{i=1}^{n} r(H_i) - v(G_i) \geq M\right) \leq \exp\left(-\rho\kappa^{-2}M\right), \tag{7}$$

which goes to 0 as M goes to infinity. We need a bound that forces it to 0 as $n \to \infty$. In what follows, we write $b = \rho\kappa^{-2}$ for conciseness.

Observe that $\texttt{relscore}_n(z,j,m) \leq \sum_{i=1}^{t_n} (r(H_i) - v(G_i) - \rho\varepsilon/2m^2)$, where t_n is the number of times $\texttt{testseq}_n(z,\, j,\, m)$ outputs. Thus, the probability that $\texttt{relscore}_n(z,j,m) \geq \Lambda$ for any Λ is upper-bounded by the probability that, for some n,

$$\left(\sum_{i=1}^{t_n} r(H_i) - v(G_i)\right) - t_n\frac{\rho\varepsilon}{2m^2} \geq \Lambda. \tag{8}$$

For any given n, applying inequality (7) with $\Lambda + t_n \frac{\rho\varepsilon}{2m^2}$ for M,

$$\mathbb{P}\left(\sum_{i=1}^{t_n} r(H_i) - v(G_i) \geq \Lambda + t_n \frac{\rho\varepsilon}{2m^2}\right) \leq \exp\left(-b\left(\Lambda + \frac{t_n \rho\varepsilon}{2m^2}\right)\right). \tag{9}$$

We now see the function of the $\rho\varepsilon/2m^2$ term in `relscore`: it adds a tiny bias in favor of the forecaster being judged, such that the longer a contender waits to prove itself, the more it has to prove. Equation (9) says that, because f_j never proves itself too much in expectation, the probability that f_z's score relative to f_j goes strongly in f_j's favor gets lower as t_n gets larger.

Note that $\texttt{relscore}_n(z,j,m)$ only depends on n through t_n: If $t_{n_1} = t_{n_2}$ for some n_1 and n_2 then $\texttt{relscore}_{n_1}(z,j,m) = \texttt{relscore}_{n_2}(z,j,m)$. Thus, $\mathbb{P}\left(\exists n\colon \texttt{relscore}_n(z,j,m) > \Lambda\right)$ can be bounded by summing only over the possible values t of t_n.

$$\sum_{t=0}^{\infty} \exp\left(-b\Lambda + t\left(-\frac{b\rho\varepsilon}{2m^2}\right)\right) = \frac{\exp(-b\Lambda)}{1 - \exp\left(-\frac{b\rho\varepsilon}{2m^2}\right)}, \tag{10}$$

and $m \geq 1$, so

$$\mathbb{P}\big(\exists n\colon \texttt{relscore}_n(z,j,m) \geq \Lambda\big) \leq \frac{\exp(-b\Lambda)}{1 - \exp(-b\rho\varepsilon/2)}. \tag{11}$$

Applying inequality (11) with $\lambda + m + j$ for Λ, we see that the probability $\texttt{relscore}_n(z,j,m) \geq \lambda + m + j$ is at most

$$\sum_{m=1}^{\infty}\sum_{j=1}^{\infty} \frac{\exp(-b(\lambda + m + j))}{1 - \exp(-b\rho\varepsilon/2)} = \frac{\exp(-b(\lambda+2))}{(1 - \exp(-b\rho\varepsilon/2))(1 - \exp(-b))^2}. \tag{12}$$

This goes to 0 as λ goes to ∞. Therefore, with probability 1, there exists some bound B such that $\texttt{relscore}_n(z,j,m) - m - j < B$ for all j and $m \geq 1$. Thus, $\texttt{maxscore}_n(z)$ is almost surely bounded. □

If f_z is Bayes-optimal and makes infinitely many predictions all of which are almost surely eventually revealed, then for any f_j, with probability 1, if $y_i^z \coloneqq f_z(o_{\prec i})$ and $y_i^j \coloneqq f_j(o_{\prec i})$ are both defined on the same t infinitely often, and if $\|y_i^z - y_i^j\| \geq \delta$ infinitely often for some $\delta > 0$,

$$\lim_{n\to\infty} \texttt{maxscore}_n(j) = \infty. \tag{13}$$

Roughly speaking, the proof runs as follows. Choose m such that $1/m < \delta$. It suffices to show that $\texttt{relscore}_n(j,z,m) \to \infty$ as $n \to \infty$. The $\sum(\mathcal{L}(x_k, y_k^j) - \mathcal{L}(x_k, y_k^z))$ portion goes to infinity in expectation, and also goes to infinity with probability 1 by Lemma 2. It remains to show that the $\sum \rho\varepsilon/2m^2$ terms working in f_j's favor are not sufficient to prevent the total from going to infinity, which can be done by showing that the differences between $\mathcal{L}(x_k, y_k^j)$ and $\mathcal{L}(x_k, y_k^j)$ are at least $\rho/2m^2 > \rho\varepsilon/2m^2$ in expectation, and appealing again to Lemma 2. The proof proceeds similarly to the proof of Lemma 1, so we leave the details to Appendix B.

With these lemmas in place, we now prove that `EvOp` is eventually optimal. Recall Theorem 1:

Theorem 1. *For any Bayes-optimal $f^s \in \mathcal{F}$ defined on s,*

$$\lim_{n\to\infty} \left|\mathcal{L}\left(x_{s_n}, \texttt{EvOp}(o_{\prec s_n})\right) - \mathcal{L}\left(x_{s_n}, f^s(o_{\prec s_n})\right)\right| = 0. \tag{4}$$

Proof. Let f_z be Bayes-optimal and defined infinitely often, such that everything it predicts is almost surely eventually revealed. It suffices to show that, with probability 1, if $f_z(o_{\prec n})$ is defined then

$$\lim_{n\to\infty} \|\texttt{EvOp}(o_{\prec n}) - f_z(o_{\prec n})\| = 0. \tag{14}$$

By Lemma 1, $\texttt{maxscore}_n(z)$ is bounded with probability 1. Let B be this bound. Note that there are only finitely many i such that $\texttt{maxscore}_n(i) \leq B$, for the same reason that the min step always terminates. For each of those i, either f_i and f_z converge to the same prediction, or they only make finitely many predictions in common, or (by Lemma 4.1) $\texttt{maxscore}_n(i) \to \infty$. The latter contradicts the assumption that $\texttt{maxscore}_n(i) \leq B$. If f_i and f_z only make finitely many predictions in common, then for sufficiently large n, f_i is not defined and so will not be selected. Thus, we need only consider the case where f_i and f_z converge to the same predictions whenever they both make predictions. In this case, $\texttt{EvOp}(o_{\prec n})$ is choosing among finitely many forecasters all of which converge to $f_z(o_{\prec n})$, so $\texttt{EvOp}(o_{\prec n})$ must converge to $f_z(o_{\prec n})$. □

4.2 Bounds

The speed with which EvOp converges to optimal behavior on a subsequence depends on both (1) the sparseness of independent subsequences in the outcomes; and (2) the frequency with which forecasters make claims that differ.

Specifically, assume that all forecasters are defined everywhere and disagree infinitely often, and that $\mathcal{F}$ is finite. (The first two constraints imply the third.) We can show that, given a (potentially fast-growing) function h bounding how long it takes before predictors disagree with each other, and given another (potentially fast-growing) function g bounding the delay in feedback, and given a probability p, the time it takes before EvOp has converged on f_z with probability p is proportional to $h \circ g$ iterated a number of times proportional to $\log p$. (Note that h and g are not uniform bounds; $g(n)$ is the maximum delay between the nth prediction and feedback on the nth prediction, and delays may grow ever larger as n increases.)

Theorem 2. *Given h, g, a Bayes-optimal f_z, and a probability p, there is an $N \propto (h \circ g)^{\log p}(1)$ such that, with probability at least $1 - p$, for all $n \geq N$,*

$$\texttt{EvOp}(o_{\prec n}) = f_z(o_{\prec n}). \tag{15}$$

We prove Theorem 2 in Appendix C.

To call these bounds "weak" is an understatement. In the case where the outcomes are generated by running a universal Turing machine U on different inputs, g is infinite, because U will sometimes fail to output. It is possible to achieve *much* better bounds given certain simplifying assumptions, such as delays that are finite in expectation [2]. However, it is not yet clear which simplifying assumptions to use, or what bounds to ask for, in the setting with ever-growing delays.

5 The Deterministic Setting

Our motivation for studying online learning with unbounded delays in a stochastic setting is that this gives us a simplified model of the problem of predicting large computations from observations of smaller ones. We have already seen one instance of an issue in the stochastic setting which looks likely to have an analog in the deterministic setting. In Section 3 we gave the example of a deterministic "coin" that appears more and more often in larger and larger computations, which might (for instance) be a common subsystem in the environment of a physical simulation. Intuitively, if there are many correlated subsystems that appear "sufficiently random" to all forecasters, then forecasters might follow the strategies of f^1 and f^0 in Section 3 and achieve regular large swings in their total loss. Intuitively, the techniques used in Algorithm 1 to handle the problem in the stochastic case should

well carry over to the deterministic case, but any attempt to formalize this intuition depends on what it means for a deterministic sequence to be "sufficiently random."

For that we turn to algorithmic information theory, a field founded by [23] which studies the degree and extent to which fixed bitstrings can be called "random." In their canonical text, [3] give three different definitions of algorithmic randomness and show them all to be equivalent. The oldest of the three, given by [23], is rooted in the idea that an algorithmically random sequence should satisfy all computably verifiable properties that hold with probability 1 on randomly generated sequences.

It is with this definition in mind that we note that Lemma 1 and Lemma 4.1 are both stated as properties that are true of randomly generated sequences with probability 1. Lemma 1 says that if the outputs of the environment are generated randomly, then with probability 1, the score of a Bayes-optimal predictor does not go to infinity. Lemma 4.1 says that if the outputs of the environment are generated randomly, then with probability 1, a predictor that disagrees by $\delta > 0$ with a Bayes-optimal predictor infinitely many times has its score going to infinity. Both these computable properties hold for random sequences with probability 1, so they hold for Martin-Löf-random sequences.

This means that if $\mathcal{F}$ is the class of all Turing machines, and `EvOp` is predicting an algorithmically random sequence (such as Chaitin's Ω, the fraction of Turing machines which halt), then Theorem 1 holds and `EvOp` will converge on optimal predictions on subsequences of that sequence. However, this does us no good: There are no computable patterns in Chaitin's Ω; computable forecasters won't be able to do any better than predicting a 50% chance of a 1. Besides, the goal is not to predict uncomputable sequences by running all Turing machines. The goal is to predict large computations using efficient (e.g., polynomial-time) experts.

What we need is a notion of algorithmic randomness *with respect to a restricted class of experts*. For example, if $\mathcal{F}$ is the class of polynomial-time forecasters, we would like a notion of sequences which are algorithmically random with respect to polynomial-time forecasters.

The authors do not yet know of a satisfactory definition of algorithmic randomness with respect to resource constraints. However, the obvious analog of Martin-Löf's original definition [23] is that a sequence should be defined as algorithmically random with respect to a class of bounded experts if, and only if, it satisfies all properties that hold of randomly generated sequences with probability 1 *and that can be checked by one of those experts*. On sequences that are algorithmically random with respect to $\mathcal{F}$ in this sense, Lemma 1 and Lemma 4.1 must apply: Assume f_z is a Bayes-optimal predictor on a subsequence that is algorithmically random with respect to $\mathcal{F}$; any forecaster $f_j \in \mathcal{F}$ that outperforms f_z infinitely often would be identifying a way in which the sequence fails to satisfy a property that randomly generated sequences satisfy with probability 1, which contradicts the assumption. This gives strong reason to expect that `EvOp` would be eventually optimal when predicting sequences that are algorithmically random with respect to $\mathcal{F}$, even though formalizing such a notion remains an open problem.

Even so, this does not mean that `EvOp` would perform *well* at the actual task of predicting large computations from the observation of small ones. Eventual optimality provides no guarantees about the ability of the algorithm to converge on good but non-optimal predictors, and the bounds that we have on how long it takes `EvOp` to converge on good behavior are weak (to say the least).

Furthermore, there are other notions of what it means to "predict computations well" that are not captured by eventual optimality. For example, [21] discusses the problem of computably assigning probabilities to the outputs of computations and refining them in such a way that they are "coherent," drawing on inspiration from the field of mathematical logic that dates at least back to [22]. The intuition is that given two statements "this computation will halt and output 1" and "this computation will fail to halt or output something besides 1," a good reasoner should assign those claims probabilities that sum to roughly 1. We have no reason to expect that `EvOp` has any such property.

6 Conclusions

We have studied online learning in a setting where delays between prediction and observation may be unbounded, in attempts to explore the general problem of predicting the behavior of large computations from observations of many small ones. We found that, in the stochastic setting, the unbounded delays give rise to difficulties: Total regret and average regret are not good measures of forecaster success, and consistency is not possible to achieve in general. However, it is possible to converge on good predictions by comparing forecasters according to their performance only on sparse and independent subsequences of the observations, and we have reason to expect that some of the techniques used to achieve good performance in the stochastic setting will carry over into the deterministic setting. We have proposed an algorithm `EvOp` that converges to optimal behavior. It is not a practical algorithm, but it does give a preliminary model of online learning in the setting where the delay between prediction and feedback is ever-growing.

Our results suggest a few different paths for future research. `EvOp` handles the problem of learning in the face of potentially unbounded delays by comparing forecasters only on subsequences that are potentially very sparse, and this means that it converges to optimal behavior quite slowly. Speeding up convergence without falling prey to the problems described in Section 3 might prove difficult. Furthermore, `EvOp` only guarantees convergence on forecasters that are Bayes-optimal; it is not yet clear how to converge on the best available forecaster (even if it is non-optimal) in the face of unbounded delays. As mentioned in Section 5, a formal notion of algorithmic randomness with respect to a bounded class of experts would make it easier to study the problem of using online learning to predict the behavior of large computations in a deterministic setting. `EvOp` is only a first step towards a predictor that can learn to predict the behavior of large computations from the observation of small ones, and the problem seems ripe for further study.

Acknowledgements

Thanks to Jessica Taylor for the proof of Lemma 2, and to Benya Fallenstein for helpful discussions. This work was supported by the Future of Life Institute.

A Proof of Lemma 2

Lemma 2. *Let $\mathcal{G}$ and $\mathcal{H}$ be sets, and let $G_1, H_1, G_2, H_2, ..., G_n, H_n$ be random variables forming a Markov chain (with each $G_i \in \mathcal{G}$ and $H_i \in \mathcal{H}$). Let there be functions $v : \mathcal{G} \to \mathbb{R}_{\geq 0}$ and $r : \mathcal{H} \to \mathbb{R}$, with $|r(H_i)| \leq a\sqrt{v(G_i)}$ and $\mathbb{E}[r(H_i)|G_i] \leq 0$. Let $\lambda > 0$. Then*

$$P\left(\sum_{i=1}^{n}(r(F_i) - v(G_i)) \geq \lambda\right) \leq \exp\left(-2/a^2\lambda\right) \tag{16}$$

Proof. This proof closely follows the standard proof of Azuma's inequality, given by, e.g., [24]. Let $b = 2/a^2$. Using Markov's inequality:

$$\begin{aligned} &P\left(\sum_{i=1}^{n}(r(H_i) - v(G_i)) \geq \lambda\right) \\ &= P\left(\exp\left(b\sum_{i=1}^{n}(r(H_i) - v(G_i))\right) \geq \exp\left(b\lambda\right)\right) \\ &\leq \exp\left(-b\lambda\right)\mathbb{E}\left[\exp\left(b\sum_{i=1}^{n}(r(H_i) - v(G_i))\right)\right] \end{aligned} \tag{17}$$

To bound the expectation, we will inductively show that for all $m \leq n$,

$$\mathbb{E}\left[\exp\left(b\sum_{i=1}^{m}(r(H_i)-v(G_i))\right)\right] \leq 1 \tag{18}$$

When $m = 0$, this is trivial. Otherwise:

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(b\sum_{i=1}^{m}(r(H_i)-v(G_i))\right)\right] \\
&= \mathbb{E}\left[\exp\left(b\sum_{i=1}^{m-1}(r(H_i)-v(G_i))\right)\exp\left(-bv(G_i)\right)\mathbb{E}\left[e^{br(F_m)}|F_{1:m-1}\right]\right] \\
&\leq \mathbb{E}\left[\exp\left(b\sum_{i=1}^{m-1}(r(H_i)-v(G_i))\right)\exp\left(-bv(G_i)\right)\exp\left(b^2a^2v(F_{m-1})/2\right)\right] \qquad (19)\\
&= \mathbb{E}\left[\exp\left(b\sum_{i=1}^{m-1}(r(H_i)-v(G_i))\right)\exp\left(-bv(F_{m-1})+bv(F_{m-1})\right)\right] \\
&= \mathbb{E}\left[\exp\left(b\sum_{i=1}^{m-1}(r(H_i)-v(G_i))\right)\right].
\end{aligned}$$

By the inductive assumption, this quantity is no more than 1, so the inductive argument goes through. Using this bound on the expectation, the given upper bound or the original probability of interest follows. □

B Proof of Lemma 4.1

If f_z is Bayes-optimal and makes infinitely many predictions all of which are almost surely eventually revealed, then for any f_j, with probability 1, if $y_i^z \coloneqq f_z(o_{\prec i})$ and $y_i^j \coloneqq f_j(o_{\prec i})$ are both defined on the same t infinitely often, and if $\|y_i^z - y_i^j\| \geq \delta$ infinitely often for some $\delta > 0$,

$$\lim_{n\to\infty} \texttt{maxscore}_n(j) = \infty. \tag{13}$$

Proof. Let $1/m < \delta$. It suffices to show that with probability 1,

$$\lim_{n\to\infty} \texttt{relscore}_n(j,\, z,\, m) = \infty. \tag{20}$$

Write t_n for the number of times that $\texttt{testseq}_n(j,\, z,\, m)$ outputs, and note that $t_n \to \infty$ as $n \to \infty$ because f_j and f_z disagree by more than δ infinitely often. We will show that $\texttt{relscore}_n(j,\, z,\, m)$ is bounded below by a bound proportional to t_n, which means that $\texttt{relscore}_n(j,\, z,\, m)$ must diverge to infinity.

Let $s = \texttt{testseq}_\infty(j,z,m)$. Define $G_1H_1G_2H_2\ldots$, $r(H_i)$, and $v(G_i)$ as in the proof of Lemma 1. Recall that $r(H_i) - v(G_i)$ is an upper bound for $\mathcal{L}(x_i, y_i^z) - \mathcal{L}(x_i, y_i^j)$, which means that $v(G_i) - r(H_i)$ is a lower bound for $\mathcal{L}(x_i, y_i^j) - \mathcal{L}(x_i, y_i^z)$. Therefore, it suffices to show that, for some $\alpha > 0$,

$$\lim_{N\to\infty} \mathbb{P}\left(\forall n > N\colon \sum_{i=1}^{t_n}\left(v(G_i) - r(H_i) - \frac{\rho\varepsilon}{2m^2}\right) \geq \alpha t_n\right) = 1. \tag{21}$$

Observe that $v(G_i) \geq \rho/2m^2$ for all i, so the positive $v(G_i)$ terms going against f_j more than compensate for the negative $\rho\varepsilon/2m^2$ terms going in its favor. Because

$\varepsilon < 1$, only a $\frac{1+\varepsilon}{2}$ portion of each $v(G_i)$ is needed to cancel out the $\rho\varepsilon/2m^2$ terms,

$$\mathbb{P}\left(\sum_{i=1}^{t_n}\left(v(G_i) - r(H_i) - \frac{\rho\varepsilon}{2m^2}\right) \geq \alpha t_n\right) \geq \mathbb{P}\left(\sum_{i=1}^{t_n}\left(\frac{1-\varepsilon}{2}v(G_i) - r(H_i)\right) \geq t_n\left(\alpha + \frac{\rho(\varepsilon-1)}{4m^2}\right)\right). \quad (22)$$

Now we apply Lemma 2. $\mathbb{E}[r(H_k) \mid G_k]$ is still 0. With $a = \frac{\kappa\sqrt{2}}{\sqrt{\rho}} \cdot \sqrt{\frac{2}{1-\varepsilon}}$,

$$|r(H_k)| \leq a\sqrt{\frac{1-\varepsilon}{2}v(G_k)}. \quad (23)$$

Therefore, by Lemma 2 we have that

$$\mathbb{P}\left(\sum_{i=1}^{t_n}\left(r(H_i) - \frac{1-\varepsilon}{2}v(G_i)\right) \geq M\right) \leq \exp\left(-\frac{\rho(1-\varepsilon)M}{2\kappa^2}\right). \quad (24)$$

Choose $\alpha = \rho(1-\varepsilon)/8m^2$ and set $M = -t_n\left(\alpha + \frac{\rho(\varepsilon-1)}{4m^2}\right) = t_n\frac{\rho(1-\varepsilon)}{8m^2}$ to get:

$$\mathbb{P}\left(\sum_{i=1}^{t_n}\left(\frac{1-\varepsilon}{2}v(G_i) - r(H_i)\right) \leq t_n\frac{\rho(\varepsilon-1)}{8m^2}\right) \leq \exp\left(-\frac{t_n\rho^2(1-\varepsilon)^2}{16m^2\kappa^2}\right). \quad (25)$$

We write $c = \rho^2(1-\varepsilon)^2/16m^2\kappa^2$ for conciseness. Observe that

$$\mathbb{P}\left(\exists n \geq N\colon \sum_{i=1}^{t_n}\left(v(G_i) - r(H_i) - \frac{\rho\varepsilon}{2m^2}\right) \leq \alpha t_n\right) \leq \sum_{t=t_N}^{\infty}\exp(-tc) = \frac{\exp(-t_N c)}{1-\exp(-c)}. \quad (26)$$

If $|s| = \infty$ then the right-hand side almost surely goes to zero as $n \to \infty$, in which case, with probability 1, there exists an N such that

$$\forall n > N\colon \sum_{i=1}^{t_n}\left(v(G_i) - r(H_i) - \frac{\rho\varepsilon}{2m^2}\right) \geq \alpha t_n. \quad (27)$$

Thus if f_z and f_j disagree by more than δ infinitely often, then with probability 1, eventually $\texttt{relscore}_n(j,z,m)$ grows proportionally to t_n. Therefore, with probability 1,

$$\lim_{n\to\infty}\texttt{relscore}_n(j,z,m) = \infty, \quad (28)$$

so $\texttt{maxscore}_n(j)$ almost surely diverges to ∞ as $n \to \infty$. □

C Proof of Theorem 2

Let f_z be a Bayes-optimal predictor and assume $\mathcal{F}$ is finite. Assume we have an increasing function h such that for some m and every f_j, for all times t, there exists a $t < t' < h_j^m(t)$ such that $y_{t'}^z \coloneqq f_z(o_{\prec t'})$ and $y_{t'}^j \coloneqq f_j(o_{\prec t'})$ are both defined and $\|y_{t'}^z - y_{t'}^j\| > 1/m$. Assume we have an increasing function g such that $o_{\prec g(t)}(t)$ is always defined. $\circ$ denotes function composition; i.e., $(h \circ g)^n(1)$ denotes $h(g(\ldots h(g(1))))$ with n calls to h and g.

Theorem 2. *Given h, g, a Bayes-optimal f_z, and a probability p, there is an $N \propto (h \circ g)^{\log p}(1)$ such that, with probability at least $1-p$, for all $n \geq N$,*

$$\texttt{EvOp}(o_{\prec n}) = f_z(o_{\prec n}). \tag{15}$$

Proof. Observe that $\texttt{testseq}_n(j, z, m)$ outputs at least t terms for some t such that $(h \circ g)^t(1) \leq n$. In the proof of Lemma 1, we prove that the probability that $\texttt{maxscore}_n(\text{z}) \geq \lambda$ for any n is at most

$$\frac{\exp(-b(\lambda + 2 - z))}{(1 - \exp(-b\rho\varepsilon/2))(1 - \exp(-b))^2}. \tag{29}$$

In the proof of Lemma 4.1, we prove that the probability that

$$\texttt{maxscore}_n(j) \leq \alpha t - m - z + j \tag{30}$$

for any n such that $\texttt{testseq}_n(j, z, m)$ outputs at least t terms is at most

$$\frac{\exp(-tc)}{1 - \exp(-c)}. \tag{31}$$

Combining these, we get that for any T, if we let t be the maximal t such that $(h \circ g)^t(1) \leq T$, then for $\lambda = \alpha t - m - z + |\mathcal{F}|$, with probability at least

$$1 - \left(\frac{\exp(-b(\lambda + 2 - z))}{(1 - \exp(-b\rho\varepsilon/2))(1 - \exp(-b))^2} + |\mathcal{F}| \frac{\exp(-tc)}{1 - \exp(-c)} \right), \tag{32}$$

$\texttt{EvOp}(o_{\prec n}) = f_z(o_{\prec n})$ for all times after T. This also gives us a weak bound on total loss: Because $\mathcal{L}$ is both Lipschitz and strongly convex, it is bounded. Let L be the bound. Then with probability as per equation (32), the total loss never goes above LT.

Reversing this process, we also get that for any p, if we let t be such that

$$\left(\frac{\exp(-b(\alpha t - m - z + |\mathcal{F}| + 2 - z))}{(1 - \exp(-b\rho\varepsilon/2))(1 - \exp(-b))^2} + |\mathcal{F}| \frac{\exp(-ct)}{1 - \exp(-c)} \right) < p, \tag{33}$$

then with probability at least $1-p$, for all $n \geq (h \circ g)^t(1)$, $\texttt{EvOp}(o_{\prec n}) = f_z(o_{\prec n})$. □

References

[1] Miroslav Dudik et al. "Efficient Optimal Learning for Contextual Bandits". In: *Proceedings of the 27th Annual Conference on Uncertainty in Artificial Intelligence (UAI-11)*. Corvallis, Oregon: AUAI Press, 2011, pp. 169–178.

[2] Pooria Joulani, András György, and Csaba Szepesvári. "Online Learning Under Delayed Feedback". In: *Proceedings of The 30th International Conference on Machine Learning*. Ed. by Sanjoy Dasgupta and David McAllester. Vol. 28. JMLR Workshop and Conference Proceedings. 2013, pp. 1453–1461.

[3] Rodney G. Downey and Denis R. Hirschfeldt. *Algorithmic Randomness and Complexity*. Theory and Applications of Computability. New York, NY: Springer, 2010. ISBN: 978-0-387-95567-4. DOI: 10.1007/978-0-387-68441-3.

[4] Nick Littlestone and Manfred K. Warmuth. "The Weighted Majority Algorithm". In: *Information and computation* 108.2 (1994), pp. 212–261.

[5] Volodimir G. Vovk. "Aggregating Strategies". In: *3rd Annual Workshop on Computational Learning Theory (COLT90), Rochester, NY, August 06-08, 1990 Proceedings*. Ed. by Mark Fulk and John Case. Morgan Kaufmann, 1990, pp. 371–383.

[6] Nicolò Cesa-Bianchi, David P. Helmbold, and Sandra Panizza. "On Bayes Methods for On-Line Boolean Prediction". In: *Algorithmica* 22 (1998), pp. 112–137.

[7] David Haussler, Jyrki Kivinen, and Manfred K. Warmuth. "Tight Worst-Case Loss Bounds for Predicting with Expert Advice". In: *Computational Learning Theory: Second European Conference (EuroCOLT '95), Barcelona, Spain, March 13-15, 1995 Proceedings*. Ed. by Paul Vitányi. Springer, 1995, pp. 69–83.

[8] Alexander Rakhlin and Karthik Sridharan. "Online Learning with Predictable Sequences". In: *Proceedings of the 26th Conference on Learning Theory (COLT 2013)*. Ed. by Shai Shalev-Shwartz and Ingo Steinwart. Vol. 30. JMLR Workshop and Conference Proceedings. 2012, pp. 1–27.

[9] Eyal Gofer et al. "Regret minimization for branching experts". In: *Proceedings of the 26th Conference on Learning Theory (COLT 2013)*. Ed. by Shai Shalev-Shwartz and Ingo Steinwart. Vol. 30. JMLR Workshop and Conference Proceedings. 2013, pp. 618–638.

[10] Antonio Piccolboni and Christian Schindelhauer. "Discrete Prediction Games with Arbitrary Feedback and Loss". In: *Computational Learning Theory: 14th Annual Conference on Computational Learning Theory (COLT 2001), and 5th European Conference on Computational Learning Theory (EuroCOLT 2001), Amsterdam, The Netherlands, July 16-19, 2001 Proceedings*. Vol. 2111. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2001, pp. 208–223.

[11] Marcelo J. Weinberger and Erik Ordentlich. "On Delayed Prediction of Individual Sequences". In: *Information Theory, IEEE Transactions on* 48.7 (2002). Ed. by Fabio Cozman and Avi Pfeffer, pp. 1959–1976.

[12] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. "Finite-Time Analysis of the Multiarmed Bandit Problem". In: *Machine Learning* 47.2-3 (2002), pp. 235–256.

[13] Gergely Neu et al. "Online Markov Decision Processes Under Bandit Feedback". In: *Advances in Neural Information Processing Systems 23 (NIPS 2010)*. Curran Associates, Inc., 2010, pp. 1804–1812.

[14] Jon Christian Mesterharm. "On-line Learning with Delayed Label Feedback". In: *Algorithmic Learning Theory: 16th International Conference (ALT 2005), Singapore, October 8-11, 2005. Proceedings*. Ed. by Sanjay Jain, Hans Ulrich Simon, and Etsuji Tomita. Vol. 3734. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2005, pp. 399–413.

[15] Jon Christian Mesterharm. "Improving On-Line Learning". PhD thesis. Rutgers, The State University of New Jersey, 2007.

[16] Thomas Desautels, Andreas Krause, and Joel W. Burdick. "Parallelizing Exploration-Exploitation Tradeoffs in Gaussian Process Bandit Optimization". In: *Journal of Machine Learning Research* 15.1 (2014), pp. 3873–3923.

[17] Venkata N. Padmanabhan and Jeffrey C. Mogul. "Using Predictive Prefetching to Improve World Wide Web Latency". In: *ACM SIGCOMM Computer Communication Review* 26.3 (1996), pp. 22–36.

[18] Martin Zinkevich, John Langford, and Alex J. Smola. "Slow Learners are Fast". In: *Advances in Neural Information Processing Systems 22 (NIPS 2009)*. Curran Associates, Inc., 2009, pp. 2331–2339.

[19] Alekh Agarwal and John C. Duchi. "Distributed Delayed Stochastic Optimization". In: *Advances in Neural Information Processing Systems 24 (NIPS 2011)*. Curran Associates, Inc., 2011, pp. 873–881.

[20] Marcus Hutter et al. "Probabilities on Sentences in an Expressive Logic". In: *Journal of Applied Logic* 11.4 (2013), pp. 386–420. ISSN: 1570-8683. DOI: http://dx.doi.org/10.1016/j.jal.2013.03.003. URL: http://www.sciencedirect.com/science/article/pii/S157086831300013X.

[21] Abram Demski. "Logical Prior Probability". In: *Artificial General Intelligence. 5th International Conference, AGI 2012, Oxford, UK, December 8–11, 2012. Proceedings* 7716 (2012). Ed. by Joscha Bach, Ben Goertzel, and Matthew Iklé, pp. 50–59. DOI: 10.1007/978-3-642-35506-6_6.

[22] Haim Gaifman. "Concerning Measures in First Order Calculi". In: *Israel Journal of Mathematics* 2.1 (1964), pp. 1–18. DOI: 10.1007/BF02759729.

[23] Per Martin-Löf. "The Definition of Random Sequences". In: *Information and Control* 9.6 (1966), pp. 602–619.

[24] Anirban DasGupta. "Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics". In: New York, NY: Springer New York, 2011. Chap. Discrete Time Martingales and Concentration Inequalities, pp. 463–504. DOI: 10.1007/978-1-4419-9634-3_14.