FALL 2026 BAIS 7900

Lecture 2: Expert Advice, Hedge, Exponential Weights, and Regret Lower Bounds

Lecturer: *Mengxiao Zhang* Date: August 31st

1 Learning with expert advice

The expert-advice problem is one of the simplest nontrivial online-learning models. It is also one of the most useful. The learner does not need to predict the future directly. Instead, the learner has access to a collection of experts, models, heuristics, or policies, and must decide how much to trust each of them over time.

There are N experts; throughout, we assume $N \geq 2$, since the case $N = 1$ is trivial. At each round $t = 1, \ldots, T$,

1. the learner chooses a distribution $p_t \in \Delta(N)$;
2. the environment determines a loss vector $\ell_t \in [0,1]^N$ in an *adversarial* way;
3. the learner suffers the mixture loss $\langle p_t, \ell_t \rangle$ and then observes the entire vector ℓ_t.

Thus this is a *full-information* problem. After round t, the learner knows how every expert performed, including experts that received little or no weight.

The cumulative loss of expert i through round t is $L_{t,i} \triangleq \sum_{s=1}^{t} \ell_{s,i}$, with $L_{0,i} = 0$. The regret against the best fixed expert in hindsight is

$$\mathrm{Reg}_T \triangleq \sum_{t=1}^{T} \langle p_t, \ell_t \rangle - \min_{p \in \Delta(N)} \sum_{t=1}^{T} \langle p, \ell_t \rangle = \sum_{t=1}^{T} \langle p_t, \ell_t \rangle - \min_{i \in [N]} L_{T,i}.$$

The second equality follows because the comparator objective is linear in p, so an optimal point of the simplex can always be chosen to be a vertex e_i.

1.1 Examples

Example 1.1 (Weather forecasting)**.** Suppose four forecasting systems predict tomorrow's temperature. After the temperature is observed, expert i receives the normalized absolute-error loss $\ell_{t,i} = \min\{1, |\widehat{y}_{t,i} - y_t|/10\}$. If on one day $\ell_t = (0.1, 0.7, 0.2, 0.4)$ and the learner uses $p_t = (0.4, 0.1, 0.3, 0.2)$, then its mixture loss is $0.4(0.1) + 0.1(0.7) + 0.3(0.2) + 0.2(0.4) = 0.25$. The learner need not decide permanently which forecasting system is best. It can update the weights every day as evidence accumulates.

Example 1.2 (Choosing among business rules)**.** A firm may have N pricing rules, demand forecasts, inventory heuristics, or machine-learning models. If historical outcomes allow the firm to evaluate the loss that *each* rule would have incurred after the outcome is revealed, then the problem has expert-style full information. The experts do not need to be statistical estimators. They may be arbitrary policies supplied by different teams, vendors, or modeling assumptions.

Example 1.3 (Combining classification models)**.** Suppose a company has N existing classifiers for detecting fraudulent transactions. Before the true label y_t is known, model i outputs a prediction $\widehat{y}_{t,i}$. After the transaction is investigated and y_t is revealed, the learner can evaluate every model, for example using the zero-one loss $\ell_{t,i} = \mathbf{1}\{\widehat{y}_{t,i} \neq y_t\}$. Thus the learner receives full information even if only one model, or one randomized choice among the models, is used to make the operational decision. Expert advice lets the learner combine the models without assuming that the same model will remain best forever.

2 Why Follow the Leader can fail

A natural idea is to use the expert with the smallest cumulative loss so far. This is the *Follow the Leader* (FTL) strategy.

Algorithm 1: Follow the Leader (FTL)

Input: experts $[N]$
Initialize $L_{0,i} = 0$ for every $i \in [N]$;
for $t = 1, \dots, T$ **do**
  choose $I_t \in \arg\min_{i \in [N]} L_{t-1,i}$;
  play expert I_t and observe $\ell_t \in [0,1]^N$;
  update $L_{t,i} = L_{t-1,i} + \ell_{t,i}$ for every i;

FTL looks reasonable at first glance: if one expert has performed best in the past, why not use it now? The problem is that the environment the learner is facing can be *adversarial*! FTL reacts too sharply to small changes in the cumulative losses.

Example 2.1 (An oblivious sequence that makes FTL lose every round)**.** Take $N = 2$ and suppose FTL breaks ties in favor of expert 1. Consider the fixed loss sequence

$$\ell_t = \begin{cases} (1,0), & t \text{ odd}, \\ (0,1), & t \text{ even}. \end{cases}$$

The first few rounds are

t	L_{t-1}	FTL choice	ℓ_t	learner loss
1	$(0,0)$	1	$(1,0)$	1
2	$(1,0)$	2	$(0,1)$	1
3	$(1,1)$	1	$(1,0)$	1
4	$(2,1)$	2	$(0,1)$	1

The pattern repeats. FTL suffers loss T. If T is even, each fixed expert suffers exactly $T/2$, so $\mathrm{Reg}_T = T/2$. Thus FTL has linear regret even against an oblivious adversary.

Question 2.2. Does there exist a deterministic algorithm whose decision at round t depends only on the previous loss vectors $\ell_1, \dots, \ell_{t-1}$ and that guarantees sublinear regret? Does the answer change if the learner must select one expert rather than a distribution over experts?

The distinction between the two questions is essential. If the learner may output a distribution, deterministic no-regret algorithms do exist; Hedge will give our first example. If the learner must

choose one expert and that choice is deterministic, however, the answer is no. The failure is not special to FTL.

Proposition 2.3 (Deterministic pure-expert algorithms have linear regret)**.** *Fix $N \geq 2$. For every deterministic algorithm that chooses one expert $I_t \in [N]$ at each round using the past full-information feedback, there exists an oblivious sequence $\ell_1, \ldots, \ell_T \in \{0,1\}^N$ such that*

$$\mathrm{Reg}_T = \sum_{t=1}^{T} \ell_{t,I_t} - \min_{i \in [N]} \sum_{t=1}^{T} \ell_{t,i} \geq T\left(1 - \frac{1}{N}\right).$$

In particular, no deterministic pure-expert algorithm is no-regret.

Proof. As the learner is deterministic and the decision is concentrated on a single expert, once $\ell_1, \ldots, \ell_{t-1}$ have been specified, the next choice I_t is completely determined. We can therefore construct a bad sequence recursively *before the interaction begins*. After computing the expert I_t that the algorithm would choose from the already constructed history, set $\ell_{t,I_t} = 1$ and $\ell_{t,j} = 0$ for every $j \neq I_t$. Repeating this for $t = 1, \ldots, T$ produces one fixed loss sequence, so the resulting adversary is oblivious during the actual interaction.

The learner suffers loss 1 on every round, hence total loss T. On the other hand, exactly one expert receives loss 1 on each round, so $\sum_{i=1}^{N} L_{T,i} = T$. Therefore at least one expert satisfies $L_{T,i} \leq T/N$, which gives the claimed lower bound. □

Remark 2.4 (What does "deterministic algorithms fail" mean?). The proposition is about algorithms that must choose a *single expert*. It does not say that every deterministic map from the history to a distribution fails. We will see in the next section that there exists an online learning algorithm, Hedge, which computes p_t deterministically from the past and suffers the mixture loss $\langle p_t, \ell_t \rangle$, for which sublinear regret is possible. If an application requires executing one expert, randomization can be used to turn the distribution into one action without making the realized action predictable in advance.

3 Hedge and exponential weights

The failure of FTL suggests that we should not put all our weight on whichever expert happens to be the current leader. Hedge keeps the same basic principle: experts with smaller past loss should receive more weight. However, Hedge replaces the hard winner-take-all choice by a smooth distribution. The algorithm is shown below and is parameterized by a single parameter $\eta > 0$. At each round t, the algorithm maintains a weight $w_{t,i}$ for each expert i and assigns each expert probability proportional to $w_{t,i}$. After seeing the loss for each expert, the algorithm updates the weight of each expert with a factor of $\exp(-\eta \ell_{t,i})$.

Algorithm 2: Hedge / Exponential Weights

Input: learning rate $\eta > 0$
Initialize $w_{1,i} = 1$ for every $i \in [N]$;
for $t = 1, \ldots, T$ **do**
  set $p_{t,i} = w_{t,i} / \sum_{j=1}^{N} w_{t,j}$ for every i;
  play p_t and observe $\ell_t \in [0,1]^N$;
  update $w_{t+1,i} = w_{t,i} e^{-\eta \ell_{t,i}}$ for every i;

Unrolling the update gives $w_{t,i} = e^{-\eta L_{t-1,i}}$, so Hedge assigns probability proportional to $e^{-\eta L_{t-1,i}}$. A lower cumulative loss therefore means a larger probability. More importantly, Hedge changes the *relative* weight of two experts according to their relative performance: $p_{t,i}/p_{t,j} = e^{-\eta(L_{t-1,i} - L_{t-1,j})}$. If expert i loses less than expert j on the current round, the odds in favor of i increase on the next round.

Example 3.1 (One numerical Hedge update)**.** Take three experts with $\eta = 1$ and cumulative losses $L_{t-1} = (0, 1, 2)$. Their unnormalized weights are $(1, e^{-1}, e^{-2}) \approx (1, 0.368, 0.135)$, which gives $p_t \approx (0.665, 0.245, 0.090)$. If the new loss vector is $\ell_t = (1, 0, 0.4)$, the learner suffers mixture loss about 0.701. After the update, the cumulative losses become $(1, 1, 2.4)$. Experts 1 and 2 are now tied, so their next probabilities are equal. Hedge reacts substantially to one bad round for expert 1, but it does not permanently eliminate that expert.

At this point, Hedge is only a plausible heuristic. The important question is whether this smoothing is actually enough to prevent the linear-regret behavior of FTL. We answer this before trying to explain whether the exponential form itself is special.

3.1 Regret analysis of Hedge

The goal of the analysis is to compare two quantities that are defined very differently: the learner's cumulative mixture loss $\sum_t \langle p_t, \ell_t \rangle$ and the loss $L_{T,i^\star}$ of the best single expert in hindsight. The potential below is useful because it can be related to both quantities.

Theorem 3.2. *For any sequence $\ell_1, \dots, \ell_T \in [0,1]^N$ and any $\eta > 0$, Hedge satisfies*

$$\mathrm{Reg}_T \le \frac{\ln N}{\eta} + \eta \sum_{t=1}^{T} \sum_{i=1}^{N} p_{t,i} \ell_{t,i}^2 \le \frac{\ln N}{\eta} + \eta T.$$

Consequently, choosing $\eta = \sqrt{(\ln N)/T}$ gives $\mathrm{Reg}_T \le 2\sqrt{T \ln N}$.

The main message is already visible before the proof: even when the loss sequence is adversarial, Hedge eventually performs almost as well on average as the best fixed expert chosen after seeing the entire sequence. The dependence on the number of experts is only logarithmic.

Before going through the algebra, it helps to know what the proof is trying to do. There are in fact many different ways of proving the above theorem. What we will introduce is the *potential method*, which is classic in online learning and can be applied to different problems as you will see in later lectures (and homeworks). Define the potential function Φ_t at round t as follows:

$$\Phi_t \triangleq \frac{1}{\eta} \ln\Big(\sum_i e^{-\eta L_{t,i}}\Big).$$

We will use this same quantity in two ways. First, its one-round change will reveal the learner's loss because the normalized exponential terms are exactly the Hedge probabilities. Second, its final value can be compared with the best expert because the sum inside Φ_T contains the term $e^{-\eta L_{T,i^\star}}$. The proof works by connecting these two views of the same potential.

Proof. The one-step change of the potential is

$$\Phi_t - \Phi_{t-1} = \frac{1}{\eta} \ln\left(\frac{\sum_i e^{-\eta L_{t,i}}}{\sum_i e^{-\eta L_{t-1,i}}}\right) = \frac{1}{\eta} \ln\left(\sum_i p_{t,i} e^{-\eta \ell_{t,i}}\right).$$

The second equality is the first place where the particular Hedge probabilities enter the analysis. For $x \geq 0$, $e^{-x} \leq 1 - x + x^2$. Applying this with $x = \eta \ell_{t,i}$ and then using $\ln(1+u) \leq u$ gives

$$\Phi_t - \Phi_{t-1} \leq \frac{1}{\eta} \ln \left(1 - \eta \langle p_t, \ell_t \rangle + \eta^2 \sum_i p_{t,i} \ell_{t,i}^2 \right) \leq - \langle p_t, \ell_t \rangle + \eta \sum_i p_{t,i} \ell_{t,i}^2.$$

This is the key one-round inequality: the change of the potential controls the learner's current loss, up to a quadratic error term.

Summing over t telescopes the potential and yields

$$\sum_{t=1}^{T} \langle p_t, \ell_t \rangle \leq \Phi_0 - \Phi_T + \eta \sum_{t=1}^{T} \sum_{i=1}^{N} p_{t,i} \ell_{t,i}^2.$$

Initially all cumulative losses are zero, so $\Phi_0 = \frac{\ln N}{\eta}$. Let $i^\star \in \arg\min_i L_{T,i}$. Because the sum defining Φ_T contains the contribution of expert $i^\star$, we have $\Phi_T \geq \frac{1}{\eta} \ln(\exp(-\eta L_{T,i^\star})) = -L_{T,i^\star}$. Substituting these two bounds into the previous inequality and subtracting $L_{T,i^\star}$ gives the first claimed regret bound. Finally, $0 \leq \ell_{t,i} \leq 1$ implies $\sum_i p_{t,i} \ell_{t,i}^2 \leq 1$, which gives the worst-case bound. □

Remark 3.3. The potential method is a very powerful and universal method in online learning. The key is how to find such a potential function :) Here Φ_t acts as a bridge. Its one-step change is related to the learner's loss, while its final value is related to the best expert's cumulative loss. The two terms $\ln N/\eta$ and ηT can then be read as the price of initially not knowing which expert will be best and the price of reacting too aggressively to new losses.

Question 3.4. If $\ell_t \in [-1, 1]^N$, what will break in the above proof? Can you fix it with a small modification of the proof?

3.2 Do we really need exponential weights?

Now that we have seen the proof, we can return to a design question that would have been difficult to answer before the analysis: *why did we use an exponential function, and is it necessary?*

The proof shows several reasons that the exponential is convenient. Cumulative losses add over time, while the exponential converts this additive history into a multiplicative update. After normalization, the ratio of two weights depends only on the difference of their cumulative losses, so adding the same constant to every expert's loss does not change the decision. Most importantly for the proof, the ratio of two consecutive potentials becomes $\sum_i p_{t,i} e^{-\eta \ell_{t,i}}$, in which the same probabilities p_t used by the learner appear automatically. The elementary approximation $e^{-x} \leq 1 - x + x^2$ then turns this expression into the learner's mixture loss plus a controllable error term.

These observations explain why exponential weights are particularly clean, but they do *not* show that the exponential is the only possible no-regret update. There are other options. First, the proof itself suggests a more useful alternative. When $\eta \ell$ is small, $e^{-\eta \ell}$ is approximately $1 - \eta \ell$. What happens if we replace the exponential multiplier by this linear one? We will analyze exactly this idea in the next section.

There is also a conceptually different possibility: instead of deciding weights from an expert's absolute cumulative loss, we can ask which experts we currently regret not following more often. Define the instantaneous regret to expert i as $r_{t,i} \triangleq \langle p_t, \ell_t \rangle - \ell_{t,i}$ and the cumulative regret as

$R_{t,i} \triangleq \sum_{s=1}^{t} r_{s,i}$. If $R_{t,i} > 0$, then expert i has beaten the learner so far. *Regret Matching* [1] assigns more probability to experts with larger positive cumulative regret, roughly using weights proportional to $[R_{t,i}]_+$. *Regret Matching+* [2] uses the same idea but clips the running regret at zero after every update, so a large negative history does not have to be worked off before an expert can become relevant again.

Regret Matching and Regret Matching+ therefore start from a different intuition than Hedge. Hedge asks, "which experts have accumulated small loss?" Regret Matching asks, "which experts have I most regretted not following?" We will not analyze these algorithms here. A useful exercise is to determine whether this positive-regret rule is enough to guarantee sublinear regret, and how its dependence on N compares with Hedge.

4 Another multiplicative update: Prod

The exponential function is not the only way to update weights multiplicatively. A particularly simple alternative, often called *Prod*, replaces the exponential factor $e^{-\eta \ell_{t,i}}$ by its linear analogue $1 - \eta \ell_{t,i}$ [4]. The complete pseudocode is shown as follows.

Algorithm 3: Prod / Linear Multiplicative Weights

Input: learning rate $0 < \eta \le 1/2$
Initialize $w_{1,i} = 1$ for every $i \in [N]$;
for $t = 1, \dots, T$ **do**
  set $p_{t,i} = w_{t,i} / \sum_{j=1}^{N} w_{t,j}$ for every i;
  play p_t and observe $\ell_t \in [0,1]^N$;
  update $w_{t+1,i} = w_{t,i}(1 - \eta \ell_{t,i})$ for every i;

Because $0 < \eta \le 1/2$ and $\ell_{t,i} \in [0,1]$, all weights remain nonnegative. Writing the update directly in terms of probabilities gives $p_{t+1,i} = p_{t,i}(1 - \eta \ell_{t,i})/(1 - \eta \langle p_t, \ell_t \rangle)$. Thus an expert with larger current loss loses a larger fraction of its weight.

Example 4.1. Suppose $p_t = (0.5, 0.3, 0.2)$, $\ell_t = (0, 1, 0.5)$, and $\eta = 0.4$. The unnormalized next probabilities are proportional to $(0.5, 0.3(0.6), 0.2(0.8)) = (0.5, 0.18, 0.16)$, so after normalization $p_{t+1} \approx (0.595, 0.214, 0.190)$. Like Hedge, Prod shifts weight away from experts that just performed poorly, but it uses a linear multiplicative factor instead of an exponential one.

The following theorem shows the regret guarantee of the Prod algorithm, which is very similar to the one proved for Hedge.

Theorem 4.2. *For any $0 < \eta \le 1/2$ and every expert $i \in [N]$, Algorithm 3 guarantees that*

$$\sum_{t=1}^{T} \langle p_t, \ell_t \rangle - L_{T,i} \le \frac{\ln N}{\eta} + \eta \sum_{t=1}^{T} \ell_{t,i}^2.$$

In particular, if $i^\star \in \arg\min_i L_{T,i}$, then

$$\mathrm{Reg}_T \le \frac{\ln N}{\eta} + \eta \sum_{t=1}^{T} \ell_{t,i^\star}^2 \le \frac{\ln N}{\eta} + \eta T.$$

Hence, whenever $T \ge 4 \ln N$, the choice $\eta = \sqrt{(\ln N)/T}$ gives $\mathrm{Reg}_T \le 2\sqrt{T \ln N} = \mathcal{O}(\sqrt{T \ln N})$.

Proof. We again use a potential-function argument. Let $W_t \triangleq \sum_{i=1}^{N} w_{t,i}$, and define the potential after round t by $\Phi_t \triangleq \frac{1}{\eta} \ln W_{t+1}$ for all $t = 0, \ldots, T$.

Since $W_1 = N$, we have $\Phi_0 = \frac{\ln N}{\eta}$.

By the update rule,

$$W_{t+1} = \sum_i w_{t,i}(1 - \eta \ell_{t,i}) = W_t \left(1 - \eta \left\langle p_t, \ell_t \right\rangle\right).$$

Therefore the one-step change of the potential is

$$\Phi_t - \Phi_{t-1} = \frac{1}{\eta} \ln \left(\frac{W_{t+1}}{W_t} \right) = \frac{1}{\eta} \ln \left(1 - \eta \left\langle p_t, \ell_t \right\rangle\right) \ \leq - \left\langle p_t, \ell_t \right\rangle,$$

where the last inequality uses $\ln(1 - x) \leq -x$.

Summing over t telescopes the potential and gives

$$\sum_{t=1}^{T} \left\langle p_t, \ell_t \right\rangle \leq \Phi_0 - \Phi_T.$$

It remains to relate the final potential to the cumulative loss of any fixed expert. Fix $i \in [N]$. Since $W_{T+1} \geq w_{T+1,i}$, we know that

$$\Phi_T \geq \frac{1}{\eta} \ln w_{T+1,i} = \frac{1}{\eta} \sum_{t=1}^{T} \ln(1 - \eta \ell_{t,i}).$$

Because $0 < \eta \leq 1/2$ and $\ell_{t,i} \in [0, 1]$, we have $0 \leq \eta \ell_{t,i} \leq 1/2$. For $0 \leq x \leq 1/2$,

$$\ln(1 - x) \geq -x - x^2.$$

Hence, we have

$$\Phi_T \geq - \sum_{t=1}^{T} \ell_{t,i} - \eta \sum_{t=1}^{T} \ell_{t,i}^2 = -L_{T,i} - \eta \sum_{t=1}^{T} \ell_{t,i}^2.$$

Combining this lower bound on Φ_T with $\Phi_0 = \ln N / \eta$ yields

$$\sum_{t=1}^{T} \left\langle p_t, \ell_t \right\rangle \leq \frac{\ln N}{\eta} + L_{T,i} + \eta \sum_{t=1}^{T} \ell_{t,i}^2.$$

Rearranging gives the result. The remaining claims follow by choosing $i = i^\star$ and using $\ell_{t,i^\star}^2 \leq 1$. □

Remark 4.3 (Two multiplicative updates). Hedge and Prod both decrease the weight of an expert multiplicatively and both give the same worst-case order $\sqrt{T \ln N}$. Their analyses, however, leave different quantities before we simplify them by the crude bound $\ell_{t,i}^2 \leq 1$.

Question 4.4. Compare the strongest bounds proved above for Hedge and Prod *before* replacing their quadratic terms by T. Which squared-loss quantity appears in the Hedge guarantee, and which squared-loss quantity appears in the Prod guarantee? Can these two quantities be very different on the same instance?

5 How good is $\sqrt{T \ln N}$? Minimax regret

The upper bounds above show what particular algorithms can guarantee. They do not tell us whether a fundamentally better algorithm exists. To answer that question, we turn to minimax regret and prove lower bounds that apply to every online algorithm.

Definition 5.1. *For the expert problem with an oblivious adversary, define*

$$V_{T,N} \triangleq \inf_{\mathcal{A}} \sup_{\ell_1,\dots,\ell_T \in [0,1]^N} \mathbb{E}_{\mathcal{A}} \left[\mathrm{Reg}_T\right],$$

where the infimum is over all legitimate online algorithms and the expectation is over the learner's internal randomness.

Hedge gives $V_{T,N} \le 2\sqrt{T \ln N}$, while the trivial bound $\mathrm{Reg}_T \le T$ gives $V_{T,N} \le T$. To show that the $\sqrt{T \ln N}$ rate is optimal in the regime where it is below the linear scale, we need the opposite statement: every online algorithm must incur large regret on at least one fixed loss sequence.

A convenient lower-bound strategy is to specify a distribution over loss sequences and analyze every learner under this randomized environment. The randomness is only a proof device. If the average regret over the random draw is large, then at least one deterministic realization of the loss sequence must already be hard.

Remark 5.2. Fix any algorithm $\mathcal{A}$ and draw the entire loss sequence in advance from a distribution $\mathcal{D}$. If $\mathbb{E}_{\ell_{1:T} \sim \mathcal{D}, \mathcal{A}}[\mathrm{Reg}_T] \ge r$, then averaging implies that at least one deterministic sequence in the support of $\mathcal{D}$ satisfies $\mathbb{E}_{\mathcal{A}}[\mathrm{Reg}_T] \ge r$. Therefore a randomized construction can establish a worst-case lower bound while the resulting hard sequence is still completely fixed before the interaction begins.

5.1 A rigorous two-expert lower bound

We first isolate the basic $\sqrt{T}$ difficulty using only two experts. Consider the following randomized oblivious environment: before the interaction begins, draw $\ell_{t,1}, \ell_{t,2} \sim \mathrm{Bernoulli}(1/2)$ independently across experts and rounds. Equivalently, the entire $T \times 2$ loss matrix is sampled in advance.

Now consider any online algorithm, possibly randomized. Conditional on the full history before round t and on the learner's current distribution p_t, the round-t losses are still fresh fair Bernoulli variables. Therefore $\mathbb{E}[\langle p_t, \ell_t \rangle \mid \text{past}, p_t] = 1/2$. Summing over rounds shows that every algorithm has expected cumulative loss exactly $T/2$ under this environment.

Let $B_1 \triangleq L_{T,1}$ and $B_2 \triangleq L_{T,2}$. Then B_1 and B_2 are independent $\mathrm{Binomial}(T, 1/2)$ random variables. The learner always pays $T/2$ in expectation, but the hindsight comparator benefits from whichever expert happens to receive the smaller cumulative loss. Using $\min\{B_1, B_2\} = (B_1 + B_2 - |B_1 - B_2|)/2$ and $\mathbb{E}[B_1 + B_2] = T$, we obtain $\mathbb{E}[\mathrm{Reg}_T] = \frac{1}{2} \cdot \mathbb{E}[|B_1 - B_2|]$. It therefore remains to show that the difference of two independent binomial variables has magnitude of order $\sqrt{T}$ with constant probability. The following inequality lower bounds the probability that a nonnegative random variable is not too small, in terms of its first two moments.

Lemma 5.3 (Paley–Zygmund inequality)**.** *If $Z \ge 0$ has finite second moment, then for every $\theta \in (0,1)$,*

$$\mathbb{P}\left(Z \ge \theta \mathbb{E}\left[Z\right]\right) \ge (1-\theta)^2 \frac{\left(\mathbb{E}\left[Z\right]\right)^2}{\mathbb{E}\left[Z^2\right]}.$$

Proof. Define the event $A = \{Z \geq \theta \mathbb{E}[Z]\}$. Then

$$\mathbb{E}\left[Z\right] = \mathbb{E}\left[Z\mathbf{1}\{A^c\}\right] + \mathbb{E}\left[Z\mathbf{1}\{A\}\right] \leq \theta \mathbb{E}\left[Z\right] + \mathbb{E}\left[Z\mathbf{1}\{A\}\right].$$

Therefore

$$(1-\theta)\mathbb{E}\left[Z\right] \leq \mathbb{E}\left[Z\mathbf{1}\{A\}\right] \leq \sqrt{\mathbb{E}\left[Z^2\right]\mathbb{P}(A)},$$

where the last step is Cauchy–Schwarz. Squaring both sides and rearranging proves the claim. □

Theorem 5.4 (Two-expert regret lower bound)**.** *There is a universal constant $c > 0$ such that, for every online algorithm, there exists a deterministic loss sequence with $\mathbb{E}[\mathrm{Reg}_T] \geq c\sqrt{T}$. In particular, $V_{T,2} = \Omega(\sqrt{T})$.*

Proof. Define $Y_t \triangleq \ell_{t,1} - \ell_{t,2}$ and $D_T \triangleq B_1 - B_2 = \sum_{t=1}^{T} Y_t$. The variables Y_t are independent with $\mathbb{P}(Y_t = 1) = \mathbb{P}(Y_t = -1) = 1/4$ and $\mathbb{P}(Y_t = 0) = 1/2$. Hence $\mathbb{E}[Y_t] = 0$, $\mathbb{E}[Y_t^2] = 1/2$, and $\mathbb{E}[Y_t^4] = 1/2$. By independence and the zero mean, $\mathbb{E}[D_T^2] = T/2$. Expanding the fourth moment gives

$$\begin{aligned}
\mathbb{E}\left[D_T^4\right] &= \mathbb{E}\left[\left(\sum_{t=1}^{T} Y_t\right)^4\right] \\
&= \sum_{t=1}^{T} \mathbb{E}\left[Y_t^4\right] + 4 \sum_{\substack{s,t\in[T] \\ s\neq t}} \mathbb{E}\left[Y_s^3 Y_t\right] + 6 \sum_{1\leq s<t\leq T} \mathbb{E}\left[Y_s^2 Y_t^2\right] \\
&\quad + 12 \sum_{r=1}^{T} \sum_{\substack{1\leq s<t\leq T \\ s,t\neq r}} \mathbb{E}\left[Y_r^2 Y_s Y_t\right] + 24 \sum_{1\leq q<r<s<t\leq T} \mathbb{E}\left[Y_q Y_r Y_s Y_t\right] \\
&= \sum_{t=1}^{T} \mathbb{E}\left[Y_t^4\right] + 6 \sum_{1\leq s<t\leq T} \mathbb{E}\left[Y_s^2 Y_t^2\right] \\
&= T\mathbb{E}\left[Y_1^4\right] + 6\binom{T}{2}\left(\mathbb{E}\left[Y_1^2\right]\right)^2 = \frac{T}{2} + \frac{3T(T-1)}{4} \leq \frac{3T^2}{4}.
\end{aligned}$$

Here the terms containing an odd power of some Y_t vanish by independence and $\mathbb{E}[Y_t] = 0$, while for $s \neq t$, independence gives $\mathbb{E}[Y_s^2 Y_t^2] = \mathbb{E}[Y_s^2]\mathbb{E}[Y_t^2]$. Apply the previous lemma to $Z = D_T^2$ with $\theta = 1/2$. Since $\mathbb{E}[Z] = T/2$, it gives $\mathbb{P}(|D_T| \geq \sqrt{T}/2) \geq 1/12$. Therefore $\mathbb{E}[|D_T|] \geq (\sqrt{T}/2)(1/12) = \sqrt{T}/24$. Since $\mathbb{E}[\mathrm{Reg}_T] = (1/2)\mathbb{E}[|D_T|]$ under the Bernoulli environment, $\mathbb{E}[\mathrm{Reg}_T] \geq \sqrt{T}/48$. Finally, by the randomized-environment argument, at least one deterministic realization of the entire loss sequence causes at least this much expected regret for the given algorithm. □

Remark 5.5 (Where the $\sqrt{T}$ comes from). The learner faces fresh fair noise on every round, so no strategy can reduce its conditional expected loss below $1/2$. The comparator is chosen only after all T rounds, however, and therefore benefits from whichever expert happened to be luckier. The gap between two independent cumulative coin-flip losses fluctuates on the $\sqrt{T}$ scale, and the hindsight comparator captures exactly this fluctuation.

5.2 The N-expert lower bound and the origin of $\sqrt{\ln N}$

Now repeat the same construction with N experts: draw $\ell_{t,i} \sim \text{Bernoulli}(1/2)$ independently across $i \in [N]$ and $t \in [T]$. As before, every online algorithm has expected cumulative loss $T/2$, while the $L_{T,i}$ are independent Binomial$(T, 1/2)$ random variables. Hence the expected regret is exactly

$$\mathbb{E}\left[\text{Reg}_T\right] = \frac{T}{2} - \mathbb{E}\left[\min_{i\in[N]} L_{T,i}\right] = \mathbb{E}\left[\max_{i\in[N]}\left(\frac{T}{2} - L_{T,i}\right)\right].$$

This identity reduces the lower bound to an extreme-value question: how far below its mean can the luckiest among N independent experts fall? The learner cannot predict this lucky expert in advance, while the hindsight comparator is allowed to select it after observing all T rounds.

Lemma 5.6 (Binomial extreme-value estimate)**.** *There are universal constants $c_0, c_1, C > 0$ such that, for $2 \le N \le e^{c_0 T}$,*

$$c_1\sqrt{T\ln N} \le \mathbb{E}\left[\max_{i\in[N]}\left(\frac{T}{2} - L_{T,i}\right)\right] \le C\sqrt{T\ln N},$$

where $L_{T,1}, \dots, L_{T,N}$ are independent Binomial$(T, 1/2)$ *random variables.*

The two-expert theorem established the $\sqrt{T}$ effect using only moment calculations. To obtain the additional $\sqrt{\ln N}$ factor, we need quantitative control of the *lower* tail of a binomial random variable. The calculation is more technical, but its interpretation is simple. For deviations u that are not too large relative to T, binomial lower tails have the exponential scale

$$\mathbb{P}\left(L_{T,i} \le \frac{T}{2} - u\right) \approx \exp\left(-c\frac{u^2}{T}\right).$$

Because the N experts are independent,

$$\mathbb{P}\left(\min_i L_{T,i} > \frac{T}{2} - u\right) = \left(1 - \mathbb{P}\left(L_{T,1} \le \frac{T}{2} - u\right)\right)^N.$$

The minimum typically reaches the level $T/2 - u$ once the probability that a single expert reaches this level is roughly $1/N$. Equating the two exponential scales, $\exp(-cu^2/T) \approx 1/N$, gives $u \asymp \sqrt{T\ln N}$. This is the source of the extra $\sqrt{\ln N}$ factor. A rigorous proof of the lemma follows from standard binomial lower-tail estimates; see, for example, the treatment of minimax regret in [3] and the non-asymptotic refinement in [7].

Theorem 5.7 (Minimax lower bound for expert advice)**.** *There is a universal constant $c > 0$ such that, in the regime $2 \le N \le e^{cT}$,*

$$V_{T,N} \ge c\sqrt{T\ln N}.$$

Together with the Hedge upper bound, this shows that $V_{T,N} = \Theta(\sqrt{T\ln N})$ up to universal constant factors throughout this regime.

Remark 5.8 (Why $\ln N$ appears). Increasing N does not make any individual expert easier to predict: every expert still has mean loss $1/2$. Instead, it strengthens the hindsight benchmark because the comparator may select the luckiest expert from a larger collection. The factor $\sqrt{\ln N}$ is precisely the statistical gain from this extreme-value selection. The restriction $N \le e^{cT}$ is also natural: regret is always at most T, so once $\ln N$ becomes comparable to or larger than T, the scale $\sqrt{T\ln N}$ itself approaches the linear regime.

References

[1] Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[2] Oskari Tammelin. Solving large imperfect information games using CFR+. arXiv:1407.5042, 2014.

[3] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[4] Nicolò Cesa-Bianchi, Yishay Mansour, and Gilles Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2–3):321–352, 2007.

[5] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

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

[7] Francesco Orabona and Dávid Pál. Optimal non-asymptotic lower bound on the minimax regret of learning with expert advice. arXiv:1511.02176, 2015.