ECE 8803: Online Decision Making in Machine Learning **Fall 2021**

Lecture 5: September 8 and September 13

Lecturer: Vidya Muthukumar

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

Last week, we learned about the fundamental tradeoff involved in prediction of a sequence in the worst case: between leveraging the information present in past data, and randomizing to avoid being exploited by an adversarially/maliciously designed sequence. We then introduced an explicit algorithm (seemingly out of nowhere) to provably navigate this tradeoff and achieve a $\mathcal{O}(\sqrt{T})$ regret with respect to the *best constant predictor* in hindsight, i.e. the predictor that gets to see the entire sequence and either predict 0 on every round, or 1 at every round — whichever choice makes fewer errors.

In these two lectures, we will learn about an entire family of algorithms that also optimally navigates this tradeoff. At the heart of this family of algorithms is a somewhat different approach that essentially follows the leader *with random perturbations added.* This algorithm is popularly called *Follow-the-Perturbed-Leader*, and is introduced below.

5.1. Recap

We first recap our relevant notion for binary sequence prediction and regret. We introduced the binary sequence prediction paradigm in which we aim to predict the next realization of a sequence, X_t, from the past realizations, denoted by $X^{t-1} : -\{X_1, \ldots, X_{t-1}\}$. Our prediction is denoted by $\widehat{X}_t$ (which, in general, can be random), and our loss function is given by $\ell(\widehat{X}_t; X_t)$. The overall goal is to choose a prediction strategy $f_{\mathsf{predict}}(\cdot)$ to minimize the total loss $H_T := \mathbb{E}\left[\sum_{t=1}^{T} \ell(\widehat{X}_t, X_t)\right]$, where the expectation is with respect to the randomness in the algorithm only. Importantly, the sequence $\{X_1, \ldots, X_T\}$ can be completely arbitrary, even generated *adversarially* to the prediction process, i.e. in order to try and maximize H_T. We introduced the metric of performance of *regret* with respect to the best constant predictor in hindsight (who is able to see the entire stream of data at once) as

$$R_T(X^T) := \underbrace{H_T}_{\text{our algorithm}} - \underbrace{\min_{x \in \{0,1\}} \sum_{t=1}^{T} \ell(x; X_t)}_{\text{best fixed prediction}} . \tag{5.1}$$

Henceforth, we denote $\min_{x \in \{0,1\}} \sum_{t=1}^{T} \ell(x; X_t) := L_T^*$ as shorthand.

It is also convenient to denote the cumulative losses incurred by each letter as

$$L_{t,x} := \sum_{s=1}^{t} \mathbb{I}[X_s \neq x] \text{ for each } x \in \{0,1\},$$

and also the loss incurred at round t by each letter as $l_{t,x} := \mathbb{I}[X_t \neq x]$ (so $L_{t,x} = \sum_{s=1}^{t} l_{s,x}$). Under this notation, it is worth noting that we can write $L_T^* := \min\{L_{T,0}, L_{T,1}\}$. Therefore, the best fixed/constant prediction in hindsight does the following: it looks at the entire sequence all at once, and picks the letter that minimizes the total loss, which simply corresponds to the letter that appeared more often (since we are measuring loss using the 0-1 loss function).

5.2. The Follow-the-Perturbed-Leader algorithm

In Lecture 3 (August 30), we introduced the Follow-the-Leader (FTL) algorithm, which was given by

$$\widehat{X}_t := \arg\min_{x \in \{0,1\}} L_{t-1,x}, \tag{5.2}$$

in other words, *predicting the letter out of* $x \in \{0,1\}$ *that was seen more often thus far.* We saw that FTL performed very well against a stochastic Bernoulli sequence, but very poorly against an adversarially designed sequence as it would incur a loss of 1 on every round. (In HW, you additionally showed that this implies linear in T regret.)

The central issue with FTL is that it is too *sensitive* to the past data. Recall the adversarial sequence that was designed for FTL, which was given by $101010\ldots$. We saw that this example induces a change of the identity of the leader on every single round! This instability of FTL is the central reason behind its high loss (and therefore, high regret) in a malicious setting. We now examine a very simple approach to mitigate this instability. This constitutes continuing to follow the leader, but adding a significant *random perturbation* to each of the losses at every round. Concretely, we draw random variables $N_{t,x}$ that are independently and identically distributed (iid) across $t = 1, \ldots, T$ *and* $x \in \{0,1\}$, and we add them to the cumulative losses for each letter. We then follow this perturbed leader in the following way:

$$\widehat{X}_t := \arg\min_{x \in \{0,1\}} \left[L_{t-1,x} + N_{t,x}\right]. \tag{5.3}$$

Clearly, $\widehat{X}_t$ is randomized because $N_{t,x}$'s are random. We can continue to denote $\widehat{P}_t := \mathbb{P}[\widehat{X}_t = 1]$, although we will not work with these probabilities directly in this set of lectures[1].

Let us now elaborate on the distribution that we choose for each of the perturbations given by $N_{t,x}$. We will consider $N_{t,x}$ to be an exponential random variable[2] with parameter equal to η; in other words, the PDF of $N_{t,x}$ is given by $p(n) := \eta \exp(-\eta n)$ for all $n \geq 0$. The parameter η is called a learning rate and its choice critically impacts the behavior of the FTPL algorithm. To see this, we note that the variance of an exponential random variable is given by $\frac{1}{\eta^2}$. Now, we can consider two extremes:

- When $\eta \to \infty$, we have variance$(N_{t,x}) = 0$, and so there is no noise introduced into the algorithm. FTPL becomes exactly FTL.

1. You did work this out in HW 1, Problem 4 for a different choice of distribution on $N_{t,x}$, which turns out to reduce to MWA.
2. See `https://en.wikipedia.org/wiki/Exponential_distribution` for basic factoids on the exponential distribution, including weird quantities like kurtosis.

- When $\eta \to 0$, we have variance$(N_{t,x}) = \infty$, and so the effect of the random perturbation completely "drowns out" the effect of the cumulative losses. Effectively, FTPL becomes like "pure guessing" in this case, making it equally likely to predict 0 or 1 on every round.

This intuition tells us that the smaller the value of η, the more randomization we are introducing into the algorithm. In fact, this parameter η is actually very related to the parameter that was chosen in MWA. Moreover, FTPL constitutes a *family* of algorithms: the exponential distribution is not the only choice we could have made. See the bibliographical notes for more details on possible alternative choices of distribution for the random perturbations.

5.3. Why it works

We will now see, through a series of steps, why FTPL can ensure low regret. We begin with a critical simplification of Equation (5.3) and go through our two canonical examples to get some intuition for what FTPL really does.

5.3.1 A critical simplification: Random perturbation at the start

Recall that we measure our performance through *expected* loss, given by $H_T := \mathbb{E}\left[\sum_{t=1}^{T} \ell(\widehat{X}_t; X_t)\right]$, where the expectation is taken over the randomness in the predictions $\{\widehat{X}_t\}$. We first notice that FTPL is identical in expectation to running FTL on a *loss sequence* of length $(T+1)$ given by:

$$\{N_x, l_{1,x}, \dots, l_{T,x}\} \text{ for each } x \in \{0,1\},$$

where for each $x \in \{0,1\}$, N_x is a single random variable (drawn according to the same distribution as what was earlier $N_{t,x}$). Note that this essentially yields updates given by

$$\widehat{X}_t = \arg\min_{x \in \{0,1\}} \left[L_{t-1,x} + N_x\right]. \tag{5.4}$$

Exercise 1 (Optional) *Verify that the predictions generated by Equation* (5.3) *and* (5.4) *would yield the exact same expected total loss, given by H_T, for any sequence $\{X_t\}_{t=1}^T$ that is fixed in advance and does not depend on the value of N_x. Use linearity of expectation (i.e. $\mathbb{E}[A+B] = \mathbb{E}[A] + \mathbb{E}[B]$), and a calculation of $\widehat{P}_t$ in terms of the cumulative distribution function (CDF) of an appropriate random variable.*

For the rest of the lecture, we will work with this simplified version of FTPL with "one-shot" noise for the reason that it preserves expected performance[3]. The same intuition

3. There is a caveat that we will not discuss in lecture that can lead to differing performance between these variants of FTPL. The high-level reason is as follows: in the "one-shot" version of FTPL, the same realization of the noise N_x influences the decisions made on all rounds. Therefore, an adversary that is allowed to adapt *after* the algorithm has started running would be able to intuit this value of N_x by observing past predictions made, and eventually exploit it. Such an adversary is more commonly called a "non-oblivious" adversary in the literature. In fact, the principal reason for needing the perturbations to be drawn independently at round t is precisely to avoid such exploitation by an adaptive adversary.

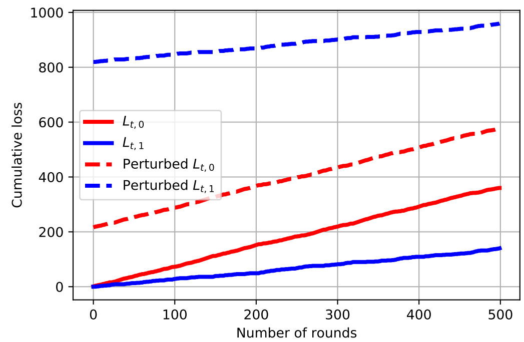


Figure 5.1: Visualizing the impact of a random perturbation on the evolution of the cumulative losses for an iid Bernoulli(0.7) sequence. Notice that the random perturbation's effect is so strong that it flips the identity of the best predictor from 1 to 0, which is a totally undesired outcome!

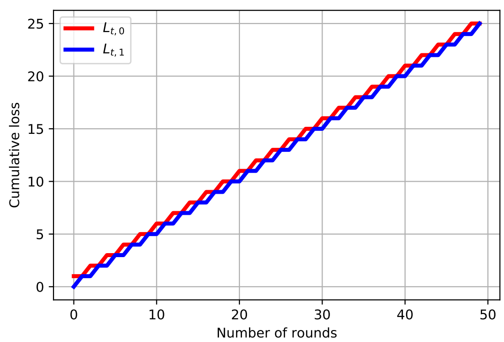
for how increased variance (randomization) in N_x's influences performance continues to hold: when $\eta \to \infty$, $N_x = 0$ and the algorithm reduces to FTL; but when $\eta \to 0$, the variance of N_x goes to ∞ and will dominate the algorithm's performance.

Just as we saw in the analysis of MWA, we neither want to over-exploit the true information present in the sequence, nor do we want to over-randomize. We will now visualize the impact of adding high-variance random perturbations $\{N_x\}_{x \in \{0,1}$ through our two canonical examples. In both examples, we will consider the total number of rounds $T = 500$.

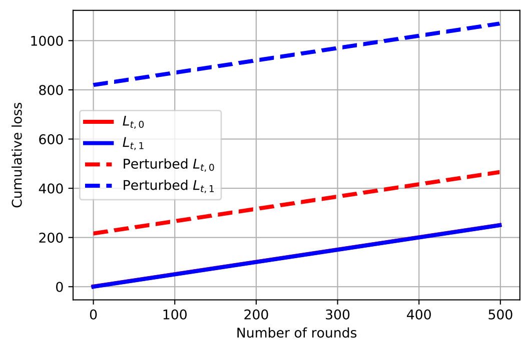
Example 1 (iid Bernoulli(0.7) sequence) *We start with our canonical example of the iid Bernoulli sequence, and first plot (for a* typical *realization of the sequence) the cumulative losses of predicting* 0 *and* 1 *(respectively marked in solid red and solid blue) as a function of the number of rounds played. Notice that we consistently have* $L_{t,1} < L_{t,0}$ *(in fact, the gap between them grows with* t*). Therefore, after a "warm-start" period*[4], 1 *is always the leader (and is also the best in hindsight). The solid lines in Figure 5.1 illustrate why FTL is such a good algorithm on this sequence: it will figure out very quickly that* 1 *is the correct letter to predict, and it will always predict it.*

Now, we evaluate what will happen when we add perturbations N_0 *and* N_1 *to the respective loss sequences. In particular, we will consider a special realization of the random variables,* N_0 *and* N_1*, drawn from the exponential distribution with parameter* $\eta = 0.01$*, such that* N_1 *is really large but* N_0 *is really small. The dashed lines (in red and blue respectively) in Figure 5.1 plot the cumulative losses of predicting* 0 *and* 1 that are perturbed by N_0 and N_1 respectively. *The figure shows that, unfortunately, these perturbations are adversely large because they create a completely different outcome: now, because* 0 *is always leading, FTPL will always predict* 0 *on every round! This is the complete opposite of the outcome that we want: the original sequence had* 1 *as the leader on almost all rounds. Thus, for*

4. You characterized the length of this warm-start period in HW 1, Problem 2.



(a) Original.



(b) Original + perturbed

Figure 5.2: Visualizing the impact of a random perturbation on the evolution of the cumulative losses for a sequence of alternating 1's and 0's. Notice that the random perturbation's effect is to separate the cumulative losses, which are intricately intertwined in the absence of the perturbation. Thus, the perturbation in this case has the positive effect of "stabilizing" the FTPL updates. Note the difference in y-axis scales between figure (a) and (b).

this realization of the random perturbation FTPL will, in fact, incur linear in T regret. It can be shown formally that the probability of this kind of an adverse realization occurring is actually significant (in particular, not close to 0) when the perturbations are of high variance. Therefore, the expected regret (over the randomness in the algorithm) can be shown to be linear.

Example 2 (Alternating 1's and 0's) *We now consider our other canonical example of alternating 1's and 0's, which was observed to make FTL incur a loss of 1 on every single round. Similar to the previous example, Figure 5.2 plots $L_{t,0}$ and $L_{t,1}$ (in solid red and blue respectively) versus t. You can see from Figure 5.2a that the identity of the leader keeps changing from round to round; therefore, FTL is a highly unstable algorithm and (like we saw in class) will incur linear regret.*

We now consider the effect of adding perturbations N_0 and N_1. As before, we will consider a special realization of the random variables, N_0 and N_1, drawn from the exponential distribution with parameter $\eta = 0.01$, such that N_1 is really large but N_0 is really small. You can see from Figure 5.2b that this disparate perturbation now has the positive effect of "separating out" the evolution of cumulative loss: in particular, 0 is always less than 1 in the perturbed loss sequence, and so FTPL will always predict 0. Thus, FTPL ***stabilizes*** *the prediction updates. Since the best predictor in hindsight can be either 0 or 1, this stabilization ensures that the regret will be low (and would be low even in the complementary case where instead N_0 were really large, but N_1 were really small).*

The two examples highlight the inherent tension in deciding how much perturbation to add to the algorithm. On one hand, Example 2 shows that if the extent of perturbation is insufficient, the updates remain unstable to an adversary (as in the original FTL algorithm).

On the other hand, Example 1 shows that if the extent of perturbation is too large, it can change the entire outcome of prediction. Nevertheless, we will now show that we can effectively trade off these two effects and achieve the optimal $\mathcal{O}(\sqrt{T})$ regret guarantee, as below.

Theorem 1 *FTPL with the learning rate $\eta = \frac{1}{\sqrt{T}}$ achieves $R_T = \mathcal{O}(\sqrt{T})$ on any sequence.*

The rest of the note will be devoted to proving Theorem 1 by bounding these two effects. To quantify this tradeoff better, we now introduce an intermediate term, which is the loss of the best predictor-in-hindsight *on the perturbed losses*. We define this by $L^*_{T,\mathsf{pert.}} := \min_{x \in \{0,1\}} [L_{T,x} + N_x]$. Then, it is easy to see that we can decompose the regret as

$$R_T := H_T - L^*_T = \underbrace{H_T - \mathbb{E}\left[L^*_{T,\mathsf{pert.}}\right]}_{\mathbb{E}[R^A_T]} + \underbrace{\mathbb{E}\left[L^*_{T,\mathsf{pert.}}\right] - L^*_T}_{\mathbb{E}[R^B_T]} \tag{5.5}$$

Above, $\mathbb{E}[R^A_T]$ reflects the regret of FTL on the *perturbed loss sequence*, and as we will show, measures how stable the perturbations make the updates of FTL. On the other hand, $\mathbb{E}[R^B_T]$ reflects the difference between the loss incurred by the best predictor on hindsight *on the perturbed sequence*, and the true best predictor in hindsight. As we will show, this measures how much the perturbations impact the eventual outcome of prediction. It is useful to note here that under Example 1, too much perturbation would lead to $\mathbb{E}[R^B_T]$ becoming prohibitively large. On the other hand, under Example 2, too little perturbation would lead to $\mathbb{E}[R^A_T]$ becoming prohibitively large.

5.3.2 Part 1: difference between best "perturbed" predictor and true best

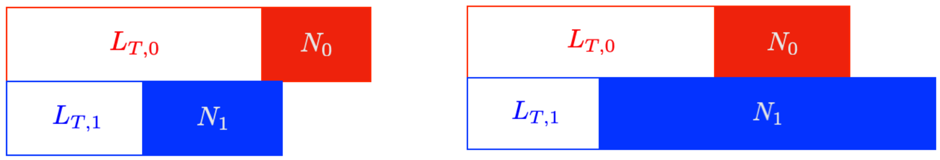
We begin by providing an upper bound on the impact of perturbation on the eventual best-predictor-in-hindsight, i.e. $\mathbb{E}[R^B_T] := \mathbb{E}[L^*_{T,\mathsf{pert.}}] - L^*_T$. Our examples intuitively showed that *less* randomization will lead to a smaller impact; thus, we expect $\mathbb{E}[R^B_T]$ to decrease with η. We will now show that this is the case in a precise quantitive manner, by showing that $\mathbb{E}[R^B_T] = \mathcal{O}\left(\frac{1}{\eta}\right)$.

Without loss of generality, we will assume that the sequence was generated such that $L_{T,1} \leq L_{T,0}$; therefore, 1 is (one of the) best predictor(s) in hindsight. We will use Figure 5.3 to guide our proof strategy. To this end, we make three important observations:

- Suppose that N_0, N_1 are realized such that the identity of the best predictor is *still* 1 under these perturbations. This is the scenario that is depicted in Figure 5.3a. In this case, we will have $L^*_{T,\mathsf{pert.}} = L_{T,1} = L^*_T$, and so the term $R^B_T = 0$.

- Therefore, we can only have $R^B_T > 0$ under the complementary event that 0 is the best predictor in hindsight under the perturbed losses. Because $L_{T,1} < L_{T,0}$, this can only be the case if $N_1 > N_0$ as visualized in Figure 5.3b.

- Finally, under this complementary event we have $R^B_T = L_{T,0} - L_{T,1}$; however, we also have $L_{T,0} + N_0 \leq L_{T,1} + N_1$, which implies that $L_{T,0} - L_{T,1} \leq N_1 - N_0$.



(a) Type 1 realization. (b) Type 2 realization.

Figure 5.3: Visualizing the impact of perturbations on the identity of the best predictor. (a) demonstrates a scenario in which the perturbations are not sufficient to change the identity of the best predictor. (b) demonstrates a scenario in which the perturbations are sufficient to change the identity of the best predictor.

Putting all of these together, we get

$$\begin{aligned}\mathbb{E}[R_T^B] &\leq \mathbb{E}\left[(N_1 - N_0) \cdot \mathbb{I}[N_1 - N_0 \geq 0]\right] \\ &= \mathbb{E}\left[N \cdot \mathbb{I}[N \geq 0]\right]\end{aligned}$$

where we defined $N := N_1 - N_0$. Now, we are going to see a remarkably general way to bound the quantity $N \cdot \mathbb{I}[N \geq 0]$ in terms of $\mathcal{O}\left(\frac{1}{\eta}\right)$ (in fact, this will work for any distribution of variance $\mathcal{O}\left(\frac{1}{\eta^2}\right)$). Because N_1, N_0 are iid, it is clear that N is a symmetric random variable around 0; therefore, we have $\mathbb{E}\left[N \cdot \mathbb{I}[N \geq 0]\right] = \mathbb{E}[|N|]$. Next, we will use Jensen's inequality to get

$$\begin{aligned}(\mathbb{E}[|N|])^2 &\leq \mathbb{E}[N^2] \\ \implies \mathbb{E}[|N|] &\leq (\mathbb{E}[N^2])^{1/2} \\ &= \mathcal{O}\left(\frac{1}{\eta^2}\right).\end{aligned}$$

Putting these steps together completes our proof of the upper bound $\mathbb{E}[R_T^B] = \mathcal{O}\left(\frac{1}{\eta}\right)$. The last inequality above uses the following reasoning:

- Because N_1, N_0 are iid, we have $\mathbb{E}[N] = 0$ and so $\mathbb{E}[N^2] = \text{variance}(N)$.
- Because N_1, N_0 are iid, we have $\text{variance}(N) = \text{variance}(N_1) + \text{variance}(N_0) = \mathcal{O}\left(\frac{1}{\eta^2}\right)$.

5.3.3 Part 2: regret with respect to best "perturbed" predictor

Now, we turn to bounding $\mathbb{E}[R_T^A]$, which is the regret with respect to the best "perturbed leader", i.e. leader on the perturbed sequence. This is equivalent to bounding the regret of FTL on the perturbed loss sequence given by

$$\{N_x, l_{1,x}, \ldots, l_{T,x}\} \text{ for each } x \in \{0, 1\}.$$

Essentially, this tells us that the regret with respect to the best perturbed leader measures the *instability* of the FTPL algorithm, i.e. how many times it changes its mind about the leader in the sequence. Based on the intuition that we explored in Examples 1 and 2, we would expect that more randomization in the algorithm (i.e. lower η) will lead to greater stability, and therefore a smaller regret $\mathbb{E}[R_T^A]$.

Indeed, in this section we will prove that $\mathbb{E}[R_T^A] \leq \eta T$. We will prove this in two steps:

- First, we show that for any realization of the perturbations N_0, N_1, we have $R_T^A \leq \sum_{t=1}^T \mathbb{I}[\widehat{X}_t \neq \widehat{X}_{t-1}$.
- Next, we show that for any round t, we have $\mathbb{P}[\widehat{X}_t \neq \widehat{X}_{t-1}] \leq \eta$. Note that this directly implies the upper bound $\mathbb{E}[R_T^A] \leq \eta T$. (Recall that all expectations and probabilities are with respect to the randomness in the algorithm (i.e. the randomness in the perturbations N_0, N_1).)

We will start by showing the first statement, i.e.

$$R_T^A \leq \sum_{t=1}^{T} \mathbb{P}[\widehat{X}_t \neq \widehat{X}_{t-1}]. \tag{5.6}$$

Proof [of Equation (5.6)] We now introduce one more piece of notation: let $\widetilde{x}^* := \arg\min_{x \in \{0,1\}} L_{T,x}$. We note that, in other words, we have $\widehat{X}_{T+1} = \widetilde{x^*}$. Adding and subtracting terms, we get

$$L^*_{T,\mathsf{pert.}} = L_{T,\widehat{X}_{T+1}} = \sum_{t=1}^{T} (L_{t,\widehat{X}_{t+1}} - L_{t-1,\widehat{X}_t}).$$

Plugging in the definition of R_T^A, we then get

$$\begin{aligned}
R_T^A &= \sum_{t=1}^{T} l_{t,\widehat{X}_t} - \left(\sum_{t=1}^{T} (L_{t,\widehat{X}_{t+1}} - L_{t-1,\widehat{X}_t}) \right) \\
&= \sum_{t=1}^{T} (l_{t,\widehat{X}_t} - (L_{t,\widehat{X}_{t+1}} - L_{t-1,\widehat{X}_t})).
\end{aligned}$$

We will now upper bound each of the terms $(l_{t,\widehat{X}_t} - (L_{t,\widehat{X}_{t+1}} - L_{t-1,\widehat{X}_t}))$. There are two cases:

- The first case is the one for which the leader does not change, i.e. $\widehat{X}_{t+1} = \widehat{X}_t$. Then, we simply have $L_{t,\widehat{X}_{t+1}} - L_{t-1,\widehat{X}_t} = L_{t,\widehat{X}_t} - L_{t-1,\widehat{X}_t} = l_{t,\widehat{X}_t}$ and so the term becomes exactly equal to 0.
- The second and more interesting case is the one for which there is a leader change, i.e. $\widehat{X}_{t+1} \neq \widehat{X}_t$. In that case, we get

$$\begin{aligned}
(l_{t,\widehat{X}_t} - (L_{t,\widehat{X}_{t+1}} - L_{t-1,\widehat{X}_t})) &= L_{t,\widehat{X}_t} - L_{t,\widehat{X}_{t+1}} \\
&= (L_{t-1,\widehat{X}_t} - L_{t-1,\widehat{X}_{t+1}}) + (l_{t-1,\widehat{X}_t} - l_{t-1,\widehat{X}_{t+1}}) \\
&\leq 0 + 1,
\end{aligned}$$

where the last inequality follows because a) $\widehat{X}_t$ is the leader at round t, therefore we need to have $L_{t-1,\widehat{X}_t} \leq L_{t-1,\widehat{X}_{t+1}}$, and b) $l_{t,x} \in \{0,1\}$, therefore, $l_{t,x} - l_{t,x'} \leq 1$ for any x, x'.

Putting these cases together directly gives us Equation (5.6) and completes the proof. ■

Thus, we have shown this somewhat nifty statement: that the regret of FTL is at most the number of leader changes. Now, we will show that the *expected* number of leader changes is bounded above by ηT, i.e. that

$$\sum_{t=1}^{T} \mathbb{P}\left[\widehat{X}_t \neq \widehat{X}_{t-1}\right] \leq \eta T. \tag{5.7}$$

Putting this together with Equation (5.6) clearly gives the desired bound on $\mathbb{E}[R_T^A]$.

Proof [of Equation (5.7)] We will, in fact, show that $\mathbb{P}\left[\widehat{X}_{t+1} \neq \widehat{X}_t\right] \leq \eta$; by linearity of expectation this suffices to show Equation (5.7). Equivalently, we will show that $\mathbb{P}\left[\widehat{X}_{t+1} = \widehat{X}_t\right] \geq 1 - \eta$.

We will show that this reduces to a conditional probability, again, on the random variable $N_1 - N_0$. Let us start with the case $\widehat{X}_t = 0$, which means that

$$\begin{aligned} & L_{t-1,0} + N_0 \leq L_{t-1,1} + N_1 \\ \implies & N_1 - N_0 \geq L_{t-1,0} - L_{t-1,1}. \end{aligned}$$

Then, our goal is to lower-bound $\mathbb{P}[\widehat{X}_{t+1} = 0 | \widehat{X}_t = 0]$. Note that $\widehat{X}_{t+1} = 0$ iff

$$\begin{aligned} & L_{t,0} + N_0 \leq L_{t,1} + N_1 \\ \implies & N_1 - N_0 \geq L_{t,0} - L_{t,1} := L_{t-1,0} + L_{t-1,1} + (l_{t,0} - l_{t,1}). \end{aligned}$$

Denoting $N_1 - N_0 := N$, $L_{t-1,0} - L_{t-1,1} := v$ and $l_{t,0} - l_{t,1} := c$ as shorthand, we then get

$$\begin{aligned} \mathbb{P}[\widehat{X}_{t+1} = 0 | \widehat{X}_t = 0] &= \mathbb{P}\left[N \geq v + c | N \geq v\right] \\ &\geq \mathbb{P}\left[N \geq v + 1 | N \geq v\right] \end{aligned}$$

where the last step follows by the monotonicity of the CDF and noting that $c \leq 1$. Now, we will use the special property of two iid exponential random variables: in particular, N follows a Laplace distribution[5] with parameter $\frac{1}{\eta}$. Therefore, we get

$$\begin{aligned} \mathbb{P}\left[N \geq v + 1 | N \geq v\right] &= \frac{\mathbb{P}\left[N \geq v + 1\right]}{\mathbb{P}\left[N \geq v\right]} \\ &= \frac{0.5e^{-\eta(v+1)}}{0.5e^{-\eta v}} \\ &= e^{-\eta} \geq 1 - \eta. \end{aligned}$$

5. For the precise definition of the Laplace distribution that we use in this note, see: `https://en.wikipedia.org/wiki/Laplace_distribution`.

An identical argument to the above gives us $\mathbb{P}[\widehat{X}_{t+1} = 1|\widehat{X}_t = 1] \geq 1 - \eta$. Putting these together then gives us

$$\mathbb{P}[\widehat{X}_{t+1} = \widehat{X}_t] = \mathbb{P}[\widehat{X}_t = 1] \cdot \mathbb{P}[\widehat{X}_{t+1} = 1|\widehat{X}_t = 1] + \mathbb{P}[\widehat{X}_t = 0] \cdot \mathbb{P}[\widehat{X}_{t+1} = 0|\widehat{X}_t = 0] \geq 1 - \eta,$$

and this completes the proof of Equation (5.7). ■

5.3.4 Putting it all together

Thus, in summary we have

$$R_T = \mathcal{O}\left(\frac{1}{\eta} + \eta T\right),$$

and substituting $\eta = 1/\sqrt{T}$ gives us $R_T = \mathcal{O}(\sqrt{T})$, just as in the case of MWA!

It is interesting to note that the choice of distribution on the perturbation is quite flexible. In the first part of the proof (upper bounding $\mathbb{E}[R_T^B]$), we only used the fact that the perturbations are of variance $\frac{1}{\eta^2}$!

In fact, only in the proof of Equation (5.7) did we actually use the fact that the distribution of the perturbations is exponential. Even this proof could be adapted to other choices (such as the Gumbel distribution, which you explored in HW 1, Problem 4). It may be a useful exercise to try and prove a version of Equation (5.7) for these other cases as well.

This shows that the FTPL algorithm is flexible in nature (because several choices of perturation distribution can be leveraged), and satisfies an intuitive principle of trading off randomization in the algorithm and exploiting past information in the sequence.