EE 194/290S: ML for Sequential Decisions under Uncertainty Fall 2018

Lecture 7: Follow-the-Perturbed-Leader Analysis

Instructors: Anant Sahai, Vidya Muthukumar *Scribes: Sayan Paul, Hitesh Yalamanchili*

1 Follow the Perturbed Leader (FTPL) Recap

As a recap, FTPL is a modification of the traditional Follow the Leader (FTL) algorithm where, at each time step, we perturb the cumulative loss. In other words, we determine our prediction at time t using the following objective:

$$\hat{x}_{t,\texttt{FTPL}} = \arg\min_{x\in\{0,1\}} \underbrace{L_{t-1,x}}_{\substack{\text{original loss}\\ \text{objective}}} + \underbrace{N_{t,x}}_{\substack{\text{random}\\ \text{perturbation}}} \tag{1}$$

We also noted that if the perturbation is distributed according to a Gumbel distribution with parameters 0 and $\frac{1}{\eta}$, the probability that the t-th prediction is x is the same as if we had used the multiplicative weights update.

2 FTPL Regret

Our definition of **regret** is the difference between the loss we incurred using our current strategy and the best loss achieved for choosing a single value and sticking with it, i.e.

$$R_T = \left(\sum_{t=1}^{T} l_{t,\hat{x}_t}\right) - L_{t,x^*}$$

Claim. For all possible sequences of losses where the loss is bounded and the noise is *spread out* (i.e. the noise has a sufficiently high variance), FTPL achieves an expected regret that is bounded by

$$\mathbb{E}[R_T] \in \mathcal{O}(\sqrt{T})$$

In the context of a binary sequence, where we incur a Hamming loss at each time step, and with perturbations that are distributed according to a Gumbel distribution, this bound manifests itself as

$$\mathbb{E}[R_T] \lesssim \sqrt{T \ln 2}$$

Consider the sequences of losses

$$\{l_{1,0}, l_{2,0}, \ldots, l_{T,0}\}$$
$$\{l_{1,1}, l_{2,1}, \ldots, l_{T,1}\}$$

where $l_{t,x}$ is the loss incurred at time t for value x as defined by the Hamming loss function.

In expectation, running FTPL on this sequence is the same as running FTL on

$$\{N_0, l_{1,0}, l_{2,0}, \ldots, l_{T,0}\}$$
$$\{N_1, l_{1,1}, l_{2,1}, \ldots, l_{T,1}\}$$

where N_x is some value sampled from a perturbation distribution. Throughout our analysis of this alternative FTPL view, we are assuming that the adversary generating the sequence is offline in that it decides its sequence at the very beginning and doesn't adapt to our predictions. Notice that even with the perturbation we add at the beginning, an adaptive adversary will still be able to trick the algorithm. Thus, we are working with a weakened adversary where this is not the case.

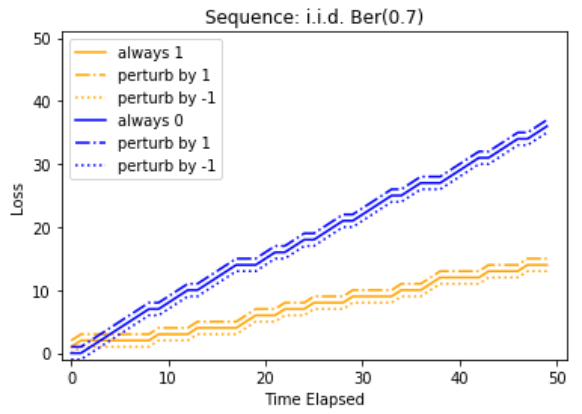
Note. Running FTL on this sequence is different from FTPL because we only sample from the perturbation distribution once at the beginning. In other words, we are shifting the cumulative loss, $L_{t,x}$, by a constant.

$$L_{t,x} = \left(\sum_{i=1}^{t} l_i \right) + N_x$$

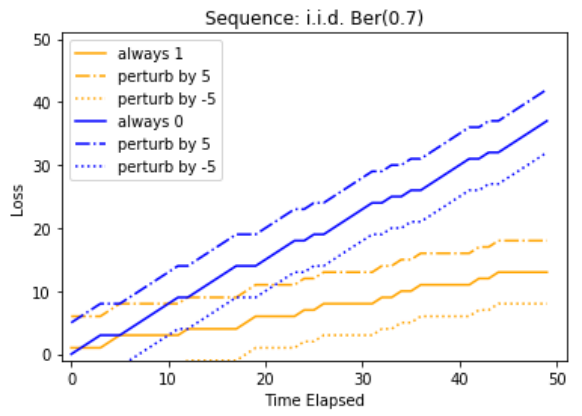
However, this is the same as FTPL in expectation. We can see this by observing that in the case of FTPL, we would normally have a randomly sampled perturbation coming in with each incremental loss, as we can see in equation 1. When all of these perturbation terms are summed up, due to linearity of expectation, the expectation of the sum of these t randomly sampled perturbations is equivalent to the expected value of some N_x sampled from a distribution. Therefore, we can use this simplified setting to bound the expected regret.

The crucial idea is that our simplified view of FTPL is *stubborn*. Since we are adding a perturbation at the beginning for each action (and assuming that the difference of these perturbations is well spread out), we are separating the losses from the start. At every time step, we can only ever move one of the losses at a time, and even than at most by 1. So, our algorithm is "stubborn" in the sense that given the clear choice at the beginning, it may take some time for the other choice to catch up and have its loss dip below. Only at that point would the prediction change. Now, we can further understand the effects of this FTPL through some specific examples with some arbitrarily chosen perturbation to illustrate the benefits and disadvantages of separating the losses at the start.

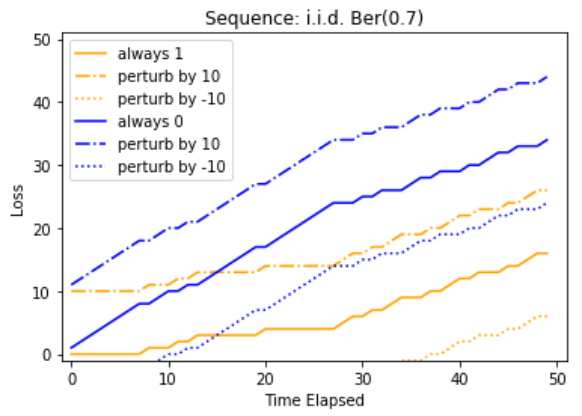
Example. Let's consider a sequence of i.i.d. Bernoulli random variables with parameter 0.7, which is something that vanilla FTL would do fairly well on, and analyze the effect our choice of perturbation has on what we think is the best action under the perturbed losses.



If we have a very small perturbation, then the effect on the cumulative loss is naturally also small. We can see in the plot above that 1 was the best action in hindsight in the original situation (since it had a lower cumulative loss), and that remained true even after perturbing the losses. In other words, $\tilde{x}^* = x^*$.



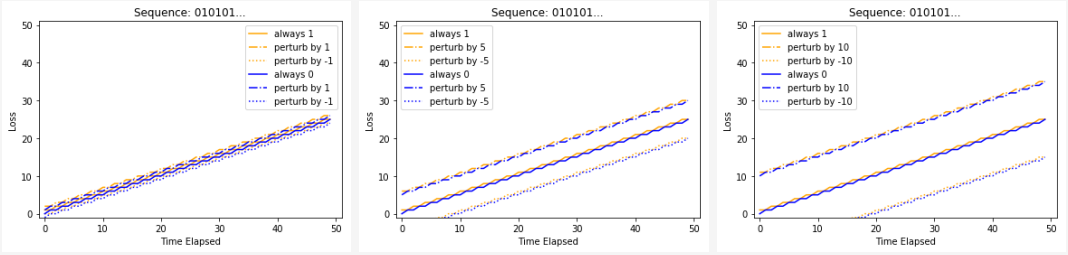
Now, since we've chosen a higher perturbation, some of the loss plots are growing quite close. Specifically, if we shift the loss from choosing 1 by 5, and that of choosing 0 by -5, the cumulative losses are much closer than they would be before. In this scenario, $\tilde{x}^*$ is still equal to x^*, but the lead that 1 had over 0 under perturbations shrank in some scenarios.



Our noise is pretty huge in this scenario, so much so that 1 now actually looks *worse* than 0 if we perturbed its loss by 10 and 0's loss by -10. In this case, $\tilde{x}^* \neq x^*$.

In essence, in this case where FTL performs fairly well, we want to make sure that the initial perturbation of the losses isn't too much, otherwise FTPL will do poorly.

Now, let's consider the effect of noise on a sequence of alternative 0's and 1's, which FTL does very poorly on.



Here, the crucial thing is that as long as the initial perturbation that we apply to the losses makes them sufficiently separated, then FTPL will do well. Unlike before, a larger noise is in fact beneficial.

Here lies the core tension—we want to pick a perturbation distribution that is both spread out enough

to differentiate two losses for sequences where FTL does poorly, but remains small enough that we don't do too much worse than FTL on sequences for which it does well.

We proceed by splitting the regret expression into two terms

$$R_T = \underbrace{\sum_{t=1}^{T} l_{t,\hat{x}_t} - L_{T,\tilde{x}^*}}_{R_T^{(A)} \text{ (est. error)}} + \underbrace{L_{T,\tilde{x}^*} - L_{T,x^*}}_{R_T^{(B)} \text{ (approx. error)}} , \tag{2}$$

where $\tilde{x}^*$ is the best action in hindsight under the *perturbed losses*, i.e.,

$$\tilde{x}^* = \arg\min_x L_{T,x} + N_x$$

and x^* is the best action under the *original losses*, i.e.

$$x^* = \arg\min_x L_{T,x}$$

Notice that these two values are different (i.e. $\tilde{x}^* \neq x^*$) if the amount by which we perturbed the cumulative loss is large enough to overcome the lead that the L_{T,x^*} had over $L_{T,\tilde{x}^*}$. To bound our expression for $\mathbb{E}[R_T]$, we will bound $\mathbb{E}[R_T^{(A)}]$ and $\mathbb{E}[R_T^{(B)}]$ separately.

2.1 Analysis of Approximation Error, $R_T^{(B)}$

The approximation error we have is the error we incur by aiming at the wrong target, that is, by choosing $\tilde{x}^*$ as our reference class rather than x^*. This is just the difference between $L_{T,\tilde{x}^*} - L_{T,x^*}$.

Lemma 1. We can bound the expected value of $R_T^{(B)}$ using the Gumbel distribution for our perturbations as:

$$\mathbb{E}[R_T^{(B)}] \leq \frac{\ln(2)}{\eta}$$

Proof. We can split up our analysis into two cases since

$$\mathbb{E}[R_T^{(B)}] = \mathbb{E}[R_T^{(B)} \mid \tilde{x}^* = x^*]\mathbb{P}[\tilde{x}^* = x^*] + \mathbb{E}[R_T^{(B)} \mid \tilde{x}^* \neq x^*]\mathbb{P}[\tilde{x}^* \neq x^*]$$

Case 1 ($\tilde{x}^* = x^*$): Since these two are equal, the cumulative losses are necessarily equal, so the $R_T^{(B)}$ given that $\tilde{x}^* = x^*$ is 0. Therefore, this term disappears from the expectation.

Case 2 ($\tilde{x}^* \neq x^*$): Since our optimal choice after perturbation is different from that before, that means that the perturbation we applied for that value (recalling that $\tilde{x} \in \{0, 1\}$) made its loss *dip below* that of the opposite value. We can write this as an explicit inequality

$$L_{T,\tilde{x}^*} + N_{\tilde{x}^*} \leq L_{T,\,\mathtt{not}(\tilde{x}^*)} + N_{\mathtt{not}(\tilde{x}^*)}$$

By the assumption we made earlier (that $\tilde{x}^* \neq x^* \Rightarrow \mathtt{not}(\tilde{x}^*) \equiv x^*$), this inequality becomes

$$L_{T,\tilde{x}^*} + N_{\tilde{x}^*} \leq L_{T,x^*} + N_{x^*}$$

We can rearrange terms to see that

$$L_{T,\tilde{x}^*} - L_{T,x^*} \leq \underbrace{N_{x^*} - N_{\tilde{x}^*}}_{\text{diff. in perturbations}}$$

We define N to be the difference in perturbations, i.e. $N \coloneqq N_{x^*} - N_{\tilde{x}^*}$.

Note. In the context of case 2, N must be at least 0. We prove this by contradiction. We are in case 2, so $\tilde{x}^* \neq x^*$. If N were less than 0, then that means that the loss for the optimal solution without perturbations, L_{T,x^*}, was shifted up less than the loss for the perturbed solution was shifted down. But in that case, x^* would look like a better solution, even after perturbations,

than $\texttt{not}(x^*)$, which means that $\tilde{x}^* = x^*$. Contradiction.

The left hand side of the equation is exactly the expression for $R_T^{(B)}$. If we take the expectation of both sides of the inequality, using the fact that N must be greater than 0 for $\tilde{x}^* \neq x^*$, then

$$\mathbb{E}[R_T^{(B)} \mid \tilde{x}^* \neq x^*] \leq \mathbb{E}[N \mid N \geq 0]$$

We have already proved that if $N_x \sim \text{Gumbel}(0, \frac{1}{\eta})$, FTPL has a similar expected regret as multiplicative weights. We'll use this as the distribution for our perturbations and recall that the difference of two Gumbel distributions (i.e. N_{x^*} and $N_{\tilde{x}^*}$) is a Logistic distribution (i.e. $N \sim \text{Logistic}(0, \frac{1}{\eta})$). Now, we can find an expression for the right side of the inequality (expressions for the PDF for a logistic distribution are from Wolfram MathWorld[1]).

$$\begin{aligned}
\mathbb{E}[N \mid N \geq 0] &= \frac{1}{\mathbb{P}(N \geq 0)} \int_0^\infty x f_N(x)\, \mathrm{d}x = \frac{1}{1/2} \int_0^\infty x \frac{\eta \mathrm{e}^{-x\eta}}{(1+\mathrm{e}^{-x\eta})^2}\, \mathrm{d}x \\
&= 2 \int_{\frac{1}{2}}^1 -\frac{1}{\eta} \ln \frac{1-u}{u}\, \mathrm{d}u \quad \left(\text{where } u = \frac{1}{1+\mathrm{e}^{-x\eta}}\right) \\
&= -\frac{2}{\eta} \int_{\frac{1}{2}}^1 \ln(1-u) - \ln(u)\, \mathrm{d}u \\
&= -\frac{2}{\eta} ((u-1)\ln(1-u) - u - u\ln(u) + u) \Big|_{1/2}^1 \\
&= -\frac{2}{\eta} \ln\left(\frac{(1-u)^{u-1}}{u^u}\right) \Big|_{1/2}^1 \\
&= -\frac{2}{\eta} \left(\lim_{u\to 1} \ln\left(\frac{(1-u)^{u-1}}{u^u}\right) - \ln\left(\frac{(1/2)^{-1/2}}{(1/2)^{1/2}}\right)\right) \\
&= -\frac{2}{\eta} \left(\ln(1) - \ln\left((1/2)^{-1}\right)\right) \\
&= \frac{2\ln(2)}{\eta}
\end{aligned}$$

In the second step we used the fact that N is symmetric about 0 by construction, since it is the difference between two i.i.d. variables. Therefore, probability that it is larger than 0 is $\frac{1}{2}$. Now, we can actually bound the expected value of $R_T^{(B)}$

$$\begin{aligned}
\mathbb{E}[R_T^{(B)}] &= \mathbb{E}[R_T^{(B)} \mid \tilde{x}^* = x^*]\mathbb{P}[\tilde{x}^* = x^*] + \mathbb{E}[R_T^{(B)} \mid \tilde{x}^* \neq x^*]\mathbb{P}[\tilde{x}^* \neq x^*] \\
&= 0 + \mathbb{E}[R_T^{(B)} \mid \tilde{x}^* \neq x^*]\mathbb{P}[\tilde{x}^* \neq x^*] = \mathbb{E}[R_T^{(B)} \mid \tilde{x}^* \neq x^*]\mathbb{P}[\tilde{x}^* \neq x^*]
\end{aligned}$$

In the above note, we proved that the $\tilde{x}^* \neq x^*$ implies that $N > 0$. In other words, the difference in perturbations between the true x^* and $\texttt{not}(x^*)$ must be greater than 0 for there to be any possibility of having a different optimal leader at time T. However, there could be cases where the difference is positive but no leader change happened. This happens in the plot for the sequence of i.i.d. Ber(0.7) variables. If we isolate the 0 with negative perturbation and 1 with positive perturbation. The difference in perturbations between the true optimal, 1, and 0 are clearly positive, but there was no leader change in the end. Therefore, we can conclude that

$$\mathbb{P}[\tilde{x}^* \neq x^*] \leq \mathbb{P}[N > 0] = \frac{1}{2}$$

Therefore,

$$\mathbb{E}[R_T^{(B)}] = \mathbb{E}[R_T^{(B)} \mid \tilde{x}^* \neq x^*]\mathbb{P}[\tilde{x}^* \neq x^*] \leq \frac{2\ln 2}{\eta} \cdot \frac{1}{2}$$

$$\mathbb{E}[R_T^{(B)}] \leq \frac{\ln 2}{\eta}$$

□

[1] http://mathworld.wolfram.com/LogisticDistribution.html

2.2 Analysis of Estimation Error, $R_T^{(A)}$

The estimation error component of the regret reflects the error in our ability to find the best possible sequence of $\hat{x}_t$ with respect to the best loss we can achieve through our reference model, which in this case is a constant choice of x. Specifically, we defined $R_T^{(A)}$ as

$$R_T^{(A)} = \sum_{t=1}^{T} l_{t,\hat{x}_t} - L_{T,\tilde{x}^*},$$

which has two components. The sum is the loss that we accumulate from using FTL at each time step on the perturbed losses. The second term is the total loss if we had taken the best action in hindsight (on the perturbed loss) and stuck with it. We would like to bound $R_T^{(A)}$, but before we make any claims, let's refer back to our understanding of the simplified view of FTPL as being "stubborn."

As we explicitly state below, we will bound the estimation error by measuring the "stubbornness" of our algorithm, i.e., the number of times it changes its mind about the best action in hindsight. This is a natural measure for the estimation error since our reference class is the single best-in-hindsight action at time T. Meanwhile, in our actual simplified view of the FTPL algorithm, we start off with an arbitrary perturbation or "stubbornness" in a certain direction, which if large enough would force us to predict the same value at each time for a potentially longer period, essentially mirroring the reference class model. Thus, to understand the difference between our predictor and the reference class, which is what estimation error represents, we must understand how many times our algorithm changes its mind i.e. returns a different prediction from the previous iteration.

Claim. $R_T^{(A)}$ is *at most* the total number of leader changes while executing the FTL algorithm on the perturbed losses, where the losses are bounded by 1 due to our use of the Hamming function. At any time step t, a **leader change** happens when $\hat{x}_{t+1} \neq \hat{x}_t$, so we can bound as follows:

$$R_T^{(A)} \leq \sum_{t=1}^{T} \mathbb{1}[\hat{x}_{t+1} \neq \hat{x}_t]$$

Proof. To begin, we can re-express the cumulative loss term of the $R_T^{(A)}$ equation as a telescoping sum of the difference between consecutive cumulative losses at each time step. Furthermore, we can let $\tilde{x}^* = \hat{x}_{T+1}$, since $\hat{x}_{t+1}$ is just the best x we could have chosen over the total perturbed loss $L_{T,x} + N_{T+1}$. This exactly matches the definition of $\tilde{x}^*$, i.e. the optimal x in hindsight under the perturbed cumulative loss. Thus, we have:

$$L_{T,\tilde{x}^*} = \sum_{t=2}^{T+1} (L_{t-1,\hat{x}_t} - L_{t-2,\hat{x}_{t-1}})$$

Thus, rewriting $R_T^{(A)}$, we have:

$$\begin{aligned} R_T^{(A)} &= \sum_{t=1}^{T} l_{t,\hat{x}_t} - \left(\sum_{t=2}^{T+1} (L_{t-1,\hat{x}_t} - L_{t-2,\hat{x}_{t-1}}) \right) \\ &= \sum_{t=1}^{T} l_{t,\hat{x}_t} - \left(\sum_{t=1}^{T} (L_{t,\hat{x}_{t+1}} - L_{t-1,\hat{x}_t}) \right) \\ &= \sum_{t=1}^{T} (l_{t,\hat{x}_t} - (L_{t,\hat{x}_{t+1}} - L_{t-1,\hat{x}_t})) \end{aligned}$$

From here, we can split up our analysis again into two cases:

Case 1: Suppose that $\hat{x}_{t+1} = \hat{x}_t$. In this case, the chosen $\hat{x}$ didn't change. Thus, we can simplify the difference of the cumulative losses term to:

$$L_{t,\hat{x}_{t+1}} - L_{t-1,\hat{x}_t} = L_{t,\hat{x}_t} - L_{t-1,\hat{x}_t} = l_{t,\hat{x}_t}$$

Thus in this case, the term in the summation of $R_T^{(A)}$ above will be $l_{t,\hat{x}_t} - l_{t,\hat{x}_t} = 0$.

Case 2: Suppose that $\hat{x}_{t+1} \neq \hat{x}_t$. In this case, we cannot make any exact conclusions, but we can still use what we know about the losses to bound the term of the summation. Specifically, we know that $l_{t,\hat{x}_t} \leq 1$, since the loss can only either be 1 or 0 at any specific time step. Now, we can rewrite this term by simply taking an incremental loss term out of the first cumulative loss term.

$$L_{t,\hat{x}_{t+1}} - L_{t-1,\hat{x}_t} = l_{t,\hat{x}_{t+1}} + L_{t-1,\hat{x}_{t+1}} - L_{t-1,\hat{x}_t}$$

Now, we note that based on the definition of the Hamming loss, we have that $l_{t,\hat{x}_{t+1}} \geq 0$. Furthermore, since we know that these cumulative losses are the ones over which the argmin was computed to determine $\hat{x}_t$, we know that the cumulative loss from using $\hat{x}_t$ must be lower at time step $t-1$, so we have that

$$L_{t-1,\hat{x}_{t+1}} - L_{t-1,\hat{x}_t} \geq 0$$

Consequently, putting these both together, $L_{t,\hat{x}_{t+1}} - L_{t-1,\hat{x}_t}$ must also be greater than or equal to 0. Using the upper bound on the incremental loss term and the lower bound on the difference of the cumulative losses, we thus have:

$$l_{t,\hat{x}_t} - (L_{t,\hat{x}_{t+1}} - L_{t-1,\hat{x}_t}) \leq 1$$

Thus, using the equalities/bounds from both of these cases, we can then bound $R_T^{(A)}$ as follows:

$$R_T^{(A)} \leq \sum_{t=1}^{T} \mathbb{1}[\hat{x}_{t+1} \neq \hat{x}_t]$$

□

Now that we have bounded $R_T^{(A)}$ itself, we can determine the probability of a leader change happening ($\hat{x}_{t+1} \neq \hat{x}_t$) to bound the expectation.

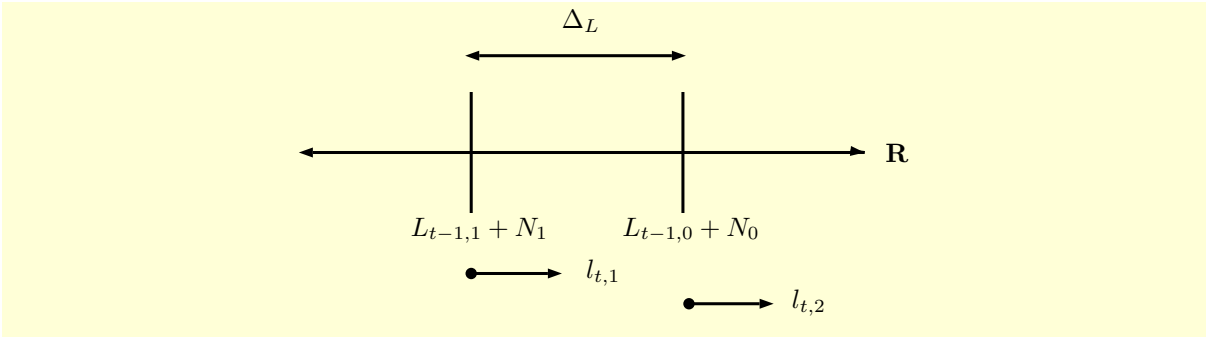
Lemma 2. We can bound the expected value of $R_T^{(A)}$ as follows:

$$\mathbb{E}[R_T^{(A)}] \leq \eta T$$

Proof. Starting from the bound we deduced earlier, in order to be able to bound the expected value of $R_T^{(A)}$, we need to be able to bound the probability of a leader change happening, which is $\mathbb{P}[\hat{x}_{t+1} \neq \hat{x}_t]$.

Note. Let's think about what it means for a leader change to occur. Recall that in this setting, we added a noise term at the beginning of the sequence. The effect of this was that, instead of starting the cumulative losses at the same place at 0, we have separated them by the difference in the noise values for both actions. Concretely, the difference between our cumulative losses at the beginning is $N_0 - N_1$.

If the initial difference is sufficiently large, then FTL has a clear winner at the beginning. Furthermore, the losses at each time step can only increase the cumulative loss of one of the actions by 1 (which we show in the figure below). So, the clear winner will remain the same until its cumulative loss becomes neck-and-neck with the other action. Only at this point can leader changes occur. Therefore, finding a bound on the probability of a leader change boils down to determining whether the initial separation between the cumulative losses was sufficiently large.



To come up with an upper bound for this probability, we can also equivalently find a lower bound for the complementary probability as follows:

$$\begin{aligned}1 - \mathbb{P}[\hat{x}_{t+1} \neq \hat{x}_t] &= \mathbb{P}[\hat{x}_{t+1} = \hat{x}_t] \\ &= \mathbb{P}[\hat{x}_{t+1} = 1 \mid \hat{x}_t = 1]\mathbb{P}[\hat{x}_t = 1] + \mathbb{P}[\hat{x}_{t+1} = 0 \mid \hat{x}_t = 0]\mathbb{P}[\hat{x}_t = 0]\end{aligned}$$

Here, we recall how we select the $\hat{x}_t$ and $\hat{x}_{t+1}$ actions themselves in the first place to help us find these probabilities. Specifically, we let $\hat{x}_t = 1$ when:

$$L_{t-1,1} + N_1 \leq L_{t-1,0} + N_0$$

Rewriting the above, we have:

$$N_0 - N_1 \geq L_{t-1,1} - L_{t-1,0} =: v_1$$

Similarly, for $\hat{x}_{t+1} = 1$, we have:

$$L_{t,1} + N_1 \leq L_{t,0} + N_0$$

Again, rewriting the above, we have:

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

Here, we can define $v_1 := L_{t-1,1} - L_{t-1,0}$ and $c_1 := l_{t,1} - l_{t,0}$, so we have:

$$N_0 - N_1 \geq v_1 + c_1$$

Now, using the definitions of v_1 and c_1, we thus have:

$$\mathbb{P}[\hat{x}_{t+1} = 1 \mid \hat{x}_t = 1] = \mathbb{P}[N_0 - N_1 \geq v_1 + c_1 \mid N_0 - N_1 \geq v_1]$$

Performing a similar process for the cases where $\hat{x}_{t+1} = 0$ and $\hat{x}_t = 0$, we will get the following, where $v_2 = L_{t-1,0} - L_{t-1,0}$ and $c_2 = l_{t,0} - l_{t,1}$:

$$\mathbb{P}[\hat{x}_{t+1} = 0 \mid \hat{x}_t = 0] = \mathbb{P}[N_1 - N_0 \geq v_2 + c_2 \mid N_1 - N_0 \geq v_2]$$

Thus, putting these together and recalling that the difference of two Gumbel distributions (i.e. N_0 and N_1) is a Logistic distribution (i.e. $N \sim \text{Logistic}(0, \frac{1}{\eta})$), we have that:

$$\mathbb{P}[\hat{x}_{t+1} = \hat{x}_t] = \mathbb{P}[N \geq v_1 + c_1 \mid N \geq v_1]\mathbb{P}[\hat{x}_t = 1] + \mathbb{P}[N \geq v_2 + c_2 \mid N \geq v_2]\mathbb{P}[\hat{x}_t = 0]$$

Now, we can find a lower bound for these probabilities.

$$\begin{aligned}\mathbb{P}[N \geq v_1 + c_1 \mid N \geq v_1] &\geq \frac{\mathbb{P}[N \geq v_1 + |c_1|]}{\mathbb{P}[N \geq v_1]} \\ \mathbb{P}[N \geq v_2 + c_2 \mid N \geq v_2] &\geq \frac{\mathbb{P}[N \geq v_2 + |c_2|]}{\mathbb{P}[N \geq v_2]}\end{aligned}$$

Note. In the above bounds, we have equality when $c_1 \geq 0$ and $c_2 \geq 0$ by the definition of conditional probability. If $c_1 < 0$ or $c_2 < 0$, then the left side essentially asks for the probability that N is greater than some value given that it is greater than some larger value, so the respective probabilities are just 1. Since we consider the absolute values of c_1 and c_2 on the right side, that means the right side is less than or equal to 1. Therefore, the inequality covers both cases.

We can take advantage of the definition of the CDF of the Logistic distribution to find an explicit lower bound.

$$\begin{aligned}
\mathbb{P}[N \geq v_1 + c_1 \mid N \geq v_1] &\geq \frac{\mathbb{P}[N \geq v_1 + |c_1|]}{\mathbb{P}[N \geq v_1]} \\
&\geq \frac{e^{-\eta(v_1+|c_1|)}}{1+e^{-\eta(v_1+|c_1|)}} \frac{1+e^{-\eta v_1}}{e^{-\eta v_1}} \\
&\geq e^{-\eta|c_1|} \\
&\geq 1 - \eta|c_1|
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}[N \geq v_2 + c_2 \mid N \geq v_2] &\geq \frac{\mathbb{P}[N \geq v_2 + |c_2|]}{\mathbb{P}[N \geq v_2]} \\
&\geq \frac{e^{-\eta(v_1+|c_1|)}}{1+e^{-\eta(v_2+|c_2|)}} \frac{1+e^{-\eta v_2}}{e^{-\eta v_2}} \\
&\geq e^{-\eta|c_2|} \\
&\geq 1 - \eta|c_2|
\end{aligned}$$

Thus, we have that:

$$\begin{aligned}
\mathbb{P}[\hat{x}_{t+1} = \hat{x}_t] &\geq \frac{1}{2}(1-\eta|c_1|) + \frac{1}{2}(1-\eta|c_2|) \\
&= 1 - \frac{\eta}{2}(|c_1| + |c_2|)
\end{aligned}$$

Thus, we have that the probability of a leader change is:

$$\mathbb{P}[\hat{x}_{t+1} = \hat{x}_t] \leq \frac{\eta}{2}(|c_1| + |c_2|)$$

Both c_1 and c_2 can take any value in $\{-1, 0, 1\}$ since they represent the difference in loss between choosing 0 or 1 in a given round. Therefore, the absolute values of both are upper bounded by 1, which means the individual probabilities are upper-bounded by $\frac{\eta}{2} \cdot 2 = \eta$. Now, for the entire summation, we get:

$$\mathbb{E}[R_T^{(A)}] \leq \sum_{t=2}^{T+1} \mathbb{1}[\hat{x}_{t+1} \neq \hat{x}_t] \leq \eta T$$

□

2.3 Putting It All Together

Recalling how we defined R_T as the sum of $R_T^{(A)}$ and $R_T^{(B)}$ in Equation 2, we can now easily bound the total regret using the linearity of expectation and the bounds from Lemma 1 and Lemma 2.

$$\begin{aligned}
\mathbb{E}[R_T] &= \mathbb{E}[R_T^{(A)} + R_T^{(B)}] \\
&= \mathbb{E}[R_T^{(A)}] + \mathbb{E}[R_T^{(B)}] \\
&\leq \eta T + \frac{\ln 2}{\eta}
\end{aligned}$$

We want to choose the tightest bound possible for $\mathbb{E}[R_T]$, and since we have control over η, we can optimize the right side over η.

$$\frac{\mathrm{d}}{\mathrm{d}\eta}\eta T + \frac{\ln 2}{\eta} = T - \frac{\ln 2}{\eta^2} = 0$$

$$\Rightarrow \eta^2 = \frac{\ln 2}{T}$$
$$\Rightarrow \eta = \sqrt{\frac{\ln 2}{T}}$$

Plugging that into our expression for the bound, we see that

$$\mathbb{E}[R_T] \leq T\sqrt{\frac{\ln 2}{T}} + \sqrt{\frac{T}{\ln 2}}\ln 2$$
$$\Rightarrow \mathbb{E}[R_T] \leq 2\sqrt{T \ln 2}$$
$$\Rightarrow \mathbb{E}[R_T] \lesssim \sqrt{T \ln 2}$$

This is a very important result since it shows that the expected value of the regret grows sublinearly (i.e. $o(T)$), so we can be sure that the regret when using the FTPL algorithm is still bounded. This also tells us that the perturbations we choose can give us real benefits by giving FTPL a clearer choice of which action to choose. Moreover, we also see that in the case that FTL does very well, and perturbations might actually hurt us, we still also do reasonably well in those circumstances.