Sequential Learning with full information (Part 1)

Pierre Gaillard

1 Useful information

– **Pierre Gaillard** (INRIA Grenoble), **Rémy Degenne** (INRIA Lille)
– Email: pierre.gaillard@inria.fr, remy.degenne@inria.fr Webpage for the class:

`http://pierre.gaillard.me/teaching/mva.html`

– Final grade: approximately 70% final exam, 30% homeworks. A single two-sided sheet of handwritten notes (with any content) will be allowed for the exam.
– Relevant references: Cesa-Bianchi and Lugosi [2006], Shalev-Shwartz et al. [2012], Hazan et al. [2016], Lattimore and Szepesvári [2020]
– Content of the class: mostly theoretical (algorithms and proofs), sequential learning with adversarial data, stochastic bandits, adversarial bandits

2 Introduction

In many applications, the data set is not available from the beginning to learn a model but it is observed sequentially as a flow of data. Furthermore, the environment may be so complex that it is unfeasible to choose a comprehensive model and use classical statistical theory and optimization. A classic example is the spam detection which can be seen as a game between spammer and spam filters. Each trying to fool the other one. Another example, is the prediction of processes that depend on human behaviors such as the electricity consumption. These problems are often not adversarial games but cannot be modeled easily and are surely not i.i.d.
There is a necessity to take a robust approach by using a method that learns as ones goes along, learning from experiences as more aspects of the data and the problem are observed. This is the goal of online learning. The curious reader can know more about online learning in the books Cesa-Bianchi and Lugosi [2006], Hazan et al. [2016], Shalev-Shwartz et al. [2012].

Setting In online learning, a player sequentially makes decisions based on past observations. After committing the decision, the player suffers a loss (or receives a reward depending on the problem). Every possible decision incurs a (possibly different) loss. The losses are unknown to the player beforehand an may be arbitrarily chosen by some adversary. More formally, an online learning problem can be formalized as in Figure 1.

Example 2.1 (Multi-armed bandit)**.** In K-armed bandit, the decision set are K actions (or arms) $\Theta = \{1, \dots, K\}$ and the player only observes the performance of the chosen action (bandit feedback). In this problem, there is an exploration-exploitation trade-off: the player wants to select the best arm as often as possible but he also needs to explore all arms to estimate their performance.

This problem takes his name from slot machines (also known as one-armed bandits because they were originally operated by one lever on the side of the machine) in which some player explores several slot machines and tries to maximize his cumulative gain (or more likely minimize his loss!).

At each time step $t = 1, \dots, T$
- the player observes a context $x_t \in \mathcal{X}$ (optional step)
- the player chooses an action $\theta_t \in \Theta$ (compact decision/parameter set);
- the environment chooses a loss function $\ell_t : \Theta \to [-1, 1]$;
- the player suffers loss $\ell_t(\theta_t)$ and observes
  - the losses of every actions: $\ell_t(\theta)$ for all $\theta \in \Theta$ $\quad\to\quad$ full-information feedback
  - the loss of the chosen action only: $\ell_t(\theta_t)$ $\quad\to\quad$ bandit feedback.

The goal of the player is to minimize his cumulative loss:

$$\widehat{L}_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(\theta_t)\,.$$

Figure 1: Setting of an online learning problem/online convex optimization

Originally, multi-armed bandit setting was introduced by Thomson in 1933 and motivated by clinical trials. For the t-th patient in some clinical study, one needs to choose the treatment to assign to this patient and observe the response. The goal is to maximize the number of patients healed during the study.

Nowadays, multi-armed bandit is motivated by many applications coming from internet (recommender systems, online advertisements,...). We will see more on multi-armed bandit in next lectures.

Example 2.2 (Prediction with expert advice)**.** In prediction with expert advice, there is some sequence of observations $y_1, \dots, y_T \in [0, 1]$ to be predicted step by step with the help of expert forecasts. The setting can be formalized as follows: at each time step $t \geq 1$
- the environment reveals experts forecasts $x_t(k)$ for $k = 1, \dots, K$
- the player chooses a weight vector $p_t \in \Delta_K \stackrel{\text{def}}{=} \{p \in [0, 1]^K : \sum_{k=1}^K p_k = 1\}$ (here θ_t is denoted p_t and $\Theta = \Delta_K$)
- the player forecasts $\widehat{y}_t = \sum_{k=1}^K p_t(k) x_t(k)$
- the environment reveals $y_t \in [0, 1]$ and the player suffers loss $\ell_t(p_t) = \ell(\widehat{y}_t, y_t)$ where $\ell : [0, 1]^2 \to [0, 1]$ is a loss function.

Considering $\Theta := \Delta_K$ and $\theta_t := p_t$, this setting can be recovered by the online learning setting of Figure 1. The inputs correspond to the expert advice $x_t(k)$ that are often revealed before the learner makes his decision p_t.

Player's performance is then measured via a loss function $\ell_t(p_t) = \ell(\widehat{y}_t, y_t)$ which measures the distance between the prediction $\widehat{y}_t$ and the output y_t. Typical loss functions are the squared loss $\ell(\widehat{y}_t, y_t) = (\widehat{y}_t - y_t)^2$, the absolute loss $\ell(\widehat{y}_t, y_t) = |\widehat{y}_t - y_t|$ or the absolute percentage of error $\ell(\widehat{y}_t, y_t) = |\widehat{y}_t - y_t|/|y_t|$. All these loss functions are convex, which will play an important role in the analysis.

How to measure the performance: the regret Of course, if the environment chooses large losses $\ell_t(x)$ for all decisions $\theta \in \Theta$, it is impossible for the player to ensure small cumulative loss. Therefore, one needs a relative criterion: the regret of the player is the difference between the cumulative loss he incurred and that of the best fixed decision in hindsight.

Definition 1 (Regret)**.** *The regret of the player with respect to a fixed parameter $\theta^* \in \Theta$ after T time steps is*

$$R_T(\theta^*) \stackrel{def}{=} \sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \ell_t(\theta^*)\,.$$

The regret (or uniform regret) is defined as $R_T \stackrel{def}{=} \sup_{\theta^* \in \Theta} R_T(\theta^*)$.

We have some bias-variance decomposition:

$$\sum_{t=1}^{T} \ell_t(\theta_t) \;=\; \underbrace{\inf_{\theta\in\Theta} \sum_{t=1}^{T} \ell_t(\theta)}_{\text{Approximation error = how good the possible actions are.}} \;+\; \underbrace{R_T}_{\text{Sequential estimation error of the best action}}$$

We will focus on the regret in these lectures. The goal of the player is to ensure a sublinear regret $R_T = o(T)$ as $T \to \infty$ and this for any possible sequence of losses $\ell_1, \ldots, \ell_T$. In this case, the average performance of the player will approach on the long term the one of the best decision.

Remarks Let us makes some remarks:

- Except in the stochastic bandit part, we will not make any random assumption on the process generating the losses ℓ_t. The latter are deterministic and may be chosen by some adversary. Typically, the problem can be seen as a game between the player who aims at optimizing with respect to $\theta_1, \ldots, \theta_T$ against an environment who aims at mazimizing with respect to $\ell_t, \ldots, \ell_T$ and θ^*. Player's goal is to approach the quantity:
$$\inf_{\theta_1} \sup_{\ell_1} \inf_{\theta_2} \sup_{\ell_2} \ldots \inf_{\theta_T} \sup_{\ell_T} \sup_{\theta^*\in\Theta} R_T(\theta^*)\,.$$
- Note that the loss functions ℓ_t depend on the round t. This may be caused by many phenomena. We provide here some possible reasons. This may be because
  - of some observation to be predicted if $\ell_t(x) = \ell(x, y_t)$. For instance, if the goal is to predict the evolution of the temperature $y_1, \ldots, y_T$, the latter changes over time and a prediction x is evaluated with $\ell_t(x) = (x - y_t)^2$.
  - the environment is stochastic and the variation over time t models some noise effect.
  - of a changing environment. For instance, if the player is playing a game against some adversary that evolves and adapts to its strategy. A typical example is the case of spam detections. If the player tries to detect spams, while some spammers (the environment) try at the same time to fool the player with new spam strategies.

Exercise 2.1. Instead considering the regret with respect to a fixed $\theta^* \in \Theta$, one would be tempted to minimize the quantity

$$R_T^* \;\stackrel{\text{def}}{=}\; \sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \inf_{\theta\in\Theta} \ell_t(\theta)$$

where the infimum is inside the sum. Show that the environment can ensure R_T^* to be linear in T by choosing properly the loss functions ℓ_t.

3 Full information feedback with linear loss functions

We will start with the simple case where the decision set Θ is the K-dimensional simplex

$$\Delta_K \;\stackrel{\text{def}}{=}\; \{p \in [0,1]^K : \sum_{k=1}^{K} p_k = 1\}\,. \qquad \text{(decision set)}$$

Since the decisions θ_t are probability distributions in $\Theta = \Delta_K$, in this part we will denote them by p_t instead of θ_t. We assume that the loss functions ℓ_t are linear

$$\forall p \in \Theta, \qquad \ell_t(p) = \sum_{k=1}^{K} p(k) g_t(k) \in [-1, 1] \qquad \text{(linear loss)}$$

where $g_t = (g_t(1), \ldots, g_t(K)) \in [-1,1]^K$ is a loss vector chosen by the environment at round t.

3.1 The exponentially weighed average forecaster

How to choose the weights p_t? At round t the player needs to choose a weight vector $p_t \in \Delta_K$. The question is how to choose it? The idea is to give more weight to actions that performed well in the past. But we should not give all the weight to the current best action, otherwise it would not work (see exercises). The exponentially weighted average forecaster (EWA) also called Hedge performs this trade-off by choosing a weight that decreases exponentially fast with the past errors.

The Exponentially weighted average forecaster (EWA)

Parameter: $\eta > 0$
Initialize: $p_1 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$
For $t = 1, \dots, T$
- select p_t; incur loss $\ell_t(p_t) = p_t^\top g_t$ and observe $g_t \in [-1,1]^K$;
- update for all $k \in \{1, \dots, K\}$

$$p_{t+1}(k) = \frac{e^{-\eta \sum_{s=1}^t g_s(k)}}{\sum_{j=1}^K e^{-\eta \sum_{s=1}^t g_s(j)}} \,.$$

Exercise 3.1. Consider the strategy, called "Follow The Leader" (FTL) that puts all the mass on the best action so far:

$$p_t \in \arg\min_{p \in \Theta} \sum_{s=1}^{t-1} \ell_s(p) \,. \tag{FTL}$$

1. Show that $p_t(k) > 0$ implies that $k \in \arg\min_j \sum_{s=1}^{t-1} g_s(j)$
2. Show that the regret of FTL might be linear: i.e., there exists a sequence $g_1, \dots, g_T \in [-1,1]^K$ such that $R_T \geq (1 - 1/K)T$.

The following theorem proves that EWA, which is a smoothed version of FTL, achieves sublinear regret.

Theorem 1. *Let $T \geq 1$. For all sequences of loss vectors $g_1, \dots, g_T \in [-1,1]^K$, EWA achieves the bound*

$$R_T \overset{def}{=} \sum_{t=1}^T \ell_t(p_t) - \min_{p \in \Delta_K} \sum_{t=1}^T \ell_t(p) \leq \eta \sum_{t=1}^T \sum_{k=1}^K p_t(k) g_t(k)^2 + \frac{\log K}{\eta} \,, \tag{1}$$

where we recall $\ell_t : p \in \Delta_K \mapsto p^\top g_t$. Therefore, for the choice $\eta = \sqrt{\frac{\log K}{T}}$, EWA satisfies the regret bound $R_T \leq 2\sqrt{T \log K}$.

This regret bound is optimal (see Cesa-Bianchi and Lugosi [2006]).

Exercise 3.2. Generalize the above theorem when the losses $\ell_1, \dots, \ell_T \in [-B, B]$ for some $B > 0$.

Proof. First, we remark that by definition of $\ell_t : p \mapsto p \cdot g_t$ we have

$$\begin{aligned} R_T &\overset{\text{def}}{=} \sum_{t=1}^T \ell_t(p_t) - \min_{p \in \Delta_K} \sum_{t=1}^T \ell_t(p) \\ &= \sum_{t=1}^T p_t \cdot g_t - \min_{p \in \Delta_K} \sum_{t=1}^T p \cdot g_t \\ &= \sum_{t=1}^T p_t \cdot g_t - \min_{p \in \Delta_K} \sum_{k=1}^K \sum_{t=1}^T p(k) g_t(k) \,. \end{aligned}$$

Now, we can see that the minimum over $p \in \Delta_K$ is reached on a corner of the simplex. Therefore

$$R_T = \sum_{t=1}^{T} p_t \cdot g_t - \min_{1 \le k \le K} \sum_{t=1}^{T} g_t(k)\,.$$

We denote $W_t(j) = e^{-\eta \sum_{s=1}^{t} g_t(j)}$ and $W_t = \sum_{j=1}^{K} W_t(j)$. The proof will consist in upper-bounding and lower-bounding W_T. We have

$$\begin{aligned}
W_t &= \sum_{j=1}^{K} W_{t-1}(j) e^{-\eta g_t(j)} && \leftarrow \quad W_t^{(j)} = W_{t-1}(j) e^{-\eta g_t(j)} \\
&= W_{t-1} \sum_{j=1}^{K} \frac{W_{t-1}(j)}{W_{t-1}} e^{-\eta g_t(j)} \\
&= W_{t-1} \sum_{j=1}^{K} p_t(j) e^{-\eta g_t(j)} && \leftarrow \quad p_t(j) = \frac{e^{-\eta \sum_{s=1}^{t-1} g_s(j)}}{\sum_{k=1}^{K} e^{-\eta \sum_{s=1}^{t-1} g_s(k)}} = \frac{W_{t-1}(j)}{W_{t-1}} \\
&\le W_{t-1} \sum_{j=1}^{K} p_t(j) \big(1 - \eta g_t(j) + \eta^2 g_t(j)^2\big) && \leftarrow \quad e^x \le 1 + x + x^2 \text{ for } x \le 1 \\
&= W_{t-1}(1 - \eta p_t \cdot g_t + \eta^2 p_t \cdot g_t^2)\,,
\end{aligned}$$

where we assumed in the inequality $-\eta g_t(j) \le 1$ and where we denote $g_t = (g_t(1), \ldots, g_t(K))$, $g_t^2 = \big(g_t(1)^2, \ldots, g_t(K)^2\big)$ and $p_t = (p_t(1), \ldots, p_t(K))$. Now, using $1 + x \le e^x$, we get:

$$W_t \le W_{t-1} \exp\big(-\eta p_t \cdot g_t + \eta^2 p_t \cdot g_t^2\big)\,.$$

By induction on $t = 1, \ldots, T$, this yields using $W_0 = K$

$$W_T \le K \exp\Big(-\eta \sum_{t=1}^{T} p_t \cdot g_t + \eta^2 \sum_{t=1}^{T} p_t \cdot g_t^2\Big)\,. \tag{2}$$

On the other hand, upper-bounding the maximum with the sum,

$$\exp\Big(-\eta \min_{j \in [K]} \sum_{t=1}^{T} g_t(j)\Big) \le \sum_{j=1}^{K} \exp\Big(-\eta \sum_{t=1}^{T} g_t(j)\Big) \le W_T\,.$$

Combining the above inequality with Inequality (2) and taking the log, we get

$$-\eta \min_{j \in [K]} \sum_{t=1}^{T} g_t(j) \le -\eta \sum_{t=1}^{T} p_t \cdot g_t + \eta^2 \sum_{t=1}^{T} p_t \cdot g_t^2 + \log K\,. \tag{3}$$

Dividing by η and reorganizing the terms proves the first inequality:

$$R_T = \sum_{t=1}^{T} p_t \cdot g_t - \min_{1 \le j \le K} \sum_{t=1}^{T} g_t(j) \le \eta \sum_{t=1}^{T} p_t \cdot g_t^2 + \frac{\log K}{\eta}$$

Optimizing η and upper-bounding $p_t \cdot g_t^2 \le 1$ concludes the second inequality. □

Anytime algorithm (the doubling trick) The previous algorithm EWA depends on a parameter $\eta > 0$ that needs to be optimized according to K and T. For instance, for EWA using the value

$$\eta = \sqrt{\frac{\log K}{T}}\,.$$

the bound of Theorem 1 is only valid for horizon T. However, the learner might not know the time horizon in advance and one might want an algorithm with guarantees valid simultaneously for all $T \geq 1$. We can avoid the assumption that T is known in advance, at the cost of a constant factor, by using the so-called *doubling trick*. The general idea is the following. Whenever we reach a time step t which is a power of 2, we restart the algorithm (forgetting all the information gained in the past) setting η to $\sqrt{\log K/t}$. Let us denote EWA-doubling this algorithm.

Theorem 2 (Anytime bound on the regret)**.** *For all $T \geq 1$, the pseudo-regret of EWA-doubling is then upper-bounded as:*

$$R_T \leq 7\sqrt{T \log K}\,.$$

The same trick can be used to turn most online algorithms into anytime algorithms (even in more general settings: bandits, general loss,...). We can use the *doubling trick* whenever we have an algorithm with a regret of order $\mathcal{O}(T^\alpha)$ for some $\alpha > 0$ with a known horizon T to turn it into an algorithm with a regret $\mathcal{O}(T^\alpha)$ for all $T \geq 1$.

Another solution is to use time-varying parameters η_t replacing T with the current value of t. The analysis is however less straightforward.

Exercise 3.3. Prove a regret bound for the time-varying choice $\eta_t = \sqrt{\log K/t}$ in EWA.

Proof of Theorem 2. For simplicity we assume $T = 2^{M+1} - 1$. The regret of EWA-doubling is then upper-bounded as:

$$\begin{aligned}
R_T &= \sum_{t=1}^{T} \ell_t(p_t) - \min_{p \in \Delta_K} \sum_{t=1}^{T} \ell_t(p) \\
&\leq \sum_{t=1}^{T} \ell_t(p_t) - \sum_{m=0}^{M} \min_{p \in \Delta_K} \sum_{t=2^m}^{2^{m+1}-1} \ell_t(p) \\
&= \sum_{m=0}^{M} \underbrace{\sum_{t=2^m}^{2^{m+1}-1} \ell_t(p_t) - \min_{p \in \Delta_K} \sum_{t=2^m}^{2^{m+1}-1} \ell_t(p)}_{R_m}\,.
\end{aligned}$$

Now, we remark that each term R_m corresponds to the expected regret of an instance of EWA over the 2^m rounds $t = 2^m, \ldots, 2^{m+1} - 1$ and run with the optimal parameter $\eta = \sqrt{\log K/2^m}$. Therefore, using Theorem 1, we get $R_m \leq 2\sqrt{2^m \log K}$, which yields:

$$R_T \leq \sum_{m=0}^{M} 2\sqrt{2^m \log K} \leq 2(1+\sqrt{2})\sqrt{2^{M+1} \log K} \leq 7\sqrt{T \log K}\,.$$

□

Improvement for small losses The first inequality in Theorem 1 is sometimes called improvement for small losses when losses take values in $[0, 1]$. Let's define $\widehat{L}_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} \ell_t(p_t)$ the loss of the algorithm and

$L_T^* \stackrel{\text{def}}{=} \min_{p\in\Delta_K} \sum_{t=1}^T \ell_t(p)$. Then, the regret is upper-bounded by

$$\begin{aligned} R_T \stackrel{\text{def}}{=} \widehat{L}_T - L_T^* &\leq \frac{\log K}{\eta} + \eta \sum_{t=1}^T p_t \cdot g_t^2 \\ &\leq \frac{\log K}{\eta} + \eta \sum_{t=1}^T p_t \cdot g_t = \frac{\log K}{\eta} + \eta \widehat{L}_T \,. \end{aligned}$$

Therefore, rearranging the terms

$$(1-\eta)\widehat{L}_T - (1-\eta)L_T^* \leq \frac{\log K}{\eta} + \eta L_T^* \,,$$

which implies

$$R_T \leq \frac{\log K}{\eta(1-\eta)} + \frac{\eta}{1-\eta} L_T^* \,.$$

Optimising in $\eta \approx \sqrt{(\log K)/L_T^*}$ we get $R_T \lesssim \sqrt{(\log K) L_T^*}$ which is small whenever some parameter achieves a small cumulative loss.

3.2 Application to prediction with expert advice

The preceding section considers linear loss functions. Yet, it can yield non-trivial regret bounds for general convex losses. We consider here an application to the setting of prediction with expert advice detailed in Example 2.2. The goal is to minimize the regret with respect to the best expert

$$R_T^{\text{expert}} \stackrel{\text{def}}{=} \sum_{t=1}^T \ell(\widehat{y}_t, y_t) - \min_{1\leq k\leq K} \sum_{t=1}^T \ell(x_t(k), y_t) \,,$$

where $\widehat{y}_t = p_t \cdot x_t$ are the prediction of the algorithm and y_t the observations to be predicted sequentially.

Convex loss function ℓ. We state bellow a corrolary to Theorem 1 when the loss functions $\ell(\cdot,\cdot)$ are convex in there first argument.

Corollary 1 (Regret of EWA for prediction with expert advice and convex loss)**.** *Let $T \geq 1$. Assume that the loss function $\ell : (x,y) \in \mathbb{R}\times\mathbb{R} \mapsto \mathbb{R}$ is convex and takes values in $[-1,1]$. Then, EWA applied with the vector vectors $g_t = \big(\ell(x_t(1), y_t), \ldots, \ell(x_t(K), y_t)\big) \in [-1,1]^K$ has a regret upper-bounded by*

$$R_T^{expert} \leq 2\sqrt{T \log K}$$

where $\widehat{y}_t = p_t \cdot x_t$ and were $\eta > 0$ is well-tuned.

Therefore, the average error of the algorithm will converge to the average error of the best expert. This is the case for the square loss, the absolute loss or the absolute percentage of error.

Proof. It suffices to remark that by convexity of $\ell(\cdot,\cdot)$ in its first argument

$$\begin{aligned} R_T^{\text{expert}} &= \sum_{t=1}^T \ell(p_t \cdot x_t, y_t) - \min_{1\leq k\leq K} \sum_{t=1}^T \ell(x_t(k), y_t) \\ &\leq \sum_{t=1}^T p_t \cdot g_t - \min_{1\leq k\leq K} \sum_{t=1}^T g_t(k) \stackrel{\text{def}}{=} R_T. \end{aligned}$$

The result is then obtained by Theorem 1. □

Exp-concave loss function Here, we show that a faster rate can be obtained (with EWA) if the loss function are exp-concave.

Definition 2 (η-exp-concavity)**.** *For $\eta \in \mathbb{R}$, a function f is said to be η-exp-concave if $x \mapsto e^{-\eta f(x)}$ is concave.*

Exp-concavity is stronger than convexity but weaker than strong convexity. Indeed, exp-concave functions are convex because $-\log$ is convex and decreasing. Furthermore, any η-exp-concave function is also η'-exp-concave for $0 \leq \eta' \leq \eta$.
In prediction with expert advice, if the loss are generated from a fixed loss function $\ell_t(p) = \ell(p \cdot \ell_t, y_t)$, then ℓ_t are η-expconcave if $\widehat{y} \mapsto \ell(\widehat{y}, y_t)$ are η-exp-concave for all y_t. We can compute η for some common loss functions:

- *the squared loss*: $\ell : (\widehat{y}, y) \in [0,1]^2 \mapsto (\widehat{y} - y)^2$, then ℓ_t are 1/2-exp-concave. Indeed, let $y \in [0,1]$ and denote $G : \widehat{y} \mapsto \exp\big(-\eta(\widehat{y} - y)^2\big)$. Then, $G''(\widehat{y}) = G(\xi)\big(4\eta^2(\xi - y)^2 - 2\eta\big)$. Thus G is concave if and only if $(\xi - y)^2 \leq 1/(2\eta)$ which is satisfied for $\eta = 1/2$. This is also the case in higher dimensions with $\ell(\widehat{y}, y) = \|\widehat{y} - y\|^2$. If the observations and prediction $\widehat{y}, y \in [0, B]$, then the ℓ_t are $1/(2B^2)$-exp-concave
- *the relative entropy* (or Kullback–Leibler divergence): $\ell : (\widehat{y}, y) \in [0,1]^2 \mapsto y \log(y/\widehat{y}) - (1-y)\log((1-y)/(1-\widehat{y}))$. Then the functions ℓ_t are 1-exp-concave. This loss can for instance used for density estimation of the sequence $y_1, \ldots, y_T$.
- the linear loss $\ell(\widehat{y}, y) = \widehat{y} \cdot y$, the absolute loss $\ell(\widehat{y}, y) = |\widehat{y} - y|$ or the absolute percentage of error are however not η-exp-concave for any $\eta > 0$.

Corollary 2 (Regret of EWA for prediction with expert advice and exp-concave loss)**.** *In the setting of prediction with expert advice, if the loss functions $\ell(\cdot, y_t)$ are η-exp-concave for all y_t, then EWA run with vectors $g_t = \big(\ell(x_t(1), y_t), \ldots, \ell(x_t(K), y_t)\big) \in \mathbb{R}^K$ with parameter $\eta > 0$ satisfies*

$$R_T^{expert} \leq \frac{\log K}{\eta}\,,$$

for all $T \geq 1$.

The worst-case regret does not increase with T but grows logarithmically in the dimension K.

Proof. The proof is similar to the original proof of EWA. We define $W_t(i) = e^{-\eta \sum_{s=1}^t g_s(i)}$ and $W_t = \sum_{i=1}^K W_t(i)$. We have

$$\begin{aligned}
W_t &= \sum_{j=1}^N W_{t-1}(j) e^{-\eta g_t(j)} && \leftarrow \quad W_t(j) = W_{t-1}(j) e^{-\eta g_t(j)} \\
&= W_{t-1} \sum_{j=1}^N \frac{W_{t-1}(j)}{W_{t-1}} e^{-\eta g_t(j)} \\
&= W_{t-1} \sum_{j=1}^N p_t(j) e^{-\eta g_t(j)} && \leftarrow \quad p_t(j) = \frac{e^{-\eta \sum_{s=1}^{t-1} g_s(j)}}{\sum_{k=1}^N e^{-\eta \sum_{s=1}^{t-1} g_s(k)}} = \frac{W_{t-1}(j)}{W_{t-1}} \\
&\leq W_{t-1} \exp\big(-\eta \ell(p_t \cdot x_t, y_t)\big) && \leftarrow \quad \eta\text{-exp-concavity}
\end{aligned}$$

Now, by induction on $t = 1, \ldots, T$, this yields using $W_0 = K$

$$W_T \leq K \exp\left(-\eta \sum_{t=1}^T \ell(\widehat{y}_t, y_t)\right). \tag{4}$$

On the other hand, upper-bounding the maximum with the sum,

$$\exp\Big(-\eta \min_{j \in [K]} \sum_{t=1}^T g_t(j)\Big) \leq \sum_{j=1}^K \exp\Big(-\eta \sum_{t=1}^T g_t(j)\Big) \leq W_T\,.$$

Combining the above inequality with Inequality (4) and taking the log concludes the proof. □

Continuous EWA Similarly as for Section 3.2, Theorem 2 does not really control the true regret R_T since it controls the regret with respect to Dirac mass $\min_{1\le k\le K}\sum_{t=1}^T \ell_t(\delta_k)$ instead of the one with respect to all convex combinations $\min_{p\in\mathcal{X}}\sum_{t=1}^T \ell_t(p)$. A true upper-bound on the regret can be obtained by using a continuous version of EWA:

$$p_t = \frac{\int_{\mathcal{X}} p e^{-\eta\sum_{s=1}^{t-1}\ell_s(p)} d\mu(p)}{\int_{\mathcal{X}} e^{-\eta\sum_{s=1}^{t-1}\ell_s(p)} d\mu(p)}\,,$$

where μ is the uniform (Lebesgue) measure on $\mathcal{X} = \Delta_K$.

Theorem 3. *Let $T \ge 1$. For all sequences of η-exp-concave losses $\ell_1, \dots, \ell_t$ the continuous EWA forecaster satisfies*

$$R_T \stackrel{def}{=} \sum_{t=1}^T \ell_t(p_t) - \inf_{p\in\mathcal{X}} \sum_{t=1}^T \ell_t(p) \le \frac{1+(K-1)\log(T+1)}{\eta}$$

Proof. The proof starts similarly to the one of Theorem 2. Let us denote $W_t(p) = e^{-\eta\sum_{s=1}^t \ell_s(p)}$, $W_t = \int_{\mathcal{X}} W_t(p)d\mu(p)$ and $d\widehat{\mu}_t(p) = W_t(p)d\mu(p)/W_t$. Then,

$$\begin{aligned}
W_T &= \int_{\mathcal{X}} e^{-\eta\sum_{t=1}^T \ell_t(p)} d\mu(p) \\
&= W_{T-1}\int_{\mathcal{X}} \frac{W_{T-1}(p)}{W_{T-1}} e^{-\eta\ell_T(p)} d\mu(p) \\
&= W_{T-1}\int_{\mathcal{X}} e^{-\eta\ell_T(p)} d\widehat{\mu}_{T-1}(p) && \leftarrow \quad p_T = \int_{\mathcal{X}} p\, d\widehat{\mu}_{T-1}(p) \\
&\le W_{T-1}\exp\big(-\eta\ell_T(p_T)\big) && \leftarrow \quad \eta\text{-exp-concavity} \\
&\le \exp\left(-\eta\sum_{t=1}^T \ell_t(p_t)\right), && \leftarrow \quad \text{induction} \qquad (5)
\end{aligned}$$

The second part of the proof to lower-bound W_T is however less straightforward. For simplicity, let us assume that ℓ_t are continuous on $\mathcal{X}$ (do the general case as exercise). Therefore the infimum is a minimum and let $p^* \in \arg\min_{p\in\mathcal{X}}\sum_{t=1}^T \ell_t(p)$ and define

$$\mathcal{X}_\varepsilon \stackrel{\text{def}}{=} \Big\{(1-\varepsilon)p^* + \varepsilon q, \quad q\in\mathcal{X}\Big\}, \qquad \varepsilon\in(0,1)\,.$$

By expconcavity of ℓ_t, we have for all t and all $p = (1-\varepsilon)p^* + \varepsilon q$

$$e^{-\eta\ell_t(p)} \ge (1-\varepsilon)e^{-\eta\ell_t(p^*)} + \varepsilon e^{-\eta\ell_t(q)} \ge (1-\varepsilon)e^{-\eta\ell_t(p^*)}$$

Therefore, for all $p\in\mathcal{X}_\varepsilon$

$$e^{-\eta\sum_{t=1}^T \ell_t(p)} \ge (1-\varepsilon)^T e^{-\eta\sum_{t=1}^T \ell_t(p^*)}$$

Integrating both parts over $\mathcal{X}_\varepsilon$ and using $\mu(\mathcal{X}_\varepsilon) = \varepsilon^{K-1}\mu(\mathcal{X})$ (exercise) we get

$$W_T \ge \int_{\mathcal{X}_\varepsilon} e^{-\eta\sum_{t=1}^T \ell_t(p)} d\mu(p) \ge \mu(\mathcal{X})\varepsilon^{K-1}(1-\varepsilon)^T e^{-\eta\sum_{t=1}^T \ell_t(p^*)}\,.$$

Combining with (5), using $W_0 = \mu(\mathcal{X})$, taking the log and reorganizing the terms yields

$$R_T \stackrel{\text{def}}{=} \sum_{t=1}^T \ell_t(p_t) - \sum_{t=1}^T \ell_t(p^*) \le \frac{(K-1)\log\frac{1}{\varepsilon} + T\log\frac{1}{1-\varepsilon}}{\eta}\,.$$

Optimizing $\varepsilon = 1/(T+1)$ concludes the proof since

$$T \log \frac{1}{1-\varepsilon} = T \log \left(1 + \frac{1}{T}\right) \leq 1 \,.$$

□

Though the nice theoretical result, this algorithm is complicated to implement because of the integral. In practice, p_t can be computed by using $(1/T)$-discretization grid of $\mathcal{X}$ (bad complexity of order T^K!) or by using Monte-Carlo methods to approximate the integral. We will see in next lectures efficient algorithms with similar guarantees.

References

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.