Prediction of individual sequences : homework

Lucas CLARTE

06 mars 2020

Part 1 - Link between online learning and game theory

All the code is written in the Python script `homework.py`.

1. For the game "Rock Paper scissors", we define the actions rock = 1, paper = 2, scissors = 3. The loss matrix L is as follows

$$L = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

2. (a) We implement the function `rand_weighted(p)` that samples $\imath \in [|M|]$ with respect to the distribution $p \in \Delta_M$. In Python, the function is simply

```
def rand_weighted(p) :
        return np.argwhere(np.cumsum(p) > np.random.rand())[0, 0]
```

In this function, we build an array `c = np.cumsum(p)` that contains the cumulative probabilities of p i.e $c_i = p_1 + ... + p_i$. We then sample a number $x \in [0, 1)$ and determine the first index i in the cumulative array c such that $c_i > x$. It is easy to show that for all j, $i = j$ with probability p_j.

2. (b) To implement the function `EWA_update(p, l)` we simply multiply the vectors p and $\exp(-\eta l_t(i))$ component-wise and then normalize the new vector.

3. (a.)) We simulate EWA against a fixed adversary with strategy $q = (0.5, 0.25, 0.25)$, in the function `EWA(L, T, eta, q`. This function runs T iterations. At iteration t, we sample an action j_t from q using `rand_weighted` and update the weights p_t with the function `EWA_update` called with the loss l_t. The loss $l_t(i)$ of the player if he choses the action i is equal to $L(i, j_t)$.

3. (b) As asked, we simulate the game with $T = 100, \eta = 1.0$ and plot the weight vectors $p_1, ..., p_T$ in the figure (**??**) . We see that the best strategy is $p = (0.0, 1.0, 0.0)$. We can prove it rigorously. Indeed, consider the strategy $p = (x, y, z)$ that minimizes the average loss $l(p, q) = \mathbb{E}_{i \sim p, j \sim q}(L(i, j))$. We write

$$\begin{aligned} l(p, q) &= x \times (0.25 - 0.25) + y \times (0.25 - 0.5) + z \times (0.5 - 0.25) \\ &= -0.25 \times y + 0.25 \times z \end{aligned}$$

The optimal value of p that satisifies the constraint $p \in \Delta_3$ and minimizes the above expression is $x = 0, y = 1, z = 0$ which is what we wanted.

3. (c) We plot the average loss $\bar{l}_t = \frac{1}{t} \sum_{1 \leqslant s \leqslant t} l(i_s, j_s)$ as a function of t, and obtain the figure (**??**). The figure shows 10 different runs of the simulation.

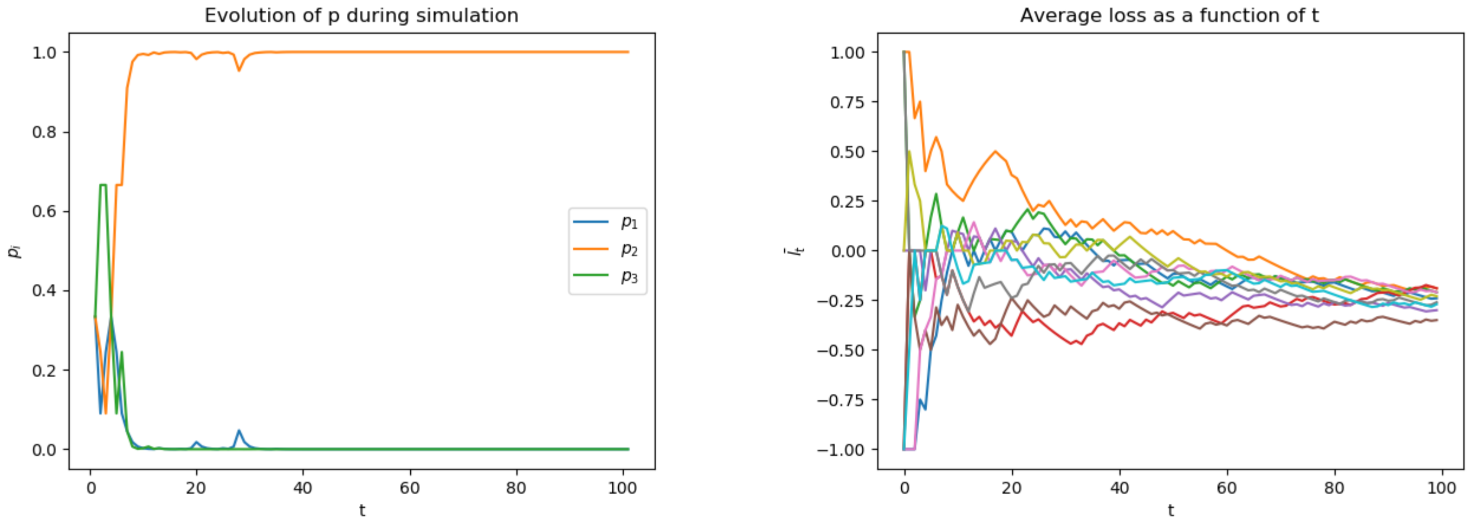


Figure 1: Left : plot of the player's stragegy p with `EWA`, when opponent has a fixed strategy (0.5, 0.25, 0.25). Right : plot of the average loss as a function of t for 10 simulations of the game with `EWA`

3. (d) Here, the cumulative regret is defined by

$$R_t = \sum_{s=1}^{t} L(i_s, j_s) - \min_i \sum_{s \leqslant t} L(i, j_s) \tag{1}$$

Figure (**??**) shows the cumulative regret for 10 trials with T = 100. We observe experimentally that for each trial, after a certain time the cumulative regret is constant.

3. (e) To estimate the stability, n = 10 simulations are executed, and we plot in figure (**??**) the n average losses as well as the minimum, maximum and mean losses. At first glance, the `EWA` seems to be stable. Indeed, the maximum and minimum both seem to converge towards the value $l_\infty = -0.25$.

3. (f) In theory, the best learning rate η with `EWA` is $\eta_{\text{EWA}} = \sqrt{2ln(K)/T}$ with $K = 3$ here and $T = 100$, so we obtain $\eta_{\text{EWA}} \simeq 0.15$. We see that in practice, the best learning rate is not necessarily equal to η_{EWA}.

4. *Simulation against an adaptive adversary* In this question, the player still uses `EWA` with a learning rate $\eta = 1.0$ while the opponent uses a learning rate $\eta = 0.05$

4.(a) When the adversary uses `EWA` like the player, we observe that the average loss seems to converge towards 0 which is the value of the game.

4. (b)

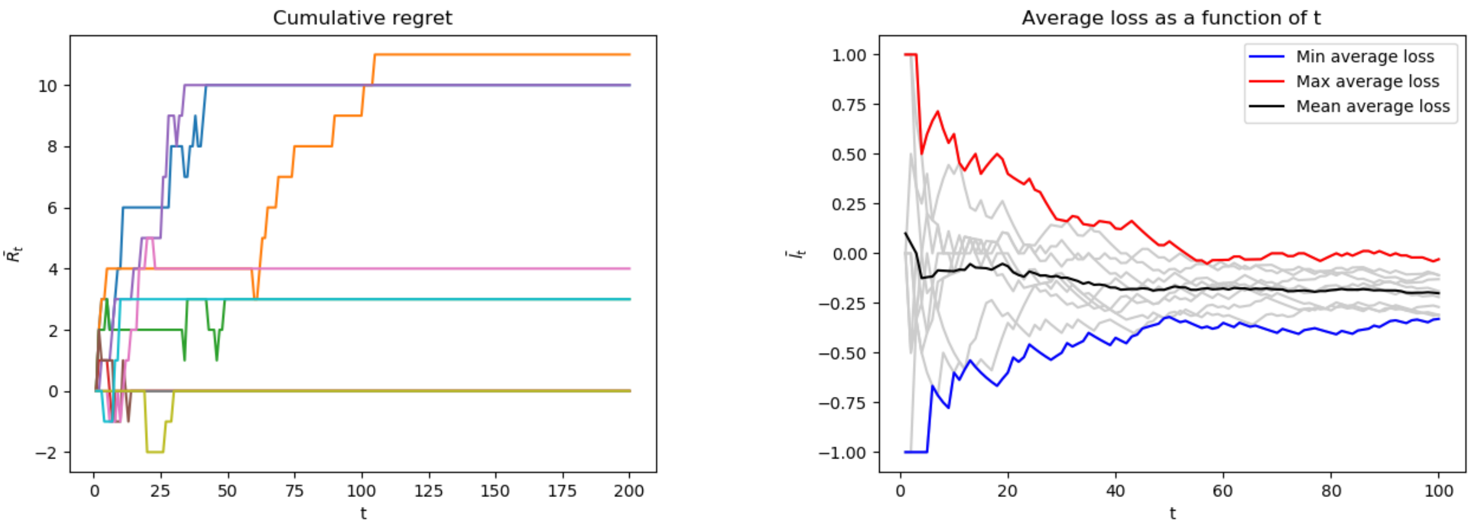


Figure 2: Left : Cumulative regret .Right : average, minimum and maximum of the average loss $\bar{l}_t$ for n = 10 simulation trials. The n average losses are plotted in grey.

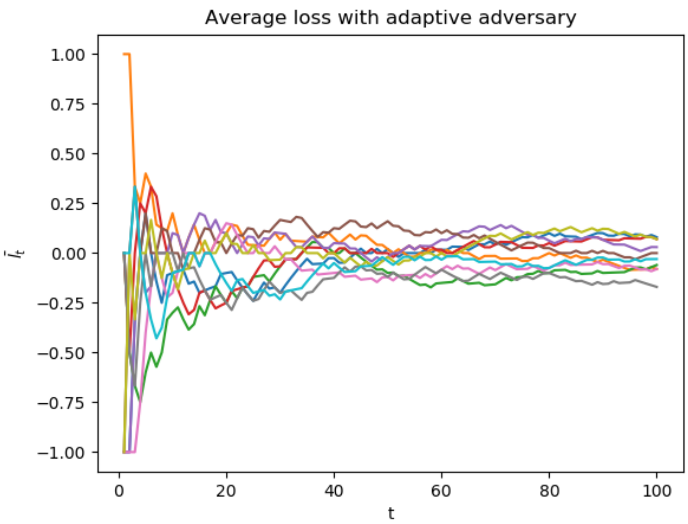
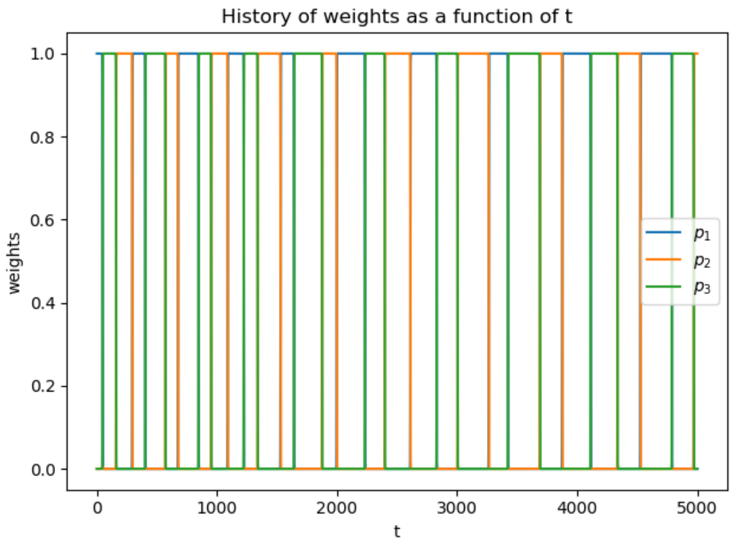
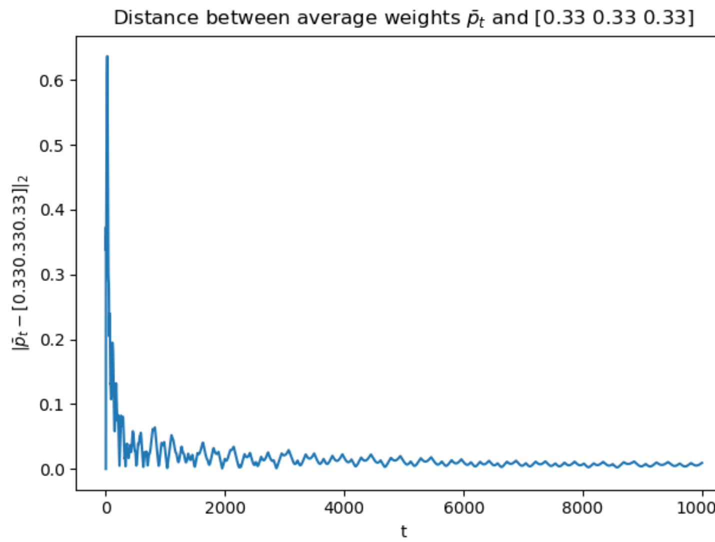


Figure 3: Average loss for n = 10 simulations with adaptive adversary





Bandit feedback We now assume that the player doesn't have access to L but only the incured loss $L(i_t, j_t)$. We will here implement the algorithm `Exp3` where at each time t, the action i_t is selected with a probability

$$\mathbb{P}(i_t = i) = \frac{\exp\left(-\eta \hat{l}_t(i)\right)}{\sum_j \exp\left(-\eta \hat{l}_t(j)\right)} \tag{2}$$

where $\hat{l}_t(i)$ is the *estimated loss* of the action i, i.e the average loss incured by the player when the action i was selected.

5.(a) To implement the function `estimated_loss`, we maintain a vector $N_t \in \mathbb{N}^M$ where for each t, $N_t(i)$ is the number of times the action i has been played until t. We also maintain the vector $S_t \in \mathbb{R}^M$ where $S_t(i) = \sum_t L(i_t, j_t)\mathbf{1}(i_t = i)$ is the sum of the losses incured by the player every time they chose the action i. The function `estimated_loss(i)` simply returns $\hat{l}_t(k) = S_t(k)/N_t(k)$.

5. (b) In the function `EXP3_update`, we compute the weights vector p_{t+1} given the weight vector

$$p_t = \frac{\exp -\eta \hat{l}_t(k)}{\sum_j \exp -\eta \hat{l}_t(k)}$$

and the loss $L(i_t, j_t)$. To do this, we use the fact that for k such that $i_t = k$, we have

$$\hat{l}_{t+1}(k) = \hat{l}_t(k) + \frac{1}{t+1}(L(i_t, j_t) - \hat{l}_t)$$

We thus define a vector $q(k) = \mathbf{1}(k = i_t)\frac{1}{t+1}(L(i_t, j_t) - \hat{l}_t)$ and use the following identity to compute p_{t+1} :

$$p_{t+1}(k) = \frac{p_t(k) \cdot \exp\left(-\eta q(k)\right)}{\sum_l p_t(l) \cdot \exp\left(-\eta q(l)\right)}$$

In essence, we do the same operation as the function `EWA_update` but here we use the vector q instead of the complete loss l_t.

6. In this question, the player uses `Exp3` with a learning rate $\eta = 1.0$ with a fixed adversary.

7. In this question, the opponent uses a strategy `EWA` with learning rate 0.05.

Part 2 - Theory - Sleeping experts

11. (a) Let us note $f(x) := log(1 + x)$ and $g(x) := x - x^2$. Let us define

$$\Delta(x) = f(x) - g(x)$$

We want to prove $\Delta(x) \geqslant 0$ for $x \geqslant -\frac{1}{2}$. Note that we have

$$\forall x, \Delta'(x) = \frac{1}{1+x} - 1 + 2x$$

It is easy to show that $\Delta'(x) < 0$ if and only if $-\frac{1}{2} \leqslant x \leqslant 0$. As a consequence, Δ is a non-increasing function on $[-0.5, 0]$ and a non-decreasing function on $[0, \infty]$. Since $\Delta(0) = 0$, this shows that $\Delta \geqslant 0$ when $x \geqslant -0.5$.

11. (b) Let $k \in \mathcal{X}$. Because the weights $w_t(j)$ are all positive, we can write

$$\begin{aligned}\log W_{T+1} &= \log \sum_j w_{T+1}(j) \geqslant \log w_{T+1}(k) \\ &= \sum_{t=1}^{T} \log\left(1 + \eta(k)(p_t \cdot l_t - l_t(k))\right)\end{aligned}$$

Because the loss is between 0 and 1, $p_t \cdot l_t - l_t(k) \in [-1, 1]$ and because $\eta(k) \leqslant 0.5$, we deduce that $1 + \eta(k)(p_t \cdot l_t - l_t(k)) \geqslant 0.5$ and we can apply the previous question. From there, we deduce the result

$$\log W_{T+1} \geqslant \eta(k) \sum_t (p_t \cdot l_t - l_t(k)) - \eta(k)^2 \sum_t (p_t \cdot l_t - l_t(k))^2$$

11. (c) Consider $k \in \mathcal{X}$. By definition of w_t, we have

$$\begin{aligned}w_{t+1}(k) &= w_t(k) \times (1 + \eta(k)(p_t \cdot l_t - l_t(k))) \\ w_{t+1}(k) - w_t(k) &= w_t(k) \times \eta(k)(p_t \cdot l_t - l_t(k))\end{aligned}$$

Thus, summing over all k, we have

$$W_{t+1} - W_t = \sum_k w_t(k) \times \eta(k)(p_t \cdot l_t - l_t(k))$$

However, developping $p_t = \frac{\sum_k \eta(k) w_t(k)}{\sum_l \eta(k) w_t(k)}$, we have

$$\sum_k w_t(k)\eta(k)(l_t \cdot p_t) = \sum_k \eta(k) w_t(k) l_t(k)$$

And thus

$$W_{t+1} - W_t = 0$$

Since we have $w_1(k) = 1$ for all k, $W_1 = K$ so at each time t, $W_t = K$.

11. (d) Using the two previous questions, we deduce that for all k, we have

$$\log K \geqslant \eta(k) \sum_t (p_t \cdot l_t - l_t(k)) - \eta(k)^2 \sum_t (p_t \cdot l_t - l_t(k))^2 \tag{3}$$

The right-hand side expression is a concave function on $\eta(k)$, its maximum is reached when its derivative w.r.t $\eta(k)$ is 0, i.e when

$$\eta(k) = \frac{1}{2} \frac{\sum_t (p_t \cdot l_t - l_t(k))}{\sum_t (p_t \cdot l_t - l_t(k))^2} \leqslant \frac{1}{2} \tag{4}$$

Substituting this expression for η in the equation (3) yields the asked inequality :

$$\sum_t (p_t \cdot l_t - l_t(k)) \leqslant 2\sqrt{(\log(K) \sum_t (p_t \cdot l_t - l_t(k))^2)}$$

12.