

Lecture 8: May 3, 2010

Lecturer: Yishay Mansour

Scribe: Itay Kirshenbaum, Mor Sela

8.1 External Regret - Reminder

Let us recall what we know about *external regret*. We are talking about the case of a single player, playing against an adversary environment.

The model (single player):

- Actions $A = \{a_1 \dots a_m\}$
- Time dependent loss function, (can also be defined as a gain function):
 l_i^t - loss of action a_i at time t

The game:

- a. For each step t , the player chooses a distribution $p^t \in \Delta(A)$
- b. The adversary decides on the losses $l^t = (l_1^t \dots l_m^t)$, where $l_i^t \in [0, 1]$
- c. The player's loss at time t is $l_{ON}^t = \sum_{i=1}^m p_i^t l_i^t$.
 The cumulative loss up to time T is

$$L_{ON}^T = \sum_{t=1}^T l_{ON}^t$$

The goal: Bring the cumulative loss to a minimum

We defined: $L_i^T = \sum_{t=1}^T l_i^t$ as the loss when playing a static action a_i .

External regret is defined as a measure of optimality (comparing our performance to the performance of the single best action):

$$ER = L_{ON}^T - \min_i L_i^T$$

Last week we showed an algorithm such that:

$$L_{ON}^T \leq L_{best}^T + 2\sqrt{T \ln m}$$

Today we show algorithms that have "stronger" guarantees than external regret.

8.2 Partial Information Model

In this model the player chooses at time t a **single action** a_i^t based on some distribution p^t . The opponent then sets the costs l^t based on p^t . The player observes only the cost of the action he selected.

8.2.1 Reduction from Full Information

We will divide our time into T/k blocks of size k , denoted by $B_1 \dots B_{T/k}$.

k	k	k	k	k	k	k
$ l^1$	l^k l^1	l^k l^1	l^k l^1	l^k l^1	l^k l^1	l^k l^1
B_1	\dots	\dots	\dots	\dots	\dots	$B_{T/k}$

During block B_t : We sample the m actions in m different times. At the rest of the time we use some distribution p^t .

At the end of block B_t : we gather the loss of the m sampled actions $x_1^t \dots x_m^t$ and give it to a full information algorithm ER . The algorithm returns a distribution p^{t+1} , which we use in block B_{t+1} during the non-sampling steps.

The ER algorithm will give us for any action $j \in A$:

$$\sum_{t=1}^{T/k} \sum_{i=1}^m p_i^t \cdot x_i^t \leq \sum_{t=1}^{T/k} x_j^t + R$$

where $R \cong \sqrt{\frac{T}{k} \log N}$

Now we compute the expected value of x_i^t :

$$E[x_i^t] = \frac{1}{k} \sum_{\tau \in B^t} l_i^\tau$$

And therefore we have:

$$\begin{aligned} E\left[\sum_{t=1}^{T/k} \sum_{i=1}^m p_i^t \cdot x_i^t\right] &\leq E\left[\sum_{t=1}^{T/k} x_j^t\right] + R \\ &\Downarrow \\ E\left[\sum_{t=1}^{T/k} \sum_{i=1}^m p_i^t \cdot x_i^t\right] &= \sum_{\tau=1}^{T/k} \frac{1}{k} \sum_{t \in B_\tau} l^t \cdot E[p^\tau(x^1 \dots x^{t-1})] \leq \frac{1}{k} \sum_{t=1}^T l_j^t + R \\ &\Downarrow \end{aligned}$$

$$\frac{1}{k}E[L_{ON}^T] \leq \frac{1}{k}L_j^T + R$$

Adding the losses from the sampling stage we get:

$$E[L_{ON}^T] \leq L_j^T + kR + \frac{1}{k}m$$

We can optimize this result over k and have:

$$k \cong T^{\frac{1}{3}}N^{\frac{2}{3}}$$

and derive:

$$Regret \sim T^{\frac{2}{3}}N^{\frac{2}{3}}$$

8.3 Correlated Equilibrium

The model:

- N players - $\{1 \dots n\}$
- Actions of player i - A_i
- Joint action - $A = A_1 \times \dots \times A_n$
- Utility function of player i - $u_i : A \rightarrow \mathbf{R}$
We will assume that the utility range is $[0, 1]$

8.3.1 Internal Regret

In this section we define *pure* and *correlated equilibria* in a different manner than before. We use a new measure, called *Internal Regret*:

$$IR_i(a, x, y) = \begin{cases} u_i(a^{-i}, y) - u_i(a) & a_1 = x \\ 0 & a_1 \neq x \end{cases}$$

The meaning is measuring the loss of player i caused by playing action x instead of action y .

Definition $a \in A$ is a *pure equilibrium* if:

$$\forall i \in N, \forall x, y \in A_i : IR_i(a, x, y) \leq 0$$

Definition Let $Q \in \Delta(A)$ be a distribution over the joint actions. We adapt the *regret* definition in the following manner:

$$IR_i(Q, x, y) = E_{a \sim Q}[IR_i(a, x, y)]$$

Using this definition, Q is a **correlated equilibrium** if:

$$\forall i \in N, \forall x, y \in A_i : IR_i(Q, x, y) \leq 0$$

The existence of a correlated equilibrium can be derivated from the existence of *Nash Equilibrium* (There is also a direct proof using zero sum games).

8.4 ϵ -Correlated Equilibrium

The demand for correlation can be relaxed, by demanding only that a player cannot gain more than ϵ , as a result of the action change.

Definition For each player $i \in N$ we'll define **deviation functions**:

$$F_i = \{f : A_i \rightarrow A_i\}$$

Definition Q is ϵ -**Correlated Equilibrium** if:

$$\forall i \in N, \forall f \in F_i : E_{a \sim Q}[u_i(a)] \geq E_{a \sim Q}[u_i(a^{-i}, f(a_i))] - \epsilon$$

8.4.1 Swap Regret

We define the *Swap Regret* to be:

$$SR(Q, i, f) = \sum_{a_i \in A_i} Pr[a_i \sim Q] \cdot IR_i(Q, a_i, f(a_i))$$

and in general:

$$SR(Q) = \max_{i \in N} \max_{f \in F_i} SR(Q, i, f)$$

Thus, Q is an ϵ -**correlated equilibrium** if:

$$SR(Q) \leq \epsilon$$

We can extend this definition for a **sequence of actions**: For a sequence of joint actions $\vec{a} = a^1 \cdots a^T$, we define in a similar manner:

$$SR(\vec{a}) = \max_{i \in N} \max_{f \in F_i} \sum_{t=1}^T IR_i(a^t, a_i^t, f(a_i^t))$$

Claim 8.1 *If for a sequence of joint actions \vec{a} , $SR(\vec{a}) \leq \epsilon \cdot T$, then the distribution Q is an ϵ -correlated equilibrium, where:*

$$Q(z) = \begin{cases} \frac{1}{T} & z = a^t \\ 0 & \text{otherwise} \end{cases}$$

8.4.2 Reduction of External Regret to Swap Regret

In order to achieve equilibrium, we need an algorithm which minimizes swap regret. In this section, we will use algorithms that use the external regret measure, as a way to create an algorithm for swap regret. For that, we will construct a reduction: external regret \mapsto swap regret.

Let's look at a single player who is only aware of his own losses (note that we switched to loss terminology instead of utility in order to be consistent with ER). Assume the number of actions of the single player is $|A_i| = m$. We will use m (possibly different) External Regret algorithms $B_1 \dots B_m$ as shown in figure 8.1.

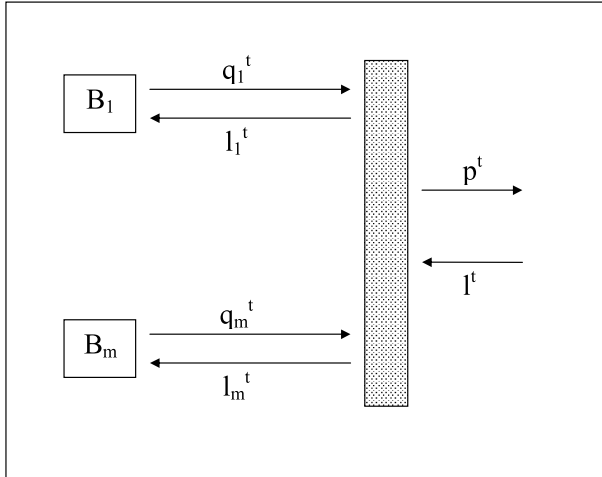


Figure 8.1: Reduction of External Regret to Swap Regret algorithm

Recall the *ER assumption*:

$$\text{For any sequence of } t \text{ losses } \{l_j^t\}, L_{ON}^T = \sum_{t=1}^T l_{ON}^t \leq L_j^T + R = \sum_{t=1}^T l_j^t + R$$

This implies that

$$\forall j \in A_i, L_{ON}^T \leq L_j^T + R$$

We will construct our algorithm using m algorithms, each guaranteeing at the most regret R . Intuitively, each will be responsible of a single action (there are m algorithms - one for each possible action of the player). Each algorithm outputs a vector of what it would like to play, and we need to return to each separate algorithm its loss. We need to wrap up these algorithms in some sort of interface which will calculate the distribution and return the loss. Thus we have two important actions to do:

- a. Calculate p^t from $\vec{q}_1^t, \dots, \vec{q}_m^t$
- b. "Distribute" \vec{l}^t - return to B_i a loss vector \vec{l}_i^t .

Let's start with distributing the loss: we simply return to B_i a loss vector $\vec{l}_i^t = p_i^t \cdot \vec{l}^t$. We need to define now how we combine the separate "recommendations" q_i to get the distribution p . We construct a stochastic matrix:

$$M_{m \times m} = \begin{pmatrix} q_{11} & \cdots & q_{1m} \\ \vdots & & \vdots \\ q_{m1} & \cdots & q_{mm} \end{pmatrix} \begin{array}{l} \leftarrow \vec{q}_1^t \\ \\ \leftarrow \vec{q}_m^t \end{array}$$

We choose p such that $pM = p$.

Intuition: p is the output of our algorithm - its meaning is the distribution over actions. We can choose an action in 2 ways:

- choose an action $a_i \in A$ directly from among all possible actions.
This is the output of the algorithm.
- choose an algorithm B_i first (according to p) and then select an action $a_i \in A$ (according to \vec{q}_i^t).

In defining p as above, we ensure that both ways are indeed equivalent. It's possible to prove in several different ways that a solution exists, $\sum p_i = 1$ and $p_i \geq 0$.

Analysis: The loss that B_i "sees" is:

$$(p_i^t \cdot \vec{l}^t) \vec{q}_i^t = p_i^t (\vec{q}_i^t \cdot \vec{l}^t)$$

As B_i is an algorithm for ER, for each B_i and for each action j we have a bound on the regret (assuming all algorithms guarantee the same regret - R):

$$\sum_{t=1}^T p_i^t (\vec{q}_i^t \cdot \vec{l}^t) \leq \sum_{t=1}^T p_i^t \cdot l_{f(j)}^t + R$$

When we sum up the losses, we get that for any point in time:

$$\sum_{i=1}^m (p_i^t \cdot \vec{q}_i^t) \cdot \vec{l}^t = \sum_{i=1}^m p_i^t (\vec{q}_i^t \cdot \vec{l}^t) = \vec{p}^t \cdot M \cdot \vec{l}^t = \vec{p}^t \cdot \vec{l}^t = l_{ON}^t$$

Therefore, the loss of *ONLINE* is the sum of all the B_i 's losses.

When we sum over time we get that:

$$L_{ON}^T = \sum_{i=1}^m L_{B_i}^T \leq \sum_{i=1}^m L_{B_i, f(i)}^T + R = L_{ON, f} + R$$

where

$$L_{ON, f}^T = \sum_{i=1}^m \sum_{t=1}^T p_i^t \cdot l_{f(i)}^t$$

and

$$L_{B_i, f(i)}^T = \sum_{t=1}^T p_i^t \cdot l_{f(i)}^t$$

Recall that last week we proved that: $R \sim \sqrt{T \log m}$ so by summing over all R_i we have that:

$$SR \leq m \sqrt{T \log m}$$

This bound can be easily improved to $SR \leq \sqrt{mT \log m}$

Lower Bound of $\Omega(\sqrt{mT})$ for SR

Let $m = 2k$ and $T = \alpha k$ for $\alpha \geq 3$. The goal is to lower bound the swap regret by $\Omega(k\sqrt{\alpha})$. Let p_i^t be the algorithm's probability of choosing action i in trial t , and let $M_i^t = \sum_{\tau=1}^t p_i^\tau$ be the expected number of trials up to time t when action i is chosen. The adversary assigns losses l_i^t for pairs of action $(2j-1, 2j)$, $j = 1, \dots, k$, as follows:

- if $M_{2j-1}^t + M_{2j}^t < \alpha/2$, then (l_{2j-1}^t, l_{2j}^t) is $(0, 1)$ or $(1, 0)$ with probability $1/2$ each.
- if $M_{2j-1}^t + M_{2j}^t \geq \alpha/2$, then $l_{2j-1}^t = l_{2j}^t = 1$

Since $\sum_{i=1}^m M_i^t = T$, the expected loss of any algorithm is at least $k\frac{\alpha}{2}\frac{1}{2} + k\frac{\alpha}{2}1 = \frac{3}{4}T$.

If there is a pair with $M_{2j-1}^t + M_{2j}^t < \alpha/2$, then either action $2j-1$ or action $2j$ has a loss of at most $T/2$. This implies the lower bound ($\frac{3}{4}T - \frac{T}{2} = \frac{T}{4} \leq \text{Regret}$).

For any pair with $M_{2j-1}^t + M_{2j}^t \geq \alpha/2$, swapping the actions of the pair appropriately given an expected improvement of $\Omega(\sqrt{\alpha})$. With k such pairs, the lower bounds $\Omega(k\sqrt{\alpha})$ follows.

8.4.3 Swap Regret Applications

8.4.4 Dominated Actions

Definition Action $a_{i,1}$ is dominated by action $a_{i,2}$ if for every a_{-i} we have

$$u_i(a_{-i}, a_{i,1}) \leq u_i(a_{-i}, a_{i,2})$$

Likewise, Action $a_{i,1}$ is ϵ -dominated by action $a_{i,2}$ if for every a_{-i} we have

$$u_i(a_{-i}, a_{i,1}) + \epsilon \leq u_i(a_{-i}, a_{i,2})$$

Clearly, we would like to avoid dominated actions. How can we guarantee it?

Theorem 8.2 *If $SR \leq R$ then in at most $\frac{R}{\epsilon}$ steps we play ϵ -dominated actions*

Proof: For every ϵ -dominated action $a_{i,1}$, there is an action $a_{i,2}$ such that by playing $a_{i,2}$ instead of $a_{i,1}$, we would gain at least ϵ .

Therefore, if we have k steps where we play ϵ -dominated actions, we would gain at least ϵk . Since $SR \leq R$, we have $\epsilon k \leq R$ - meaning no more than $\frac{R}{\epsilon}$ steps are played with ϵ -dominated actions □

8.4.5 Calibration

Each day a player gives a "forecast" q (e.g, the chance of rain for tomorrow). The quality of such a forecast may be assessed in a *Calibration test* - In a long sequence of probabilities p , the average p is expected to be close to q .

Stochastic Model: Given a constant probability q for the event to occur, then it is enough to take the average. With high probability the average until time t is close to q and the average from time t till the end is also close to q .

Adversary Model: the player is limited to forecasts of the form $\frac{i}{m}$, $0 \leq i \leq m$. After the player forecasts p^t , it observes $y^t \in \{0, 1\}$. A player is ϵ -calibrated if for all $\frac{i}{m}$:

$$\left| \rho\left(\frac{i}{m}\right) - \frac{i}{m} \right| \leq \epsilon$$

where (assuming that $\sum_{t=1}^T p_i^t = \Omega(T)$):

$$\rho\left(\frac{i}{m}\right) = \frac{\sum_{t=1}^T p_i^t y^t}{\sum_{t=1}^T p_i^t}$$

The algorithm is constructed by performing a reduction from SR . To this end, define a loss function by:

$$l(i, y^t) = \left(y^t - \frac{i}{m}\right)^2$$

And a general loss function:

$$C^T = \sum_{i=1}^m \left(\rho\left(\frac{i}{m}\right) - \frac{i}{m}\right)^2 \left(\sum_{t=1}^T p_i^t\right)$$

The following claim bounds the loss function as a function of the regret:

Lemma 8.3 $C^T \leq SR + \frac{T}{m^2}$

Proof: Using the SR, we get:

$$\begin{aligned} \sum_{t=1}^T p_i^t IR(i, j) &= \sum_{t=1}^T p_i^t [l(i, y^t) - l(j, y^t)] \\ &= \sum_{t=1}^T p_i^t \left[\left(y^t - \frac{i}{m}\right)^2 - \left(y^t - \frac{j}{m}\right)^2 \right] \\ &= \sum_{t=1}^T p_i^t \left[\frac{2(j-i)}{m} - \left(y^t - \frac{i+j}{m}\right) \right] \\ &= \frac{2(j-i)}{m} \left(\sum_{t=1}^T p_i^t \right) \left(\rho\left(\frac{i}{m}\right) - \frac{i+j}{m} \right) \\ &= \left(\sum_{t=1}^T p_i^t \right) \left[\left(\rho\left(\frac{i}{m}\right) - \frac{i}{m} \right)^2 - \left(\rho\left(\frac{i}{m}\right) - \frac{j}{m} \right)^2 \right] \end{aligned}$$

For any fixed $i = 0, 1, \dots, m$ the quantity $\max_j IR(i, j)$ is maximized for the value of j minimizing $\left(\rho\left(\frac{i}{m}\right) - \frac{j}{m}\right)^2$. Moreover, there exists j such that:

$$\left| \rho\left(\frac{i}{m}\right) - \frac{j}{m} \right| \leq \frac{1}{m}$$

Thus,

$$\begin{aligned} \left(\sum_{t=1}^T p_i^t \right) \left(\rho\left(\frac{i}{m}\right) - \frac{i}{m} \right)^2 &= \max_j IR(i, j) + \min_j \left(\sum_{t=1}^T p_i^t \right) \left(\rho\left(\frac{i}{m}\right) - \frac{j}{m} \right)^2 \\ &\leq \max_j IR(i, j) + \frac{T}{m^2} \end{aligned}$$

Summing over all the actions we get:

$$SR = \sum_{i=1}^m \left(\sum_{t=1}^T p_i^t \right) \max_j IR(i, j)$$

which implies:

$$\begin{aligned} C^T &= \sum_{i=1}^m \left(\sum_{t=1}^T p_i^t \right) \left(\rho\left(\frac{i}{m}\right) - \frac{i}{m} \right)^2 \\ &\leq \sum_{i=1}^m \max_j IR(i, j) + \frac{T}{m^2} \\ &\leq SR + \frac{T}{m^2} \end{aligned}$$

□

To complete the proof of calibration, note that if an action $\frac{i}{m}$ has $\sum_{t=1}^T p_i^t \geq \alpha T$ and $\left| \rho\left(\frac{i}{m}\right) - \frac{j}{m} \right| \geq \epsilon$, then:

$$C^T \geq \epsilon^2 \alpha T$$

We showed that:

$$C^T \leq SR + \frac{T}{m^2} \leq \sqrt{mT \log m} + \frac{T}{m^2}$$

This implies that :

$$\alpha \epsilon^2 \leq \sqrt{mT \log m} + \frac{1}{m^2}$$

For $\alpha = \Omega(1)$ and $T = m^3$ we have that:

$$\epsilon = O\left(\frac{1}{m}\right)$$