Advanced Topics in ML and AGT	**Fall Semester, 2018/9**

Lecture 10: December 17

Lecturer: Yishay Mansour *Scribe: Sagie Benaim, Itay Shamir, Daniel Dubnikov*

1. Correlated Equilibrium
2. $\epsilon-$Correlated Equilibrium
3. Swap Regret
4. Dominated Actions
5. Calibration
6. From Calibration To Swap Regret

* [1]

10.1 Correlated Equilibrium

Our model:

- N players denoted as $[N]$.
- A_i is the set of actions for the i'th player.
- $A = A_1 \times \dots A_N$ - combined actions.
- $l_i : A \to \mathbb{R}$ loss function of i'th player.
- We assume $\forall i : l_i(\vec{a}) \in [0, 1]$.
- We denote $a = [a_1, ..., a_N]$.
- We denote $a_{-i} = [a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_N]$, i.e., the strategy profile a without the i'th player action.

[1]Based on scribe notes by Roei Herzig, Barak Gross and Achiya Jerbi from 2017/18

Definition 10.1 (Correlated Equilibrium) *A distribution Q over A is* ***Correlated Equilibrium (or CE)*** *if:*

$$\forall i \in [N]\ \forall b, c \in A_i:\ \mathbb{E}_{a \sim Q}[l_i(a_{-i}, c) - l_i(a) | a_i = b] \geq 0$$

The intuition for the definition is that no player would want to deviate from the recommended strategy.

Definition 10.2 (Switch Function) $switch(a_i, b, c) = \begin{cases} c & a_i = b \\ a_i & a_i \neq b \end{cases}$

We provide another alternative definition for CE:

Definition 10.3 (Correlated Equilibrium) *A distribution Q over A is* ***Correlated Equilibrium (or CE)*** *if:*

$$\forall i \in [N]\ \forall b, c \in A_i:\ \mathbb{E}_{a \sim Q}[l_i(a_{-i}, switch(a_i, b, c)) - l_i(a)] \geq 0$$

We can compare this to a Pure Nash Equilibrium

Definition 10.4 (Pure Nash Equilibrium) *A strategy $a = (a_1, \ldots, a_N)$ over A is* ***Pure Nash Equilibrium (or PNE)*** *if :*

$$\forall i \in [N]\ \forall b, c \in A_i:\ l_i(a_{-i}, switch(a_i, b, c)) - l_i(a) \geq 0$$

Theorem 10.5 *Every game with finite number of players and actions has at least one CE.*

Proof: We build a zero-sum game.
Let the rows player choose a profile strategy from A, with the goal of maximizing the value.
The columns player chooses $switch_i(\cdot, b, c)$, with purpose of minimizing the value.
Define the game matrix $M(a, (i, b, c))$ as follows:

$$M(a, (i, b, c)) = l_i(a_{-i}, switch_i(a_i, b, c)) - l_i(a)$$

We notice that existence of CE is equal to the game having non-negative value.

A strategy for the columns player that guarantees the value is zero is (i, b, b):
$switch(a_i, b, b) = a_i$.

We will show a strategy for the rows player that guarantees the value is zero.
Given the mixed strategy D of the column player over (i, b, c), we define a matrix D_i for

every player $i \in [N]$ in the following way:

$$D_i(b,c) = \Pr[a'_i = c | a_i = b, (i,*,*) \textit{ chosen by columns player }]$$

If the probability of choosing $(i,*,*)$ is zero under D, we choose $D_i(b,c) = \frac{1}{|A_i|}$.
Given D_i, let π_i be its stationary distribution:

$$\pi_i^T D_i = \pi_i^T$$

We define the product distribution such that $\Pi = \pi_1 \times ... \times \pi_N$
By our construction of Π, sampling $a \sim \Pi$ is identical to sampling $a \sim \Pi$ and $(i,b,c) \sim D$ and considering $(a_{-i}, switch(a_i,b,c))$. This implies that the value of the game is zero.
□

10.2 Calculating Correlated Equilibrium

For each $a \in A$ we define a variable $p(a)$ and have constraint

$$\forall i \in [N]\ \forall b,c \in A_i : \sum_{a \in A} p(a)[l_i(a_{-i}, switch(a_i,b,c)) - l_i(a)] \geq 0$$

We have $N|A|^2$ constraints.
Therefore, there exists a (basic) solution where at least $N|A|^2$ of the variables $p(a)$ are not zero.

10.3 ϵ-Correlated Equilibrium

The correlation equilibrium can be relaxed by requiring only that a player cannot gain more than ϵ, as a result of consistently changing its action.

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

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

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

$$\forall i \in N, \forall f \in F_i : \mathbb{E}_{a \sim Q}[l_i(a)] - \mathbb{E}_{a \sim Q}[l_i(a_{-i}, f_i(a_i))] \leq \epsilon$$

10.3.1 Swap Regret

Given sequence of loss functions $l_1, l_2, ..., l_T$ and sequence of actions $a_1, a_2, ..., a_T$ we will define:

$$L_{SR} = \min_{f \in F} \sum_{t=1}^{T} l_t(f(a_t))$$

We will define the *Swap Regret* to be:

$$SR = \sum_{t=1}^{T} l_t(a_t) - L_{SR} = \sum_{t=1}^{T} l_t(a_t) - \min_{f \in F} \sum_{t=1}^{T} l_t(f(a_t))$$

Theorem 10.6 *If we played a game G for T steps, and SR is bound by R(T,N), then the empirical distribution Q is an ϵ-Correlated Equilibrium for $\epsilon = \frac{R(T,N)}{T}$.*

Proof: Assuming by contradiction that the empirical distribution Q is not ϵ-Correlated Equilibrium.
Then exist player i and function $f \in F$ s.t:

$$E_{a \sim Q}[l_i(a)] - E_{a \sim Q}[l_i(a_{-i}, f_i(a_i)] > \epsilon$$
$$\Rightarrow \sum_{a \in A} Q(a) \left[l_i(a) - l_i(a_{-i}, f_i(a_i)) \right] > \epsilon$$
$$\Rightarrow \sum_{t=1}^{T} \frac{1}{T} \left[l_i(a) - l_i(a_{-i}, f_i(a_i)) \right] = \frac{1}{T} SR \leq \frac{1}{T} R(T,N) = \epsilon$$

And this is a contradiction. □

10.4 Swap Regret Applications

10.4.1 Dominated actions

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

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

Likewise, action $a_{i,2}$ is $\epsilon - \textit{dominated}$ by action $a_{i,1}$ if for every a_{-i} we have:

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

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

Theorem 10.7 *Assume we play on algorithm that guarantees that the swap regret $SR \leq R$ then the number of time we play ϵ-dominated action is bound by $\frac{R}{\epsilon}$.*

Proof: For every $\epsilon - dominated$ action $a_{i,2}$, there is an action $a_{i,1}$ that dominated it by at least ϵ. Let us define function f that map $a_{i,2}$ actions to actions that are better in at least ϵ: $f(a_{i,2}) = a_{i,1}$.
If we play k times actions that are $\epsilon - dominated$, then after using f we would gain at least ϵk.
Since $SR \leq R$, then $\epsilon k \leq R$. □

10.4.2 Calibration

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

Stochastic Model: Given a constant probability q for the event to occur, then it is enough to take the average of results until time t:

$$p_t = \frac{1}{t}\sum_{i=1}^{t-1} y_i$$

.
With high probability $p_t \underset{t\to\infty}{\to} q$ and $|p_t - q| \leq \frac{1}{\sqrt{t}}$ (Chernoff bound).

Adversarial Deterministic Model: After discretization of the problem to $m+1$ possible predictions, the player is limited to forecasts of the form $\frac{i}{m}$,$0 \leq i \leq m$.
For all i, we define the following:

$$S(\frac{i}{m}) = \left\{t : p_t = \frac{i}{m}\right\} \quad \text{and} \quad \rho(\frac{i}{m}) = \frac{\sum_{t\in S(\frac{i}{m})} y_t}{\left|S(\frac{i}{m})\right|}$$

After the player forecasts p_t, it observes $y_t \in \{0,1\}$. A player is (α,ϵ)-*calibrated* if for all $i \in [m]$:

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

When:

$$\left|S\left(\frac{i}{m}\right)\right| \geq \alpha T$$

Deterministic Prediction When using the deterministic model the adversary can choose a sequence of y_t that will make any player's strategy uncalibrated. Specifically, consider the case that when $p_t \geq 1/2$ the adversary will chooses $y_t = 0$ and when $p_t < 1/2$ the adversary will chooses $y_t = 1$.
So:

- If $\frac{i}{m} < \frac{1}{2}$ then $p(\frac{i}{m}) = 1$ and $|p(\frac{i}{m}) - \frac{i}{m}| \geq \frac{1}{2}$
- If $\frac{i}{m} \geq \frac{1}{2}$ then $p(\frac{i}{m}) = 0$ and $|p(\frac{i}{m}) - \frac{i}{m}| \geq \frac{1}{2}$.

Therefore for each prediction $\frac{i}{m}$: $|p(\frac{i}{m}) - \frac{i}{m}| \geq \frac{1}{2}$, and since there a prediction that appear at least $\frac{T}{m}$ times, then the deterministic player is uncalibrated. Namely, we get a contradiction for $\epsilon = \frac{1}{2}$ and $\alpha = \frac{1}{m}$

Adversary Randomized Model: Now at each time step t the player chooses a distribution $p_t \in \Delta(\{0, \frac{1}{m}, \frac{2}{m}, ..., 1\})$. The definition for $S(\frac{i}{m})$ and $\rho(\frac{i}{m})$ changes as follows:

$$S(\frac{i}{m}) = \sum_{t=1}^{T} p_i^t \quad \text{and} \quad \rho(\frac{i}{m}) = \frac{\sum_{t \in S(\frac{i}{m})} y_t p_i^t}{S(\frac{i}{m})}$$

Existence:

Lemma 10.8 *There exists an* (α, ϵ)-calibrated *strategy for the forecaster which guarantees to pass the calibration test with probability at least* $1 - \delta$*, where* $\delta = me^{-\epsilon^2 \alpha T}$ *and* $\epsilon \geq 1/m + 1/\sqrt{\alpha T}$

Proof: We will define a zero-sum game, where the rows player will provide the y-sequence (i.e., will decide whether or not it will rain) and the columns player will choose a strategy based on the history:

$$f : history \rightarrow \Delta(\{0, \frac{1}{m}, \frac{2}{m}, ..., 1\})$$

The value of the game for (y, f) is 0 if the predictions are (ϵ, α)-calibrated and 1 otherwise. We'd like to show that the columns player can pass the calibration test with probability $1 - \delta$ for $\delta = e^{-\epsilon^2 \alpha T}$.
We'll fix a mixed strategy for the rows player (distribution over y). We can now define a strategy for the columns player that computes the probability that $y_t = 1$, given the history. Let this probability be p_t, and it is discretized to $\{0, \frac{1}{m}, \frac{2}{m}, ..., 1\}$, to be $\hat{p}_t$.
Consider all the times that the column player played $\hat{p}_t = i/m$. If the number of times it plays $\hat{p}_t$ is less then αT, this action is calibrated by definition. If it is more than αT, then with probability $1 - \delta$ the difference between the realization and $\hat{p}_t$ is at most $1/m + 1/\sqrt{t} < 1/m + 1/\sqrt{\alpha T} < \epsilon$. This implies that the value of the game is at most $m\delta$. This implies that

the forecaster has a mixed strategy that guarantees that the probability of an outcome not being calibrated is at most $m\delta$.

□

Reduction to Swap Regret: We now introduce the algorithm which is constructed by performing a reduction to SR. To this end, define a quadratic loss function by:

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

And a general cost function:

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

Note that $\sum_{t=1}^{T} p_i^t$ is the number of times i/m is played.

Lemma 10.9 *The action that minimizes the loss during the fractional times when we predict i/m is $\rho(\frac{i}{m})$*

Proof: $q^* = \arg\min_q \sum_t (y_t - q)^2 p_t(i) = \frac{\sum_{t=1}^{T} p_t(i) y_t}{\sum_{t=1}^{T} p_t(i)} = \rho(\frac{i}{m})$ □

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

Lemma 10.10 $C \leq SR + \frac{T}{m^2}$

Proof: By discretization, at time step T, for every i/m there exists a value j/m that is $\frac{1}{m} - close$ to $\rho(i/m)$, i.e.:

$$\left|\rho(\frac{i}{m}) - \frac{j}{m}\right| \leq \frac{1}{m} \tag{10.1}$$

We denote by $IR[i/m \to j/m]$ the change in loss when we replace the predictions i/m by j/m, this is also called internal regret. We now consider the regret of replacing a single probability $p_t(\frac{i}{m})$ with $p_t(\frac{j}{m})$ for all $t \in [T]$.

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

Now, if we sum the internal regret for all $i \in [m]$ we get the following:

$$\begin{aligned}
SR &= \sum_{i=1}^{m}\left(\sum_{t=1}^{T} p_t(i/m)\right)\left[\left(\rho(\frac{i}{m}) - \frac{i}{m}\right)^2 - \left(\rho(\frac{i}{m}) - \frac{j}{m}\right)^2\right] \\
&\geq \sum_{i=1}^{m}\left(\sum_{t=1}^{T} p_t(i/m)\right)\left(\left(\rho(\frac{i}{m}) - \frac{i}{m}\right)^2 - \frac{1}{m^2}\right) \\
&= \sum_{i=1}^{m}\left(\sum_{t=1}^{T} p_t(i/m)\right)\left(\rho(\frac{i}{m}) - \frac{i}{m}\right)^2 - \frac{T}{m^2} \\
&= C - \frac{T}{m^2}
\end{aligned}$$

where the inequality is because of the bound at inequality (10.1). □

Now assume by contradiction that the series is not $(\alpha,\epsilon)-$*calibrated*. Therefore, there exists an action i s.t $\sum_{t=1}^{T} p_t(i/m) \geq \alpha T$ and $\left|\rho(\frac{i}{m}) - \frac{j}{m}\right| \geq \epsilon$. Then:

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

We will show next week that:

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

This implies that:

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

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

Let $T > m^5 \log m$ and we have $\alpha\epsilon^2 \leq 2/m^2$. However, for any $\alpha, \epsilon > 0$ for $m \geq 2/\sqrt{\alpha}\epsilon$ we get a contradiction. Namely, if we select a small enough discretization using the parameter m, we are guaranteed to have (α, ϵ)-calibrated sequence of predictions.