Online Learning with Expert Advice

(Sequential-decision making framework).

Multi-round game.

Decision maker

n option available "experts".

$\underline{\ell}_t \in [0,1]^n$

Environment

$\underline{x}_t$ is a probability distribution over the n options

- At time time t, the loss incurred by the decision-maker is $\underline{x}_t \cdot \underline{\ell}_t$
- No assumption on the loss other than the fact that they are bounded.
- The loss vector $\underline{\ell}_t$ is only revealed after the choice $\underline{x}_t$ is made.
- Play this game for T rounds.

$$\text{Loss}(DM) = \sum_{t=1}^{T} \underline{x}_t \cdot \underline{\ell}_t$$

- "Best fixed strategy in hindsight" $\quad \Delta_n = \{ \underline{x} \in \mathbb{R}^n \mid \underline{x} \geq 0 \ \& \ \underline{x} \cdot \underline{1} = 1 \}$.

$$\underline{x}^* \in \underset{\underline{x} \in \Delta_n}{\text{argmin}} \sum_{t=1}^{T} \underline{x} \cdot \underline{\ell}_t = \underset{\underline{x} \in \Delta_n}{\text{argmin}} \ \underline{x} \cdot \underbrace{\sum_{t=1}^{T} \underline{x} \cdot \underline{\ell}_t}_{L_{T+1}}$$

- $$\text{Regret} = \text{Loss}(DM) - \min_{\underline{x} \in \Delta_n} \sum_{t=1}^{T} \underline{x} \cdot \underline{\ell}_t$$

<u>Goal</u>: Minimise Regret.

→ Regret grows sublinearly with T, Regret $= O(\sqrt{\log n \, T})$.

→ Regret can be negative

(Assume that the decision maker knows the time horizon T.)

(Take I): "Follow the leader".

At time t, pick $\underline{x}_t \in \arg\min_{\underline{x}} \sum_{s=1}^{t-1} \underline{x} \cdot \underline{l}_s = \arg\min_{\underline{x}} \underline{x} \cdot \underbrace{\left(\sum_{s=1}^{t-1} l_s\right)}_{L_t}$.

pick i s.t. $L_{t,i}$ is minimum, set $x_{t,i} = 1$, $x_{t,j} = 0$ $\forall j \neq i$.

<u>Only have 2 options:</u>

	l_1	l_2	—	—		...	l_T
1	$1/2$	0	1	0	1	...	
2	0	1	0	1	0		

Loss (DM) $= T$, Loss (Best in hindsight) $= T/2$, Regret $= T/2$.

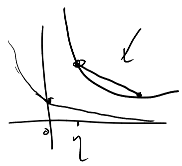
<u>Algorithm</u>: (Exponentially weighted Forecaster / Multiplicative Weight Update / Hedge ...)

- $\underline{w}_1 = (1, \dots, 1) \in \mathbb{R}^n$
- For $t = 1, \dots, T$
  - $\underline{x}_t = \frac{1}{Z_t} \underline{w}_t$, $Z_t = \sum_i w_{t,i}$
  - $w_{t+1,i} = w_{t,i} \cdot \exp(-\eta\, l_{t,i})$.

<u>Theorem</u>: Regret of MWUA with $\eta = \sqrt{\frac{2 \log n}{T}}$ is $2\sqrt{(\log n) T}$.

<u>Proof</u>:

$$\frac{Z_{t+1}}{Z_t} = \frac{\sum_i w_{t+1,i}}{Z_t} = \frac{\sum_i w_{t,i} \exp(-\eta\, l_{t,i})}{Z_t}$$



$$\leq \sum_i \frac{w_{t,i}}{Z_t} \left(1 + (e^{-\eta} - 1) l_{t,i}\right).$$

$$= 1 + (e^{-\eta} - 1) \underbrace{\sum_i x_{t,i} \cdot l_{t,i}}_{\text{loss of decision maker at time } t}$$

"Potential function"

$$\frac{Z_{t+1}}{Z_t} \leq \exp\left(\underline{x}_t \cdot \underline{l}_t (e^{-\eta} - 1)\right). \qquad (A)$$

$$\frac{Z_{T+1}}{Z_1} = \prod_{t=1}^{T} \frac{Z_{t+1}}{Z_t} \leq \exp\left((e^{-\eta} - 1) \sum_{t=1}^{T} \underline{x}_t \cdot \underline{l}_t\right) \qquad (\underline{A}).$$

$$\forall i \quad \frac{Z_{T+1}}{Z_1} \geq \frac{w_{T+1,i}}{n} = \frac{\exp\left(-\eta \sum_{t=1}^{T} l_{t,i}\right)}{n} \qquad (\underline{B})$$

Let i^* be s.t. $x^*_{i^*} = 1$, $x^*_j = 0$ for $j \neq i^*$ is in $\underset{x}{\operatorname{argmin}}\ x \cdot \sum_{t=1}^{T} \ell_t$.

$$\frac{\exp\left(-\eta \sum_{t=1}^{T} \ell_{t,i^*}\right)}{n} \leq \exp\left((e^{-\eta}-1) \sum_{t=1}^{T} x_t \cdot \ell_t\right)$$

Combining (A) & (B).

$$-\eta \sum_{t=1}^{T} \ell_{t,i^*} - \log n \leq (e^{-\eta}-1) \sum_{t=1}^{T} x_t \cdot \ell_t$$

$$\eta \Big(\underbrace{\sum_{t=1}^{T} x_t \cdot \ell_t}_{\text{Loss(DM)}} - \underbrace{\sum_{t=1}^{T} \ell_{t,i^*}}_{\text{Loss(Best in Hindsight)}} \Big) \leq \left(e^{-\eta} - (1-\eta)\right) \sum_{t=1}^{T} x_t \cdot \ell_t + \log n$$

$$\text{Regret} \leq \left(\frac{e^{-\eta} - (1-\eta)}{\eta} \right) \sum_{t=1}^{T} x_t \cdot \ell_t + \frac{\log n}{\eta}$$

$$\boxed{e^{-\eta} \leq 1 - \eta + \eta^2} \qquad \leq \eta T + \frac{\log n}{\eta} \leq 2\sqrt{T \log n} \quad \text{if } \eta = \sqrt{\frac{\log n}{T}}.$$

∎

Theorem: There exists a sequence of losses such that any algorithm suffers a regret of $\Omega(\sqrt{T})$. (or $\Omega(\sqrt{T \log n})$).

Proof ("sketch").

At time t, toss a coin. if "HEADS" $\ell_t = (0,1)$
else $\ell_t = (1,0)$

$\mathbb{E}[\text{loss of DM}] = T/2$ $\qquad$ $\mathbb{E}[\text{loss at time } t] = \frac{x_{t,1}}{2} + \frac{x_{t,2}}{2}$.

$\mathbb{E}[\text{loss of best in hindsight}] = \frac{T}{2} - c\sqrt{T}$

$\mathbb{E}[\text{Regret}] = c\sqrt{T}$.

(Can prove that with some constant prob. Regret $\geq c\sqrt{T}$ for some constant c.)