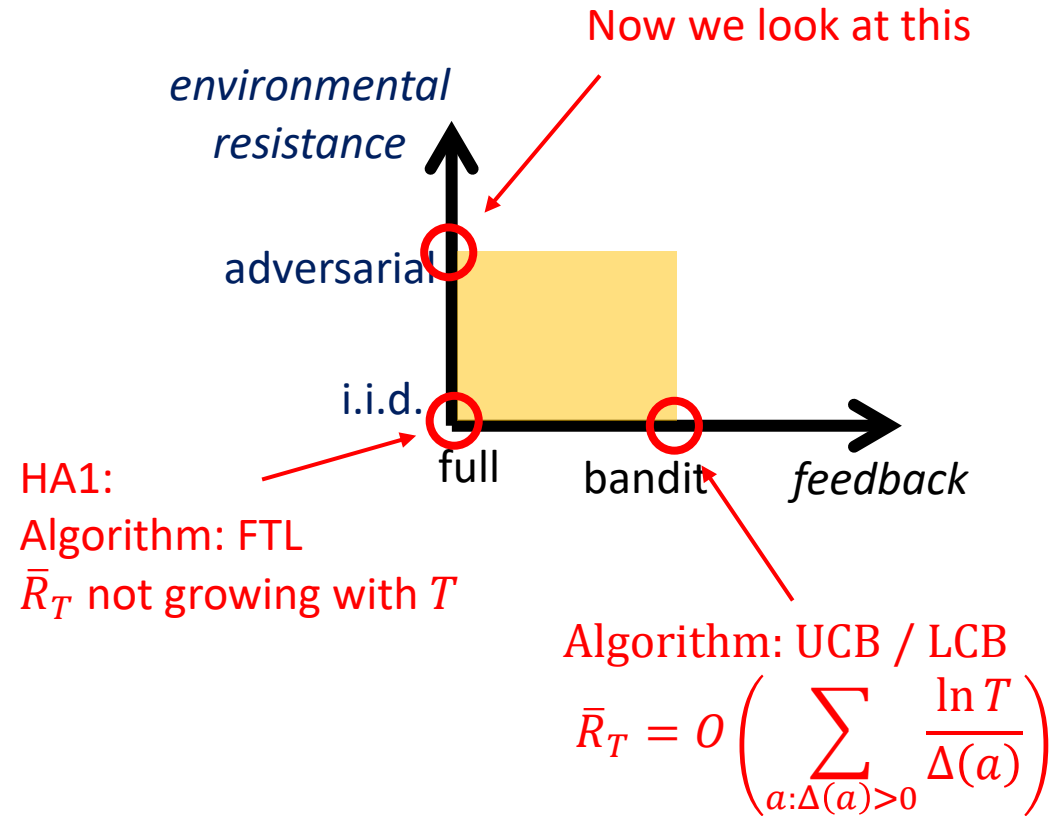


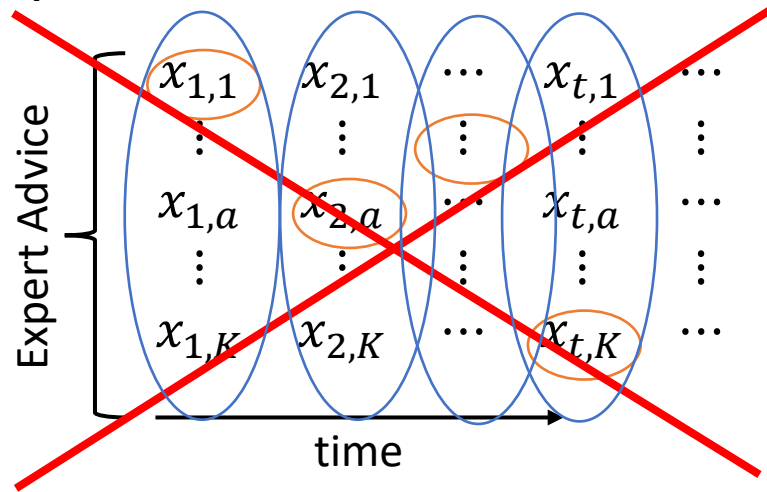
Prediction with Expert Advice and Adversarial Bandits

Yevgeny Seldin

So far



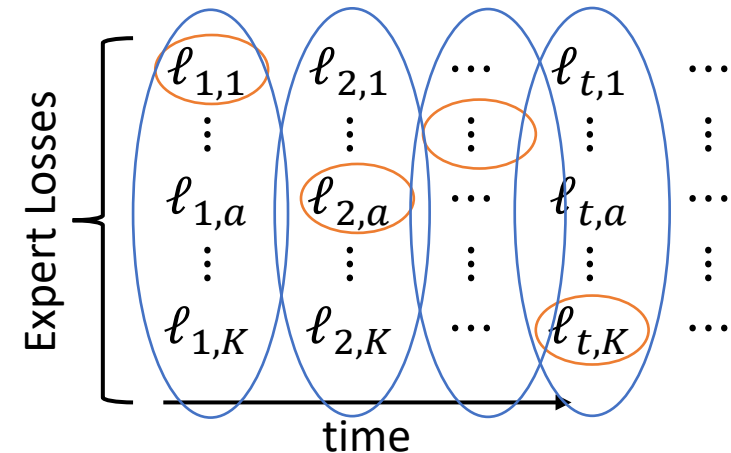
Prediction with Expert Advice (Adversarial full info game)



- Performance measures

- Regret:

$$R_T = \sum_{t=1}^T \ell_{t,A_t} - \min_a \sum_{t=1}^T \ell_{t,a}$$



- Expected regret (oblivious setting):

$$\mathbb{E}[R_T] = \mathbb{E} \left[\sum_{t=1}^T \ell_{t,A_t} \right] - \min_a \sum_{t=1}^T \ell_{t,a}$$

Algorithm for adversarial full info: Hedge / Exponential weights

- $\forall a: L_0(a) = 0$
- For $t = 1, 2, \dots$
 - $\forall a: p_t(a) = \frac{e^{-\eta_t L_{t-1}(a)}}{\sum_{a'} e^{-\eta_t L_{t-1}(a')}}$
 - $A_t \sim p_t$
 - [Observe $\ell_{t,1}, \dots, \ell_{t,K}$]
 - $\forall a: L_t(a) = L_{t-1}(a) + \ell_{t,a}$

- p_t satisfies:

$$p_t = \arg \min_p \left(\langle p, L_{t-1} \rangle + \underbrace{\frac{1}{\eta_t} \sum_a p_a \ln p_a}_{\text{Regularization}} \right)$$

- In FTL: $p_t = \arg \min_p \langle p, L_{t-1} \rangle$

$$p_t(a) = \frac{e^{-\eta L_{t-1}(a)}}{\sum_{a'} e^{-\eta L_{t-1}(a')}}$$

Analysis (for a fixed η)

- Lemma: For any sequence of non-negative $\ell_{t,a}$ and $p_t(a)$ as in Hedge

$$\underbrace{\sum_{t=1}^T \underbrace{\sum_{a=1}^K p_t(a) \ell_{t,a}}_{\text{The expected loss of Hedge at round } t}}_{\text{The expected loss of Hedge}} - \underbrace{\min_a L_T(a)}_{\text{The best loss in hindsight}} \leq \frac{\ln K}{\eta} + \underbrace{\frac{\eta}{2} \sum_{t=1}^T \underbrace{\sum_{a=1}^K p_t(a) (\ell_{t,a})^2}_{\leq 1}}_{\leq 1} \leq T$$

The expected regret of Hedge $\mathbb{E}[R_T]$

- Corollary: $\mathbb{E}[R_T] \leq \frac{\ln K}{\eta} + \frac{\eta}{2} T$
- Take $\eta = \sqrt{\frac{2 \ln K}{T}}$, then $\mathbb{E}[R_T] \leq \sqrt{2T \ln K}$

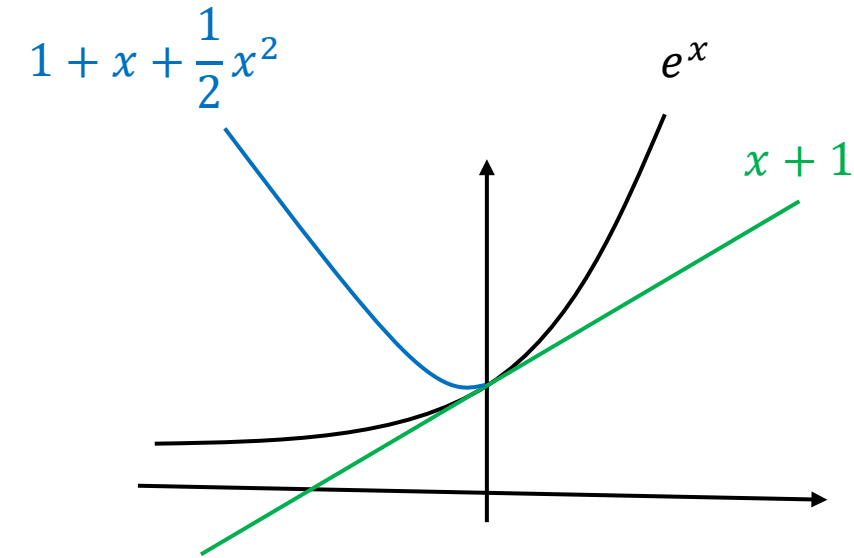
Proof of the lemma

- Define $W_t = \sum_a e^{-\eta L_t(a)}$

$$\begin{aligned}
 \frac{W_t}{W_{t-1}} &= \frac{\sum_a e^{-\eta L_t(a)}}{\sum_{a'} e^{-\eta L_{t-1}(a')}} \\
 &= \sum_a e^{-\eta \ell_{t,a}} \underbrace{\frac{e^{-\eta L_{t-1}(a)}}{\sum_{a'} e^{-\eta L_{t-1}(a')}}}_{p_t(a)} \\
 &= \sum_a e^{-\eta \ell_{t,a}} p_t(a) \\
 &\leq \sum_a \left(1 - \eta \ell_{t,a} + \frac{1}{2} \eta^2 (\ell_{t,a})^2 \right) p_t(a) \\
 &= 1 - \eta \sum_a \ell_{t,a} p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_{t,a})^2 p_t(a) \\
 &\leq e^{-\eta \sum_a \ell_{t,a} p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_{t,a})^2 p_t(a)}
 \end{aligned}$$

$$p_t(a) = \frac{e^{-\eta L_{t-1}(a)}}{\sum_{a'} e^{-\eta L_{t-1}(a')}}$$

$$\sum_{t=1}^T \sum_{a=1}^K p_t(a) \ell_{t,a} - \min_a L_T(a) \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \sum_{t=1}^T \sum_{a=1}^K p_t(a) (\ell_{t,a})^2$$



- For $x \leq 0$:

$$e^x \leq 1 + x + \frac{1}{2}x^2$$
- For any x :

$$1 + x \leq e^x$$

Proof continued

$$W_t = \sum_a e^{-\eta L_t(a)}$$
$$\frac{W_t}{W_{t-1}} \leq e^{-\eta \sum_a \ell_{t,a} p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_{t,a})^2 p_t(a)}$$

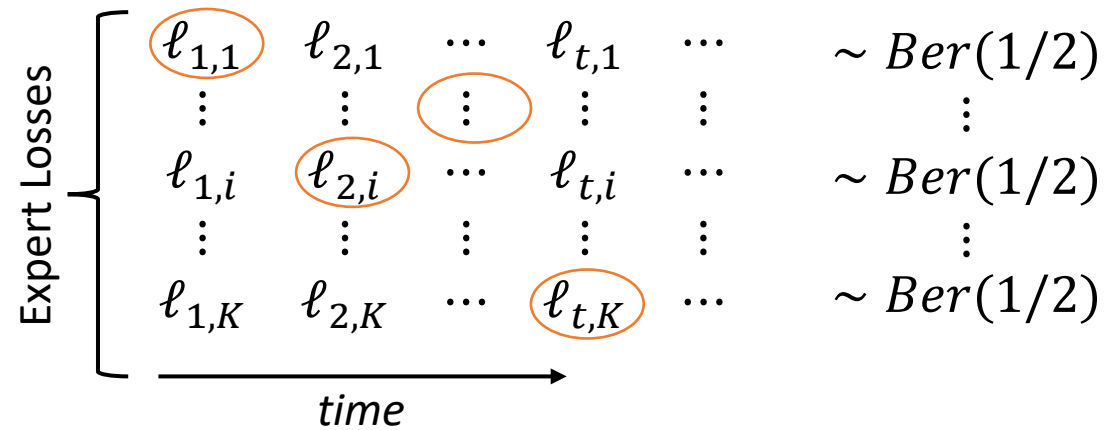
$$\frac{W_T}{W_0} = \frac{W_1}{W_0} \frac{W_2}{W_1} \frac{W_3}{W_2} \cdots \frac{W_T}{W_{T-1}} \leq e^{-\eta \sum_{t=1}^T \sum_a \ell_{t,a} p_t(a) + \frac{\eta^2}{2} \sum_{t=1}^T \sum_a (\ell_{t,a})^2 p_t(a)}$$

$$\frac{W_T}{W_0} = \frac{\sum_a e^{-\eta L_T(a)}}{\sum_a e^{-\eta L_0(a)}} = \frac{\sum_a e^{-\eta L_T(a)}}{K} \geq \frac{\max_a e^{-\eta L_T(a)}}{K} = \frac{e^{-\eta \min_a L_T(a)}}{K}$$

Put the two sides together, take a logarithm and divide by η :

$$\sum_{t=1}^T \sum_{a=1}^K p_t(a) \ell_{t,a} - \min_a L_T(a) \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \sum_{t=1}^T \sum_{a=1}^K p_t(a) (\ell_{t,a})^2$$

Full information lower bound



$$\forall a: \mathbb{E}[L_T(a)] = \frac{T}{2}$$

$$\mathbb{E} \left[\sum_t^T \ell_{t,A_t} \right] = \frac{T}{2}$$

- Lemma

$$\lim_{T \rightarrow \infty} \lim_{K \rightarrow \infty} \frac{\frac{T}{2} - \mathbb{E} \left[\min_a L_T(a) \right]}{\sqrt{\frac{1}{2} T \ln K}} = 1$$

- In the limit of large T and K :

$$\underbrace{\mathbb{E} \left[\sum_t^T \ell_{t,A_t} \right]}_{\mathbb{E}[R_T]} \approx \underbrace{\frac{T}{2} - \mathbb{E} \left[\min_a L_T(a) \right]}_{\text{Complexity of the competitor (Amount of selection)}} \approx \sqrt{\frac{1}{2} T \ln K}$$

Summary

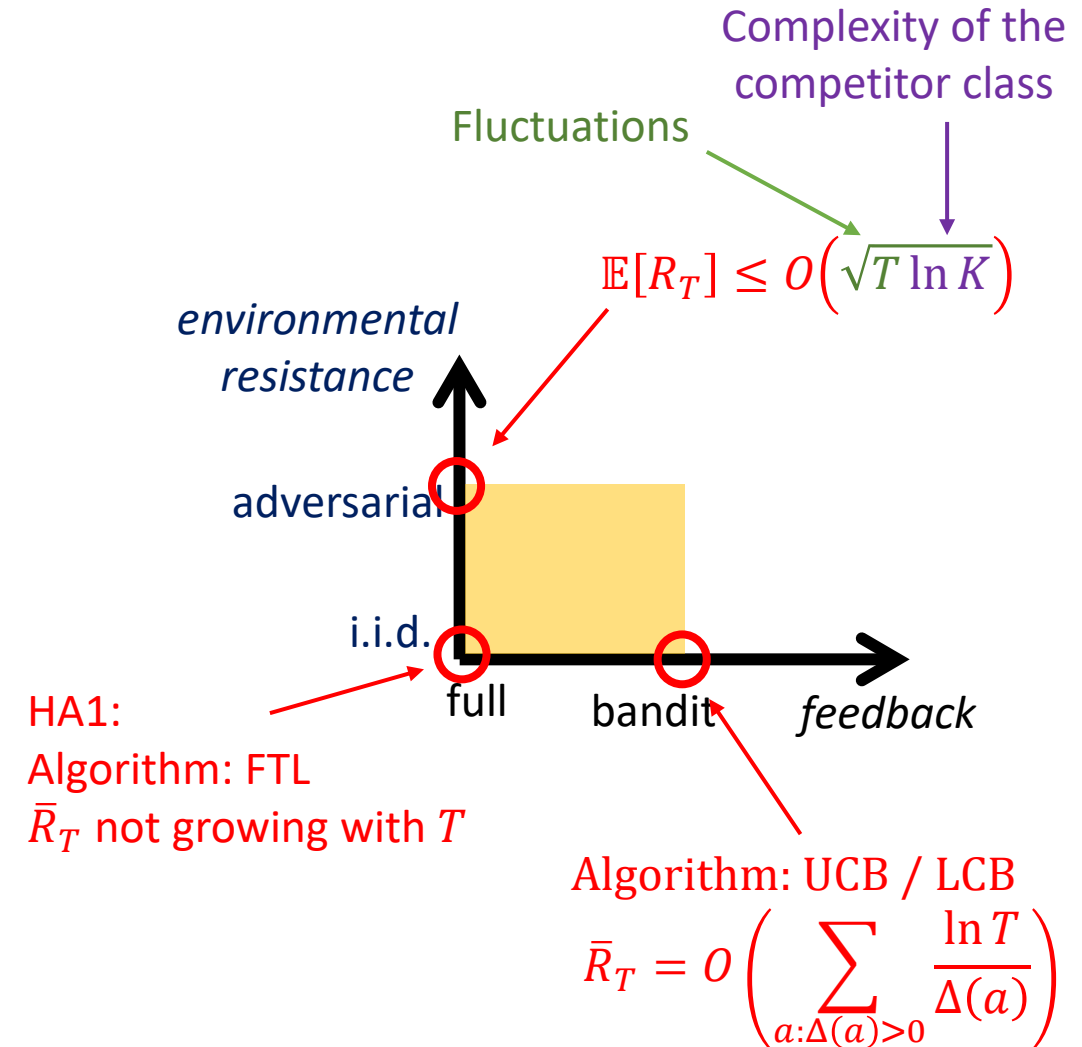
- Hedge:

- $p_t(a) = \frac{e^{-\eta_t L_{t-1}(a)}}{\sum_{a'} e^{-\eta_t L_{t-1}(a')}}$

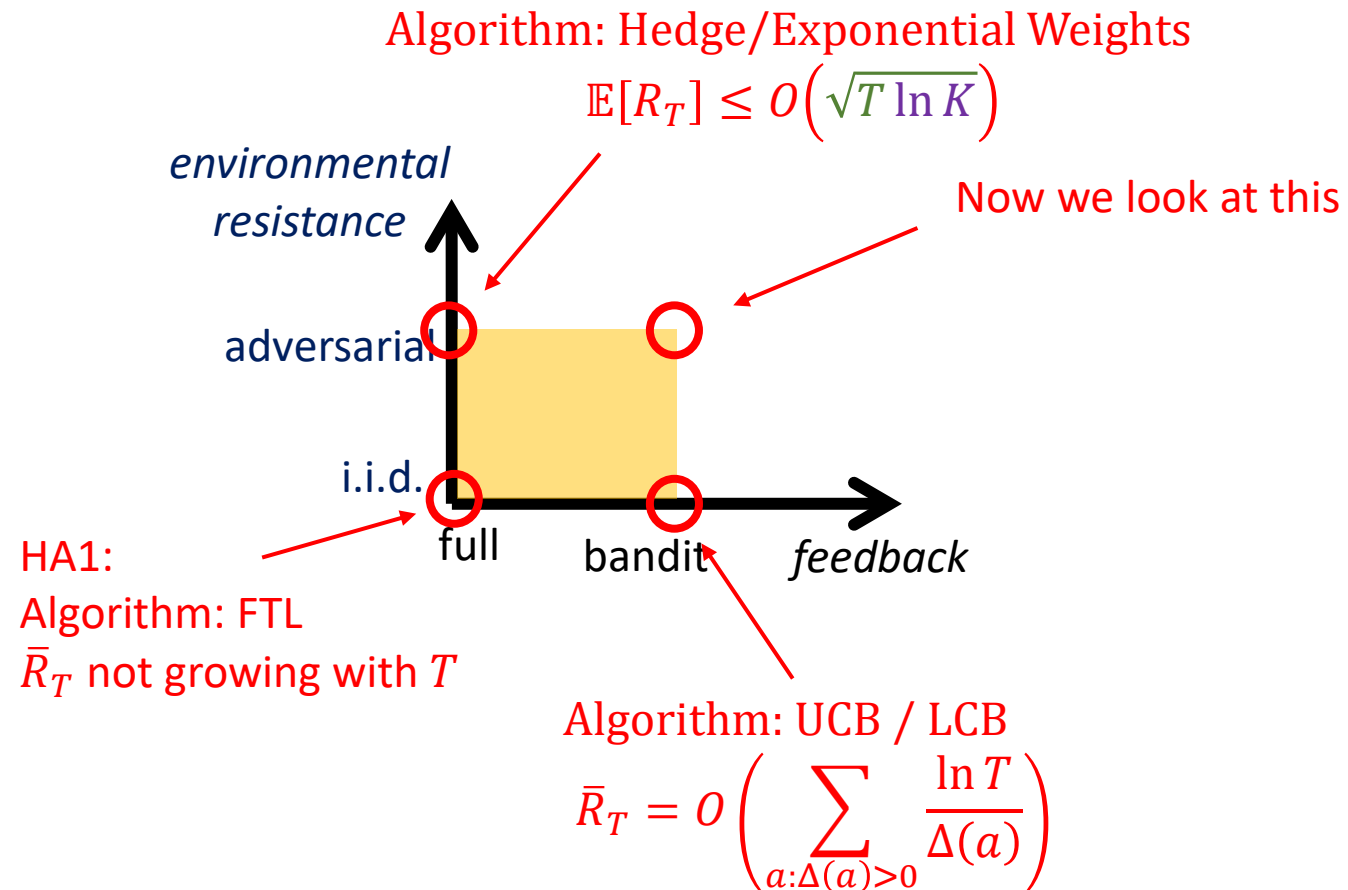
- Analysis:

- Evolution of the potential function $W_t = \sum_a e^{-\eta L_t(a)}$

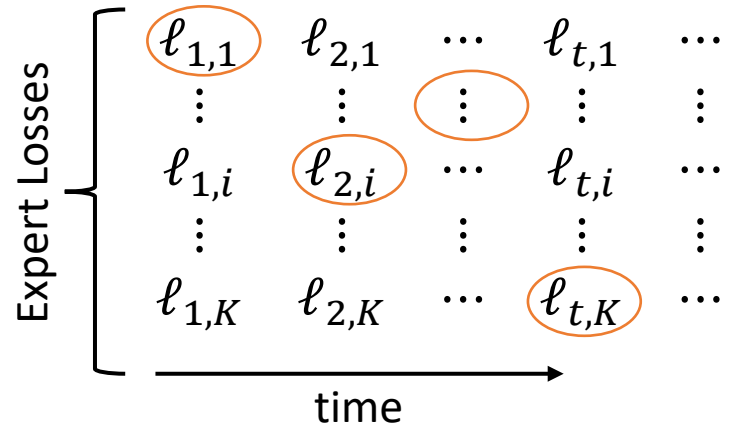
- Matching upper and lower bounds $\mathbb{E}[R_T] = \theta(\sqrt{T \ln K})$



Adversarial bandits



Adversarial bandits



- Performance measures

- Regret:

$$R_T = \sum_{t=1}^T \ell_{t,A_t} - \min_a \sum_{t=1}^T \ell_{t,a}$$

- Expected regret (oblivious setting):

$$\mathbb{E}[R_T] = \mathbb{E}\left[\sum_{t=1}^T \ell_{t,A_t}\right] - \min_a \sum_{t=1}^T \ell_{t,a}$$

Algorithm for adversarial bandits: EXP3

(Exponential Exploration Exploitation)

Hedge \rightarrow EXP3

- $\forall a: L_0(a) = 0$
- For $t = 1, 2, \dots$
 - $\forall a: p_t(a) = \frac{e^{-\eta_t L_{t-1}(a)}}{\sum_{a'} e^{-\eta_t L_{t-1}(a')}}$
 - $A_t \sim p_t$
 - ~~[Observe $\ell_{t,1}, \dots, \ell_{t,K}$]~~
 - [Observe ℓ_{t,A_t}]
 - ~~$\forall a: L_t(a) = L_{t-1}(a) + \ell_{t,a}$~~
 - $\forall a: L_t(a) = L_{t-1}(a) + \frac{\ell_{t,a} \mathbb{1}(A_t=a)}{p_t(a)}$

- Importance-weighted loss estimate

$$\tilde{\ell}_{t,a} = \frac{\ell_{t,a} \mathbb{1}(A_t = a)}{p_t(a)}$$

- Defined for all a

Properties of importance-weighted samples $\tilde{\ell}_{t,a} = \frac{\ell_{t,a} \mathbb{I}(A_t=a)}{p_t(a)}$

- **Not** independent!
 - $\tilde{\ell}_{t,1}, \dots, \tilde{\ell}_{t,K}$ are dependent (only one is nonzero)
 - $p_t(a)$ is a random variable dependent on A_1, \dots, A_{t-1}
- $\tilde{\ell}_{t,a}$ is an unbiased estimate of $\ell_{t,a}$ (meaning $\mathbb{E}[\tilde{\ell}_{t,a}] = \ell_{t,a}$)

$$\begin{aligned}\mathbb{E}[\tilde{\ell}_{t,a}] &= \mathbb{E}\left[\frac{\ell_{t,a} \mathbb{I}(A_t = a)}{p_t(a)}\right] \\ &= \ell_{t,a} \mathbb{E}\left[\frac{\mathbb{I}(A_t=a)}{p_t(a)}\right] \\ &= \ell_{t,a} \mathbb{E}_{A_1, \dots, A_{t-1}} \left[\mathbb{E}_{A_t} \left[\frac{\mathbb{I}(A_t=a)}{p_t(a)} \mid A_1, \dots, A_{t-1} \right] \right] \\ &= \ell_{t,a} \mathbb{E}_{A_1, \dots, A_{t-1}} \left[p_t(a) \frac{1}{p_t(a)} + (1 - p_t(a)) \frac{0}{p_t(a)} \right] \\ &= \ell_{t,a}\end{aligned}$$

- $\ell_{t,a} \in [0,1] \Rightarrow \tilde{\ell}_{t,a} \in \left[0, \frac{1}{p_t(a)}\right]$

Properties continued

$$\tilde{\ell}_{t,a} = \frac{\ell_{t,a} \mathbb{I}(A_t = a)}{p_t(a)}$$



- The second moment of $\tilde{\ell}_{t,a}$ is considerably smaller than the second moment of a general random variable with the same range:

$$\begin{aligned} \mathbb{E}[(\tilde{\ell}_{t,a})^2] &= \mathbb{E}\left[\left(\frac{\ell_{t,a} \mathbb{I}(A_t = a)}{p_t(a)}\right)^2\right] \\ &= \mathbb{E}\left[\frac{\overbrace{(\ell_{t,a})^2}^{\leq 1} \overbrace{(\mathbb{I}(A_t = a))^2}^{=\mathbb{I}(A_t = a)}}{p_t(a)^2}\right] \\ &\leq \mathbb{E}\left[\frac{\mathbb{I}(A_t = a)}{p_t(a)^2}\right] \\ &= \mathbb{E}\left[\frac{1}{p_t(a)}\right] \end{aligned}$$

- “The bandit magic”:

$$\begin{aligned} &\mathbb{E}\left[\sum_a p_t(a) (\tilde{\ell}_{t,a})^2\right] \\ &\leq \mathbb{E}\left[\sum_a p_t(a) \frac{1}{p_t(a)}\right] \\ &= K \end{aligned}$$

Side remark: $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
For importance-weighted sampling the second moment $\mathbb{E}[(\tilde{\ell}_{t,a})^2]$ and the variance $\text{Var}[\tilde{\ell}_{t,a}]$ are close since $\mathbb{E}[\tilde{\ell}_{t,a}]^2$ is small (smaller than 1).

Importance weighted sampling - summary

- $\tilde{\ell}_{t,a} = \frac{\ell_{t,a} \mathbb{I}(A_t=a)}{p_t(a)}$
- Defined for all a
- Unbiased estimates of the losses: $\mathbb{E}[\tilde{\ell}_{t,a}] = \ell_{t,a}$
- Dependent
- Large range $\tilde{\ell}_{t,a} \in \left[0, \frac{1}{p_t(a)}\right]$
- Second moment proportional to the range $\mathbb{E} \left[(\tilde{\ell}_{t,a})^2 \right] \leq \mathbb{E} \left[\frac{1}{p_t(a)} \right]$
 - rather than the square of the range
- The bandit magic: $\mathbb{E} \left[\sum_a p_t(a) (\tilde{\ell}_{t,a})^2 \right] \leq K$



EXP3: Expected regret bound

- By the Hedge lemma ($\tilde{\ell}_{t,a}$ satisfy $\tilde{\ell}_{t,a} \geq 0$):

$$\sum_{t=1}^T \sum_{a=1}^K p_t(a) \tilde{\ell}_{t,a} - \min_a \tilde{L}_T(a) \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \sum_{t=1}^T \sum_{a=1}^K p_t(a) (\tilde{\ell}_{t,a})^2$$

- Taking expectations on both sides:

$$\sum_{t=1}^T \mathbb{E} \left[\sum_{a=1}^K p_t(a) \ell_{t,a} \right] - \mathbb{E} \left[\min_a \tilde{L}_T(a) \right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \sum_{t=1}^T \mathbb{E} \left[\sum_{a=1}^K p_t(a) (\tilde{\ell}_{t,a})^2 \right]$$

- $\mathbb{E}[\min[\cdot]] \leq \min \mathbb{E}[\cdot]$:

$$\sum_{t=1}^T \mathbb{E} \left[\sum_{a=1}^K p_t(a) \ell_{t,a} \right] - \min_a \underbrace{\mathbb{E}[\tilde{L}_T(a)]}_{=L_T(a)} \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \sum_{t=1}^T \mathbb{E} \left[\sum_{a=1}^K p_t(a) (\tilde{\ell}_{t,a})^2 \right]$$

Expected regret bound

$$\underbrace{\sum_{t=1}^T \mathbb{E} \left[\underbrace{\sum_{a=1}^K p_t(a) \ell_{t,a}}_{\text{Expected loss at round } t} \right]}_{\text{Expected loss of EXP3}} - \underbrace{\min_a L_T(a)}_{\text{Best in hindsight}} \leq \frac{\ln K}{\eta} + \underbrace{\frac{\eta}{2} \sum_{t=1}^T \mathbb{E} \left[\sum_{a=1}^K p_t(a) (\tilde{\ell}_{t,a})^2 \right]}_{\leq KT}$$

Expected regret

- Expected regret bound:

$$\mathbb{E}[R_T] \leq \frac{\ln K}{\eta} + \frac{\eta}{2} KT$$

- Optimize with respect to η :

- $\eta = \sqrt{\frac{2 \ln K}{KT}}$

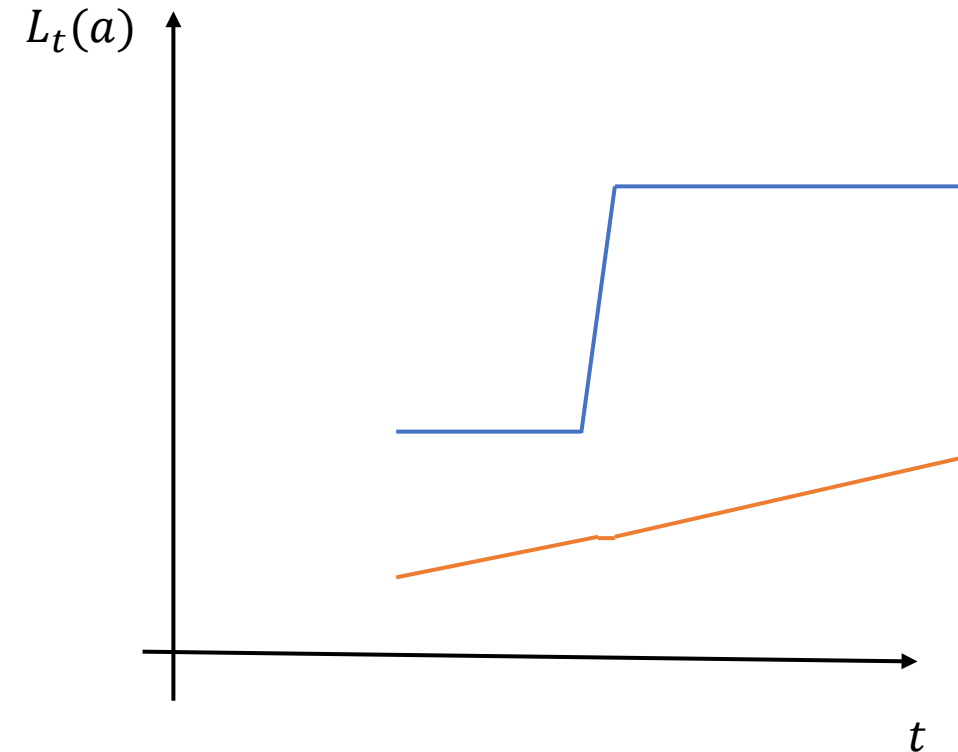
- $\mathbb{E}[R_T] \leq \sqrt{2KT \ln K}$

Algorithm's dynamics

EXP3:

- $\forall a: L_0(a) = 0$
- For $t = 1, 2, \dots$
 - $\forall a: p_t(a) = \frac{e^{-\eta_t L_{t-1}(a)}}{\sum_{a'} e^{-\eta_t L_{t-1}(a')}}$
 - $A_t \sim p_t$
 - [Observe ℓ_{t,A_t}]
 - $\forall a: L_t(a) = L_{t-1}(a) + \frac{\ell_{t,a} \mathbb{I}(A_t=a)}{p_t(a)}$

- Algorithm's dynamics ensures exploration!



Lower bound for adversarial bandits

$$\text{Game 0} \left\{ \begin{array}{ccccc} \ell_{1,1} & \ell_{2,1} & \cdots & \ell_{t,1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,a} & \ell_{2,a} & \cdots & \ell_{t,a} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,K} & \ell_{2,K} & \cdots & \ell_{t,K} & \cdots \end{array} \right. \begin{array}{l} \sim \text{Ber}(1/2) \\ \vdots \\ \sim \text{Ber}(1/2) \\ \vdots \\ \sim \text{Ber}(1/2) \end{array}$$

$$\text{Game 1} \left\{ \begin{array}{ccccc} \ell_{1,1} & \ell_{2,1} & \cdots & \ell_{t,1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,a} & \ell_{2,a} & \cdots & \ell_{t,a} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,K} & \ell_{2,K} & \cdots & \ell_{t,K} & \cdots \end{array} \right. \begin{array}{l} \sim \text{Ber}(1/2 - \varepsilon) \\ \vdots \\ \sim \text{Ber}(1/2) \\ \vdots \\ \sim \text{Ber}(1/2) \end{array}$$

$$\vdots$$

$$\text{Game } K \left\{ \begin{array}{ccccc} \ell_{1,1} & \ell_{2,1} & \cdots & \ell_{t,1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,a} & \ell_{2,a} & \cdots & \ell_{t,a} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,K} & \ell_{2,K} & \cdots & \ell_{t,K} & \cdots \end{array} \right. \begin{array}{l} \sim \text{Ber}(1/2) \\ \vdots \\ \sim \text{Ber}(1/2) \\ \vdots \\ \sim \text{Ber}(1/2 - \varepsilon) \end{array}$$

- At least one action is played at most T/K times

- For that action it is impossible to distinguish between $\text{Ber}(1/2)$ and $\text{Ber}\left(1/2 - 1/\sqrt{T/K}\right) = \text{Ber}\left(1/2 - \sqrt{K/T}\right)$

- $\mathbb{E}[R_T] \geq T\sqrt{K/T} = \sqrt{KT}$

Summary

