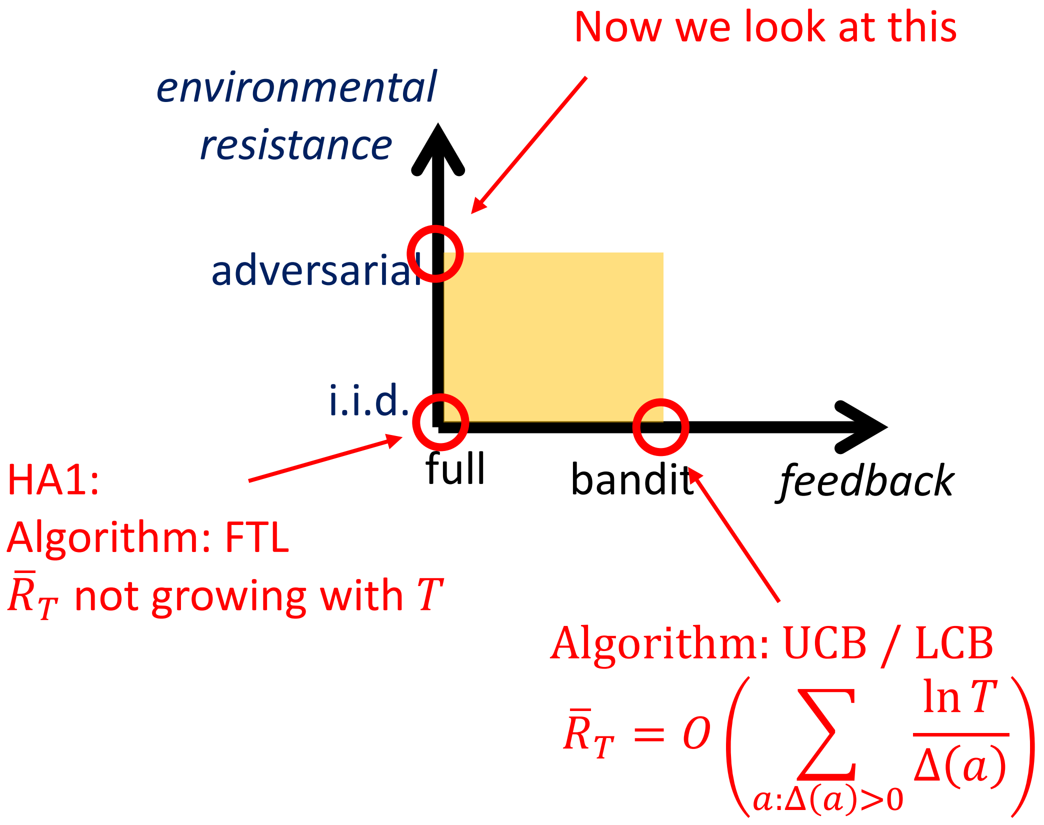
Prediction with Expert Advice and Adversarial Bandits

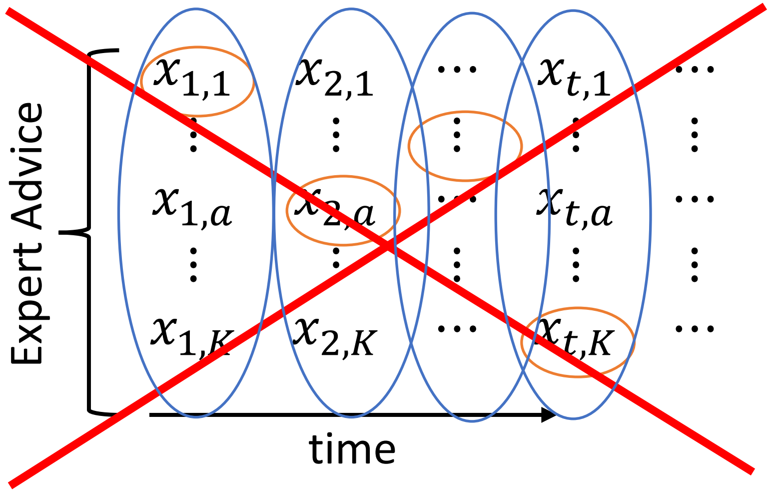
Yevgeny Seldin

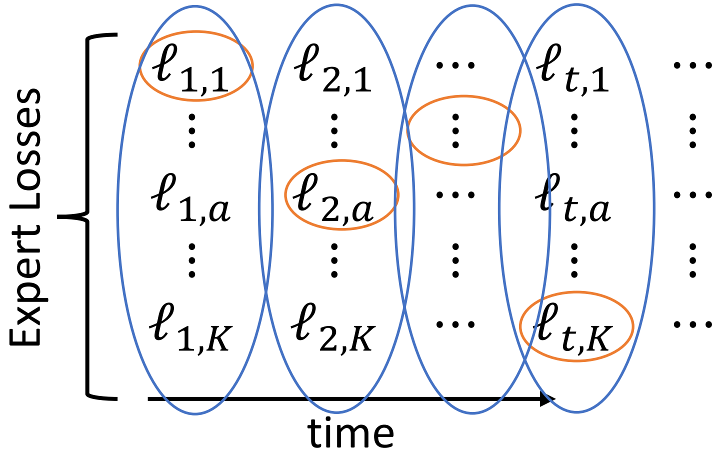
So far



Now we look at this
environmental resistance
adversarial
i.i.d.
full
bandit
feedback
HA1:
Algorithm: FTL
$\bar{R}_T$ not growing with T
Algorithm: UCB / LCB
$\bar{R}_T = O\left(\sum_{a:\Delta(a)>0} \frac{\ln T}{\Delta(a)}\right)$

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

- 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

$$\underbrace{\underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a=1}^{K} p_t(a)\ell_{t,a}}_{\text{The expected loss of Hedge at round } t}}_{\text{The expected loss of Hedge}} - \underbrace{\min_a L_T(a)}_{\text{The best loss in hindsight}}}_{\text{The expected regret of Hedge } \mathbb{E}[R_T]} \le \frac{\ln K}{\eta} + \frac{\eta}{2} \underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a=1}^{K} p_t(a) \underbrace{\left(\ell_{t,a}\right)^2}_{\le 1}}_{\le 1}}_{\le T}$$

- Corollary: $\mathbb{E}[R_T] \le \frac{\ln K}{\eta} + \frac{\eta}{2} T$

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

Proof of the lemma

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

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

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

$$\frac{W_t}{W_{t-1}} = \frac{\sum_a e^{-\eta L_t(a)}}{\sum_{a'} e^{-\eta L_{t-1}(a')}}$$

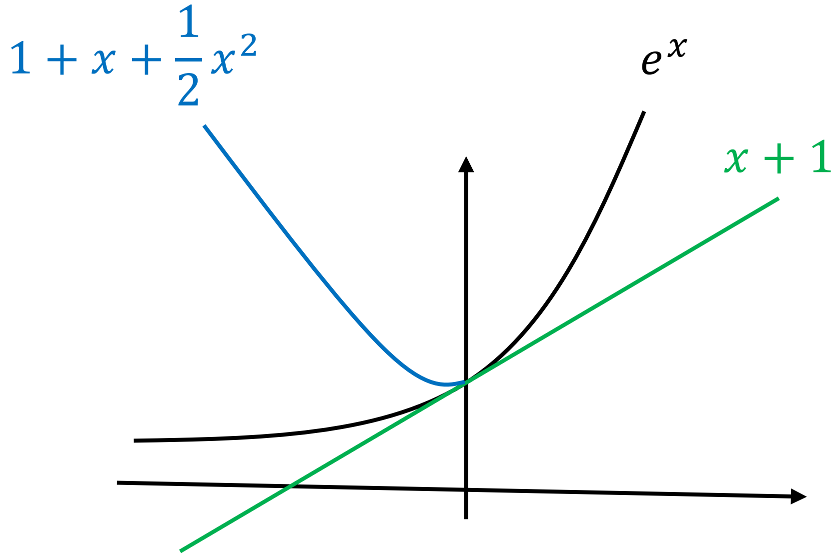
$$= \sum_a e^{-\eta \ell_{t,a}} \underbrace{\frac{e^{-\eta L_{t-1}(a)}}{\sum_{a'} e^{-\eta L_{t-1}(a')}}}_{p_t(a)}$$

$$= \sum_a e^{-\eta \ell_{t,a}} p_t(a)$$

$$\le \sum_a \left(1 - \eta \ell_{t,a} + \frac{1}{2}\eta^2 \left(\ell_{t,a}\right)^2\right) p_t(a)$$

$$= 1 - \eta \sum_a \ell_{t,a} p_t(a) + \frac{\eta^2}{2}\sum_a \left(\ell_{t,a}\right)^2 p_t(a)$$

$$\le e^{-\eta \sum_a \ell_{t,a} p_t(a) + \frac{\eta^2}{2}\sum_a \left(\ell_{t,a}\right)^2 p_t(a)}$$



- For $x \le 0$:

$$e^x \le 1 + x + \frac{1}{2}x^2$$

- For any x:

$$1 + x \le e^x$$

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

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

Proof continued

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

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

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

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

Full information lower bound

$$\text{Expert Losses}\left\{\begin{array}{cccccc} \ell_{1,1} & \ell_{2,1} & \cdots & \ell_{t,1} & \cdots & \sim Ber(1/2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,i} & \ell_{2,i} & \cdots & \ell_{t,i} & \cdots & \sim Ber(1/2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,K} & \ell_{2,K} & \cdots & \ell_{t,K} & \cdots & \sim Ber(1/2) \end{array}\right.$$

$\xrightarrow{\hspace{3cm}}$ *time*

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

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

- Lemma

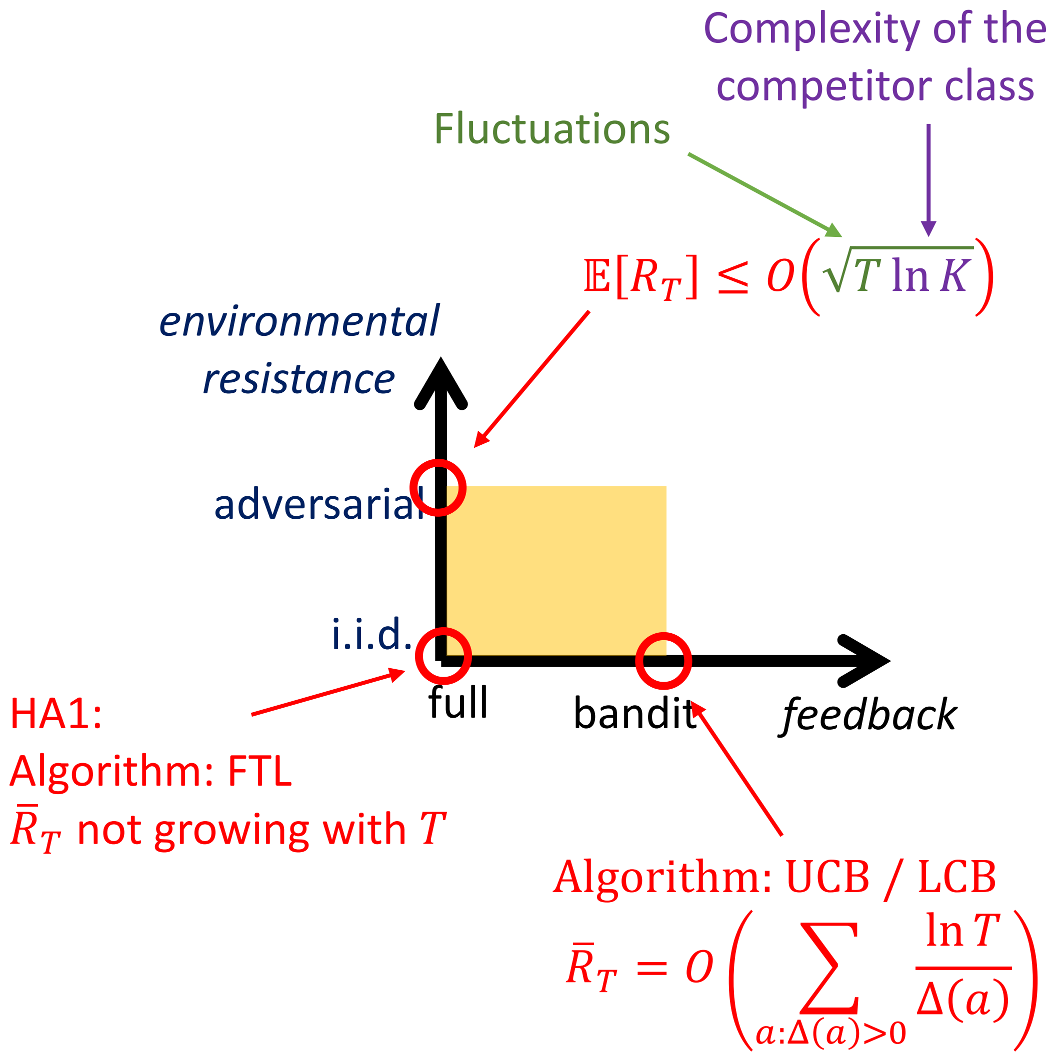
$$\lim_{T\to\infty} \lim_{K\to\infty} \frac{\frac{T}{2} - \mathbb{E}\left[\min_a L_T(a)\right]}{\sqrt{\frac{1}{2} T \ln K}} = 1$$

- In the limit of large T and K:

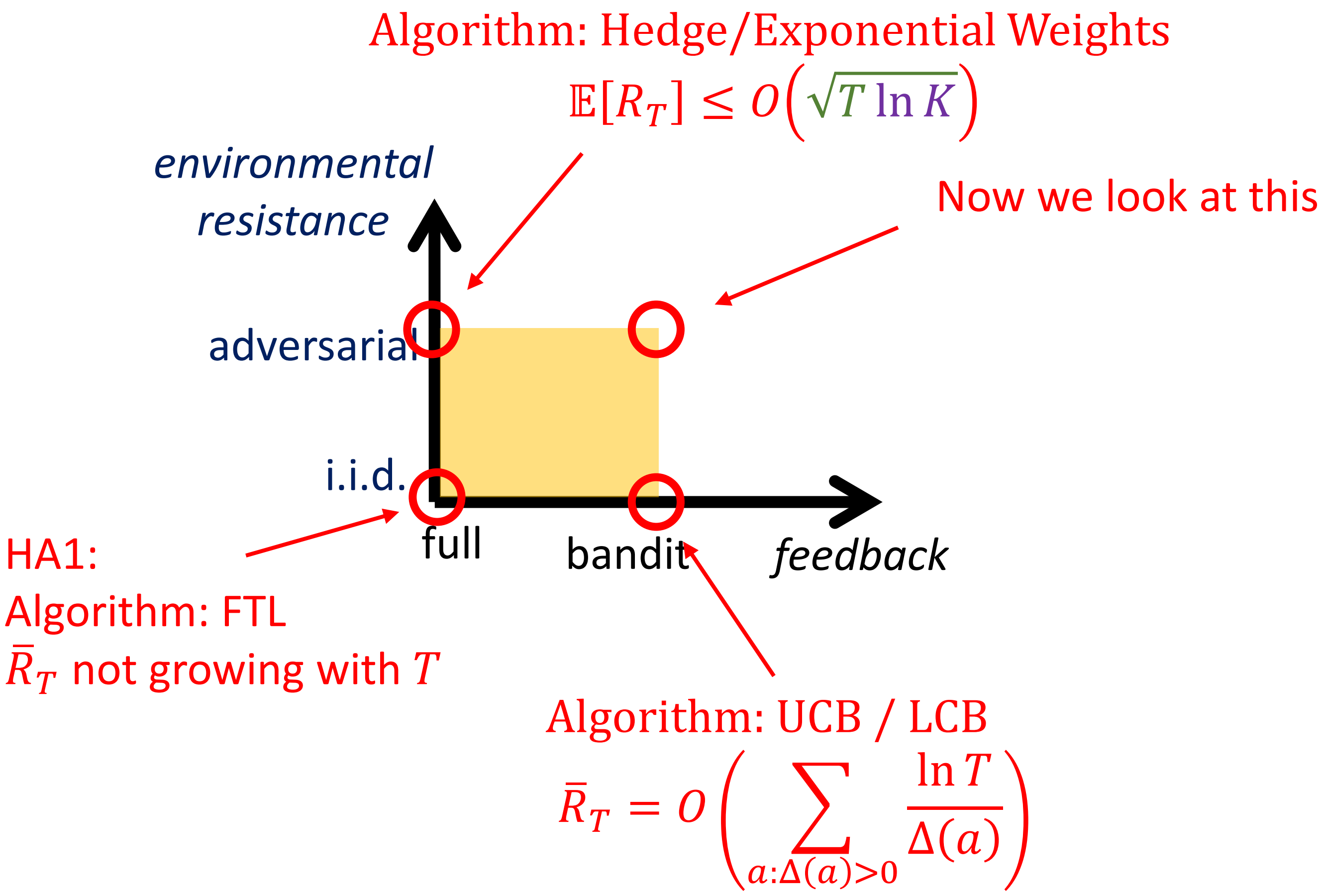
$$\underbrace{\underbrace{\frac{T}{2}}_{\mathbb{E}\left[\sum_t^T \ell_{t,A_t}\right]} - \mathbb{E}\left[\min_a L_T(a)\right]}_{\mathbb{E}[R_T]} \approx \sqrt{\frac{1}{2} T \underbrace{\ln K}_{\text{Complexity of the competitor (Amount of selection)}}}$$

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 $\mathbb{E}[R_T] = \theta\left(\sqrt{T \ln K}\right)$



Adversarial bandits



Adversarial bandits

$$\text{Expert Losses}\left\{\begin{matrix} \ell_{1,1} & \ell_{2,1} & \cdots & \ell_{t,1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,i} & \ell_{2,i} & \cdots & \ell_{t,i} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,K} & \ell_{2,K} & \cdots & \ell_{t,K} & \cdots \end{matrix}\right.$$

$$\xrightarrow[\text{time}]{}$$

- 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 → EXP3

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

- Importance-weighted loss estimate

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

- 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 $\mathbb{E}[\tilde{\ell}_{t,a}] = \ell_{t,a}$)

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

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

Properties continued

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

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

$$\mathbb{E}\left[\left(\tilde{\ell}_{t,a}\right)^2\right] = \mathbb{E}\left[\left(\frac{\ell_{t,a}\mathbb{I}(A_t = a)}{p_t(a)}\right)^2\right]$$

$$= \mathbb{E}\left[\frac{\overbrace{\left(\ell_{t,a}\right)^2}^{\leq 1}\overbrace{\left(\mathbb{I}(A_t=a)\right)^2}^{=\mathbb{I}(A_t=a)}}{p_t(a)^2}\right]$$

$$\leq \mathbb{E}\left[\frac{\mathbb{I}(A_t=a)}{p_t(a)^2}\right]$$

$$= \mathbb{E}\left[\frac{1}{p_t(a)}\right]$$

- "The bandit magic":

$$\mathbb{E}\left[\sum_a p_t(a)\left(\tilde{\ell}_{t,a}\right)^2\right]$$

$$\leq \mathbb{E}\left[\sum_a p_t(a)\frac{1}{p_t(a)}\right]$$

$$= K$$

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 $\tilde{\ell}_{t,a} \in \left[0, \frac{1}{p_t(a)}\right]$
- Variance proportional to the range $\mathbb{E}\left[\left(\tilde{\ell}_{t,a}\right)^2\right] \leq \mathbb{E}\left[\frac{1}{p_t(a)}\right]$
  - rather than the square of the range
- The bandit magic: $\mathbb{E}\left[\sum_a p_t(a)\left(\tilde{\ell}_{t,a}\right)^2\right] \leq K$

EXP3: Expected regret bound

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

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

- Taking expectations on both sides:

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

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

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

Expected regret bound

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

- Expected regret bound:

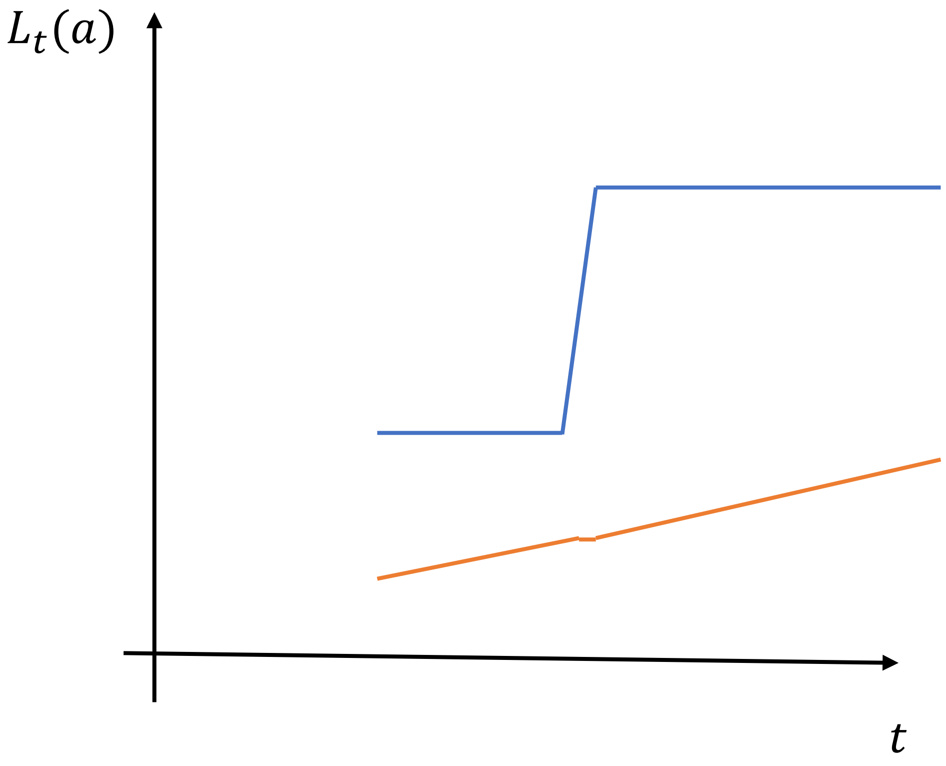
$$\mathbb{E}[R_T] \le \frac{\ln K}{\eta} + \frac{\eta}{2} KT$$

- Optimize with respect to η:
  - $\eta = \sqrt{\frac{2\ln K}{KT}}$
  - $\mathbb{E}[R_T] \le \sqrt{2KT\ln K}$

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

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

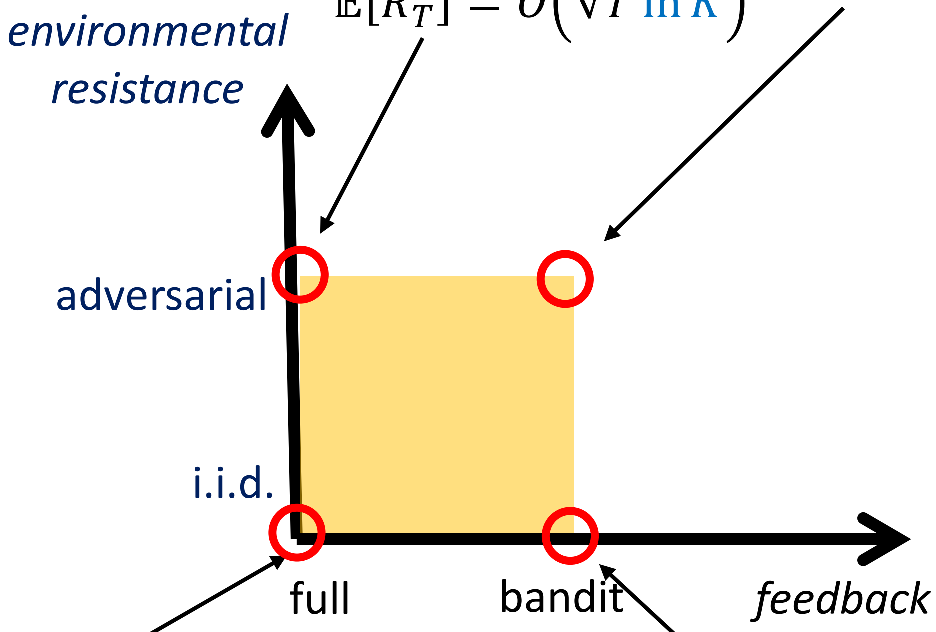
Game 1
$$\left\{\begin{array}{cccccc} \ell_{1,1} & \ell_{2,1} & \cdots & \ell_{t,1} & \cdots & \sim Ber(1/2-\varepsilon) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \ell_{1,a} & \ell_{2,a} & \cdots & \ell_{t,a} & \cdots & \sim Ber(1/2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots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

$$\vdots$$

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

- At least one action is played at most T/K times
- For that action it is impossible to distinguish between $Ber(1/2)$ and $Ber\left(1/2 - 1/\sqrt{T/K}\right) = Ber\left(1/2 - \sqrt{K/T}\right)$
- $\mathbb{E}[R_T] \geq T\sqrt{K/T} = \sqrt{KT}$

Summary



Complexity of the competitor class

$\mathbb{E}[R_T] = O\left(\sqrt{T \ln K}\right)$

weighted variance of importance-weighted estimates

$\mathbb{E}[R_T] = O\left(\sqrt{KT \ln K}\right)$

Can be removed
(but wait for the next lecture for the complexity to come back)

Constant pseudo-regret

Logarithmic pseudo-regret

$$\bar{R}_T = O\left(\sum_{a:\Delta(a)>0} \frac{\ln T}{\Delta(a)}\right)$$