Adversarial Bandits: Theory and Algorithms

Haipeng Luo

University of Southern California

Adversarial (a.k.a. Non-Stochastic) Multi-Armed Bandits

Proposed by Auer, Cesa-Bianchi, Freund, and Schapire, 2002:

Adversarial (a.k.a. Non-Stochastic) Multi-Armed Bandits

Proposed by Auer, Cesa-Bianchi, Freund, and Schapire, 2002:

For $t = 1, \ldots, T$,

- learner picks one of K arms: $i_t \in [K] \triangleq \{1, \ldots, K\}$

Adversarial (a.k.a. Non-Stochastic) Multi-Armed Bandits

Proposed by Auer, Cesa-Bianchi, Freund, and Schapire, 2002:

For $t = 1, \ldots, T$,

- learner picks one of K arms: $i_t \in [K] \triangleq \{1, \ldots, K\}$
- simultaneously adversary decides a loss vector $\ell_t \in [0,1]^K$ ($\ell_{t,i}$ denotes the loss for arm i)

Adversarial (a.k.a. Non-Stochastic) Multi-Armed Bandits

Proposed by Auer, Cesa-Bianchi, Freund, and Schapire, 2002:

For $t = 1, \ldots, T$,

- learner picks one of K arms: $i_t \in [K] \triangleq \{1, \ldots, K\}$
- simultaneously adversary decides a loss vector $\ell_t \in [0,1]^K$ ($\ell_{t,i}$ denotes the loss for arm i)
- learner suffers and only observes loss ℓ_{t,i_t}

Adversarial (a.k.a. Non-Stochastic) Multi-Armed Bandits

Proposed by Auer, Cesa-Bianchi, Freund, and Schapire, 2002:

For $t = 1, \ldots, T$,

- learner picks one of K arms: $i_t \in [K] \triangleq \{1, \ldots, K\}$
- simultaneously adversary decides a loss vector $\ell_t \in [0, 1]^K$ ($\ell_{t,i}$ denotes the loss for arm i)
- learner suffers and only observes loss ℓ_{t,i_t}

Goal: minimize regret

$$\mathrm{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star})$$

Adversarial (a.k.a. Non-Stochastic) Multi-Armed Bandits

Proposed by Auer, Cesa-Bianchi, Freund, and Schapire, 2002:

For $t = 1, \ldots, T$,

- learner picks one of K arms: $i_t \in [K] \triangleq \{1, \ldots, K\}$
- simultaneously adversary decides a loss vector $\ell_t \in [0,1]^K$ ($\ell_{t,i}$ denotes the loss for arm i)
- learner suffers and only observes loss ℓ_{t,i_t}

Goal: minimize regret

$$\mathrm{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star})$$

Stochastic MAB is a special case where $\ell_1, \ldots, \ell_T$ are iid generated

A Closer Look

Why adversarial?

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

Why regret? $(\mathrm{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star}))$

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

Why regret? $(\mathrm{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star}))$

- *why compare with a fixed arm while losses are changing*?

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

Why regret? $(\mathrm{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star}))$

- *why compare with a fixed arm while losses are changing*?
  - fixes: interval/switching/dynamic regret, internal/swap regret

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

Why regret? $(\mathrm{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star}))$

- *why compare with a fixed arm while losses are changing*?
  - fixes: interval/switching/dynamic regret, internal/swap regret
- *why compare with the same losses while the behavior has changed*?

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

Why regret? $(\text{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star}))$

- *why compare with a fixed arm while losses are changing*?
  - fixes: interval/switching/dynamic regret, internal/swap regret
- *why compare with the same losses while the behavior has changed*?
  - make sense for "oblivious" adversary (ℓ_t independent of $i_{1:t-1}$)

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

Why regret? $(\mathrm{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star}))$

- *why compare with a fixed arm while losses are changing*?
  - fixes: interval/switching/dynamic regret, internal/swap regret
- *why compare with the same losses while the behavior has changed*?
  - make sense for "oblivious" adversary (ℓ_t independent of $i_{1:t-1}$)
  - fix for adaptive adversary: policy regret

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

Why regret? $(\mathrm{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star}))$

- *why compare with a fixed arm while losses are changing?*
  - fixes: interval/switching/dynamic regret, internal/swap regret
- *why compare with the same losses while the behavior has changed?*
  - make sense for "oblivious" adversary (ℓ_t independent of $i_{1:t-1}$)
  - fix for adaptive adversary: policy regret
- **but studying the standard regret is still very meaningful!**

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

Why regret? $(\mathrm{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star}))$

- *why compare with a fixed arm while losses are changing*?
  - fixes: interval/switching/dynamic regret, internal/swap regret
- *why compare with the same losses while the behavior has changed*?
  - make sense for "oblivious" adversary (ℓ_t independent of $i_{1:t-1}$)
  - fix for adaptive adversary: policy regret
- **but studying the standard regret is still very meaningful!**
  - foundation for all other regret measures

A Closer Look

Why adversarial?

- remove any distributional assumptions $\Rightarrow$ more robust algorithms
- useful for playing games against arbitrary opponents

Why regret? $(\text{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star}))$

- *why compare with a fixed arm while losses are changing*?
  - fixes: interval/switching/dynamic regret, internal/swap regret
- *why compare with the same losses while the behavior has changed*?
  - make sense for "oblivious" adversary (ℓ_t independent of $i_{1:t-1}$)
  - fix for adaptive adversary: policy regret
- **but studying the standard regret is still very meaningful!**
  - foundation for all other regret measures
  - for games, implies convergence to equilibrium/optimal social welfare

A Marriage Between Two Literatures

Adversarial MAB (and other extensions) combines:

- online learning with adversarial losses
- bandit feedback (i.e. partial information)

A Marriage Between Two Literatures

Adversarial MAB (and other extensions) combines:

- online learning with adversarial losses
- bandit feedback (i.e. partial information)

Algorithms are all based on the following **recipe**:

- first come up with an algorithm that works with full-information feedback (i.e., ℓ_t is revealed at the end of round t)

A Marriage Between Two Literatures

Adversarial MAB (and other extensions) combines:

- online learning with adversarial losses
- bandit feedback (i.e. partial information)

Algorithms are all based on the following **recipe**:

- first come up with an algorithm that works with full-information feedback (i.e., ℓ_t is revealed at the end of round t)
- then come up with a loss estimator in the bandit setting, to be fed to the full-info algorithm

A Marriage Between Two Literatures

Adversarial MAB (and other extensions) combines:

- online learning with adversarial losses
- bandit feedback (i.e. partial information)

Algorithms are all based on the following **recipe**:

- first come up with an algorithm that works with full-information feedback (i.e., ℓ_t is revealed at the end of round t)
- then come up with a loss estimator in the bandit setting, to be fed to the full-info algorithm
- key challenge: **"controlling" the variance of estimators**

A Marriage Between Two Literatures

Adversarial MAB (and other extensions) combines:

- online learning with adversarial losses
- bandit feedback (i.e. partial information)

Algorithms are all based on the following **recipe**:

- first come up with an algorithm that works with full-information feedback (i.e., ℓ_t is revealed at the end of round t)
- then come up with a loss estimator in the bandit setting, to be fed to the full-info algorithm
- key challenge: **"controlling" the variance of estimators**

For this talk:

- start with the full-information case as a warm-up

A Marriage Between Two Literatures

Adversarial MAB (and other extensions) combines:

- online learning with adversarial losses
- bandit feedback (i.e. partial information)

Algorithms are all based on the following **recipe**:

- first come up with an algorithm that works with full-information feedback (i.e., ℓ_t is revealed at the end of round t)
- then come up with a loss estimator in the bandit setting, to be fed to the full-info algorithm
- key challenge: **"controlling" the variance of estimators**

For this talk:

- start with the full-information case as a warm-up
- highlight how to control the variance of estimators

A Marriage Between Two Literatures

Adversarial MAB (and other extensions) combines:

- online learning with adversarial losses
- bandit feedback (i.e. partial information)

Algorithms are all based on the following **recipe**:

- first come up with an algorithm that works with full-information feedback (i.e., ℓ_t is revealed at the end of round t)
- then come up with a loss estimator in the bandit setting, to be fed to the full-info algorithm
- key challenge: **"controlling" the variance of estimators**

For this talk:

- start with the full-information case as a warm-up
- highlight how to control the variance of estimators
- highlight the differences between full-info and bandit

Warm-Up: The Expert Problem

The Expert Problem

Freund-Schapire'97

The full-info counterpart of adversarial MAB:

For $t = 1, \ldots, T$,

- learner picks one of K arms: $i_t \in [K] \triangleq \{1, \ldots, K\}$
- simultaneously adversary decides a loss vector $\ell_t \in [0,1]^K$ ($\ell_{t,i}$ denotes the loss for arm i)
- learner suffers loss ℓ_{t,i_t} and observes ℓ_t (instead of only ℓ_{t,i_t})

The Expert Problem

Freund-Schapire'97

The full-info counterpart of adversarial MAB:

For $t = 1, \dots, T$,

- learner picks one of K arms: $i_t \in [K] \triangleq \{1, \dots, K\}$
- simultaneously adversary decides a loss vector $\ell_t \in [0, 1]^K$ ($\ell_{t,i}$ denotes the loss for arm i)
- learner suffers loss ℓ_{t,i_t} and observes ℓ_t (instead of only ℓ_{t,i_t})

Same goal: minimize regret

$$\text{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star})$$

The Expert Problem

Freund-Schapire'97

The full-info counterpart of adversarial MAB:

For $t = 1, \ldots, T$,

- learner picks one of K arms: $i_t \in [K] \triangleq \{1, \ldots, K\}$
- simultaneously adversary decides a loss vector $\ell_t \in [0,1]^K$ ($\ell_{t,i}$ denotes the loss for arm i)
- learner suffers loss ℓ_{t,i_t} and observes ℓ_t (instead of only ℓ_{t,i_t})

Same goal: minimize regret

$$\text{Reg} = \max_{i^\star \in [K]} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star})$$

Not trivial at all even with full information!

The Classical Algorithm

Littlestone-Warmuth'94, Freund-Schapire'97

At round t, sample $i_t \sim p_t \in \Delta_K$ s.t. (for some learning rate $\eta > 0$)

$$p_{t,i} \propto \exp\left(-\eta \sum_{\tau<t} \ell_{\tau,i}\right)$$

The Classical Algorithm

Littlestone-Warmuth'94, Freund-Schapire'97

At round t, sample $i_t \sim p_t \in \Delta_K$ s.t. (for some learning rate $\eta > 0$)

$$p_{t,i} \propto \exp\left(-\eta \sum_{\tau<t} \ell_{\tau,i}\right)$$

called by many names: Hedge, Multiplicative Weights Update (MWU), ...

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^K \exp(-\eta\sum_{\tau\leq t}\ell_{\tau,i})\right)$.

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^K \exp(-\eta\sum_{\tau\le t}\ell_{\tau,i})\right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta}\ln\left(\frac{\sum_{i=1}^K \exp(-\eta\sum_{\tau\le t}\ell_{\tau,i})}{\sum_{i=1}^K \exp(-\eta\sum_{\tau< t}\ell_{\tau,i})}\right)$$

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^K \exp(-\eta\sum_{\tau\leq t}\ell_{\tau,i})\right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta}\ln\left(\frac{\sum_{i=1}^K \exp(-\eta\sum_{\tau<t}\ell_{\tau,i})\exp(-\eta\ell_{t,i})}{\sum_{i=1}^K \exp(-\eta\sum_{\tau<t}\ell_{\tau,i})}\right)$$

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^{K}\exp(-\eta\sum_{\tau\leq t}\ell_{\tau,i})\right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta}\ln\left(\sum_{i=1}^{K} p_{t,i}\exp(-\eta\ell_{t,i})\right)$$

A Simple Analysis

$e^{-z} \le 1 - z + z^2, \ \forall z \ge 0$ and $\ell_{t,i} \ge 0$

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^K \exp(-\eta \sum_{\tau \le t} \ell_{\tau,i})\right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\exp(-\eta\ell_{t,i})\right) \le \frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\left(1 - \eta\ell_{t,i} + \eta^2\ell_{t,i}^2\right)\right)$$

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta} \ln \left(\sum_{i=1}^{K} \exp(-\eta \sum_{\tau \leq t} \ell_{\tau,i}) \right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta} \ln \left(\sum_{i=1}^{K} p_{t,i} \exp(-\eta \ell_{t,i}) \right) \leq \frac{1}{\eta} \ln \left(\sum_{i=1}^{K} p_{t,i} \left(1 - \eta \ell_{t,i} + \eta^2 \ell_{t,i}^2 \right) \right)$$

$$= \frac{1}{\eta} \ln \left(1 - \eta \langle p_t, \ell_t \rangle + \eta^2 \sum_{i=1}^{K} p_{t,i} \ell_{t,i}^2 \right)$$

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^{K}\exp(-\eta\sum_{\tau\le t}\ell_{\tau,i})\right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta}\ln\left(\sum_{i=1}^{K} p_{t,i}\exp(-\eta\ell_{t,i})\right) \le \frac{1}{\eta}\ln\left(\sum_{i=1}^{K} p_{t,i}\left(1-\eta\ell_{t,i}+\eta^2\ell_{t,i}^2\right)\right)$$

$$= \frac{1}{\eta}\ln\left(1-\eta\langle p_t,\ell_t\rangle+\eta^2\sum_{i=1}^{K}p_{t,i}\ell_{t,i}^2\right) \le -\langle p_t,\ell_t\rangle+\eta\sum_{i=1}^{K}p_{t,i}\ell_{t,i}^2$$

$$\ln(1+z)\le z$$

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta} \ln \left(\sum_{i=1}^{K} \exp(-\eta \sum_{\tau \leq t} \ell_{\tau,i}) \right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta} \ln \left(\sum_{i=1}^{K} p_{t,i} \exp(-\eta \ell_{t,i}) \right) \leq \frac{1}{\eta} \ln \left(\sum_{i=1}^{K} p_{t,i} \left(1 - \eta \ell_{t,i} + \eta^2 \ell_{t,i}^2 \right) \right)$$

$$= \frac{1}{\eta} \ln \left(1 - \eta \langle p_t, \ell_t \rangle + \eta^2 \sum_{i=1}^{K} p_{t,i} \ell_{t,i}^2 \right) \leq -\langle p_t, \ell_t \rangle + \eta \sum_{i=1}^{K} p_{t,i} \ell_{t,i}^2$$

Telescoping and rearranging gives:

$$\sum_{t=1}^{T} \langle p_t, \ell_t \rangle \leq \Phi_0 - \Phi_T + \eta \sum_{t=1}^{T} \sum_{i=1}^{K} p_{t,i} \ell_{t,i}^2$$

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^K \exp(-\eta\sum_{\tau\le t}\ell_{\tau,i})\right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\exp(-\eta\ell_{t,i})\right) \le \frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\left(1-\eta\ell_{t,i}+\eta^2\ell_{t,i}^2\right)\right)$$

$$= \frac{1}{\eta}\ln\left(1-\eta\langle p_t,\ell_t\rangle + \eta^2\sum^K p_{t,i}\ell_{t,i}^2\right) \le -\langle p_t,\ell_t\rangle + \eta\sum^K p_{t,i}\ell_{t,i}^2$$

Telescoping

note $\Phi_T \ge \frac{1}{\eta}\ln\exp\left(-\eta\sum_{\tau\le T}\ell_{\tau,i^\star}\right) = -\sum_{\tau\le T}\ell_{\tau,i^\star}$

$$\sum_{t=1}^T\langle p_t,\ell_t\rangle \le \Phi_0 - \Phi_T + \eta\sum_{t=1}^T\sum_{i=1}^K p_{t,i}\ell_{t,i}^2$$

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^K \exp(-\eta\sum_{\tau\le t}\ell_{\tau,i})\right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\exp(-\eta\ell_{t,i})\right) \le \frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\left(1-\eta\ell_{t,i}+\eta^2\ell_{t,i}^2\right)\right)$$

$$= \frac{1}{\eta}\ln\left(1-\eta\langle p_t,\ell_t\rangle+\eta^2\sum^K p_{t,i}\ell_{t,i}^2\right) \le -\langle p_t,\ell_t\rangle+\eta\sum^K p_{t,i}\ell_{t,i}^2$$

Telescoping

note $\Phi_T \ge \frac{1}{\eta}\ln\exp\left(-\eta\sum_{\tau\le T}\ell_{\tau,i^\star}\right) = -\sum_{\tau\le T}\ell_{\tau,i^\star}$

$$\sum_{t=1}^T\langle p_t - e_{i^\star},\ell_t\rangle \le \Phi_0 + \eta\sum_{t=1}^T\sum_{i=1}^K p_{t,i}\ell_{t,i}^2$$

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^K \exp(-\eta\sum_{\tau\le t}\ell_{\tau,i})\right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\exp(-\eta\ell_{t,i})\right) \le \frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\left(1-\eta\ell_{t,i}+\eta^2\ell_{t,i}^2\right)\right)$$
$$= \frac{1}{\eta}\ln\left(1-\eta\langle p_t,\ell_t\rangle + \eta^2\sum_{i=1}^K p_{t,i}\ell_{t,i}^2\right) \le -\langle p_t,\ell_t\rangle + \eta\sum_{i=1}^K p_{t,i}\ell_{t,i}^2$$

Telescoping and rearranging gives:

$$\sum_{t=1}^T \langle p_t - e_{i^\star},\ell_t\rangle \le \Phi_0 + \eta\sum_{t=1}^T\sum_{i=1}^K p_{t,i}\ell_{t,i}^2 = \frac{\ln K}{\eta} + \eta\sum_{t=1}^T\sum_{i=1}^K p_{t,i}\ell_{t,i}^2$$

A Simple Analysis

Define potential $\Phi_t = \frac{1}{\eta}\ln\left(\sum_{i=1}^K \exp(-\eta\sum_{\tau\le t}\ell_{\tau,i})\right)$. Then $\Phi_t - \Phi_{t-1} =$

$$\frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\exp(-\eta\ell_{t,i})\right) \le \frac{1}{\eta}\ln\left(\sum_{i=1}^K p_{t,i}\left(1-\eta\ell_{t,i}+\eta^2\ell_{t,i}^2\right)\right)$$

$$= \frac{1}{\eta}\ln\left(1-\eta\langle p_t,\ell_t\rangle+\eta^2\sum_{i=1}^K p_{t,i}\ell_{t,i}^2\right) \le -\langle p_t,\ell_t\rangle+\eta\sum_{i=1}^K p_{t,i}\ell_{t,i}^2$$

Telescoping and rearranging gives:

$$\sum_{t=1}^T\langle p_t - e_{i^\star},\ell_t\rangle \le \Phi_0+\eta\sum_{t=1}^T\sum_{i=1}^K p_{t,i}\ell_{t,i}^2 = \frac{\ln K}{\eta}+\eta\sum_{t=1}^T\sum_{i=1}^K p_{t,i}\ell_{t,i}^2$$

Since $\ell_{t,i}^2 \le 1$, picking the best η gives $\mathrm{Reg} = \mathcal{O}(\sqrt{T\ln K})$ (optimal)

A More Modern View

Hedge is a special case of Follow-the-Regularized-Leader (FTRL):

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \ell_\tau \right\rangle + \frac{1}{\eta} \psi(p)$$

A More Modern View

Hedge is a special case of Follow-the-Regularized-Leader (FTRL):

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \ell_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

where $\psi(p) = \frac{1}{\eta}\sum_i p_i \ln p_i$ is the (negative) Shannon entropy regularizer.

A More Modern View

Hedge is a special case of Follow-the-Regularized-Leader (FTRL):

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \ell_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

where $\psi(p) = \frac{1}{\eta}\sum_i p_i \ln p_i$ is the (negative) Shannon entropy regularizer.

Under some conditions, FTRL (with general ψ) ensures for any $p^\star \in \Delta_K$:

$$\sum_{t=1}^T \langle p_t - p^\star, \ell_t \rangle \lesssim \frac{\psi(p^\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^T \|\ell_t\|_{p_t}^2$$

stability term

penalty term

A More Modern View

Hedge is a special case of Follow-the-Regularized-Leader (FTRL):

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \ell_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

where $\psi(p) = \frac{1}{\eta}\sum_i p_i \ln p_i$ is the (negative) Shannon entropy regularizer.

Under some conditions, FTRL (with general ψ) ensures for any $p^\star \in \Delta_K$:

$$\sum_{t=1}^T \langle p_t - p^\star, \ell_t \rangle \lesssim \frac{\psi(p^\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^T \|\ell_t\|_{p_t}^2$$

stability term

penalty term

- $\|\ell_t\|_{p_t}^2 = \ell_t^\top \nabla^{-2}\psi(p_t)\ell_t$ (important **local norm**)

A More Modern View

Hedge is a special case of Follow-the-Regularized-Leader (FTRL):

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \ell_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

where $\psi(p) = \frac{1}{\eta}\sum_i p_i \ln p_i$ is the (negative) Shannon entropy regularizer.

Under some conditions, FTRL (with general ψ) ensures for any $p^\star \in \Delta_K$:

$$\sum_{t=1}^T \langle p_t - p^\star, \ell_t \rangle \lesssim \frac{\psi(p^\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^T \|\ell_t\|_{p_t}^2$$

stability term

penalty term

- $\|\ell_t\|_{p_t}^2 = \ell_t^\top \nabla^{-2}\psi(p_t)\ell_t$ (important **local norm**)
- for Shannon entropy: $\|\ell_t\|_{p_t}^2 = \sum_i p_{t,i}\ell_{t,i}^2$

From Full-Info to Bandit

The Exp3 Algorithm

Auer-CesaBianchi-Freund-Schapire'02

Obvious issue in MAB: only one coordinate of ℓ_t is observed

The Exp3 Algorithm

Auer-CesaBianchi-Freund-Schapire'02

Obvious issue in MAB: only one coordinate of ℓ_t is observed

Solution: construct an importance-weighted estimator $\widehat{\ell}_t$ with

$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\}$$

The Exp3 Algorithm

Auer-CesaBianchi-Freund-Schapire'02

Obvious issue in MAB: only one coordinate of ℓ_t is observed

Solution: construct an importance-weighted estimator $\widehat{\ell}_t$ with

$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\}$$

- non-zero only when $i = i_t$ (the selected arm), thus computable

The Exp3 Algorithm

Auer-CesaBianchi-Freund-Schapire'02

Obvious issue in MAB: only one coordinate of ℓ_t is observed

Solution: construct an importance-weighted estimator $\widehat{\ell}_t$ with

$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\}$$

- non-zero only when $i = i_t$ (the selected arm), thus computable
- clearly **unbiased** ($\mathbb{E}[\widehat{\ell}_{t,i}] = \ell_{t,i}$) since $\mathbb{E}[\mathbf{1}\{i_t = i\}] = p_{t,i}$

The Exp3 Algorithm

Auer-CesaBianchi-Freund-Schapire'02

Obvious issue in MAB: only one coordinate of ℓ_t is observed

Solution: construct an importance-weighted estimator $\widehat{\ell}_t$ with

$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\}$$

- non-zero only when $i = i_t$ (the selected arm), thus computable
- clearly **unbiased** ($\mathbb{E}[\widehat{\ell}_{t,i}] = \ell_{t,i}$) since $\mathbb{E}[\mathbf{1}\{i_t = i\}] = p_{t,i}$

Exp3 (**Exp**onential weight for **Exp**loration and **Exp**loitation) = feeding Hedge with loss estimator $\widehat{\ell}_t$: $p_{t,i} \propto \exp\left(-\eta \sum_{\tau<t} \widehat{\ell}_{\tau,i}\right)$

The Exp3 Algorithm

Auer-CesaBianchi-Freund-Schapire'02

Obvious issue in MAB: only one coordinate of ℓ_t is observed

Solution: construct an importance-weighted estimator $\widehat{\ell}_t$ with

$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\}$$

- non-zero only when $i = i_t$ (the selected arm), thus computable
- clearly **unbiased** ($\mathbb{E}[\widehat{\ell}_{t,i}] = \ell_{t,i}$) since $\mathbb{E}[\mathbf{1}\{i_t = i\}] = p_{t,i}$

Exp3 (**Exp**onential weight for **Exp**loration and **Exp**loitation) = feeding Hedge with loss estimator $\widehat{\ell}_t$: $p_{t,i} \propto \exp\left(-\eta \sum_{\tau<t} \widehat{\ell}_{\tau,i}\right)$

Where is the exploration?

The Exp3 Algorithm

Auer-CesaBianchi-Freund-Schapire'02

Obvious issue in MAB: only one coordinate of ℓ_t is observed

Solution: construct an importance-weighted estimator $\widehat{\ell}_t$ with

$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\}$$

- non-zero only when $i = i_t$ (the selected arm), thus computable
- clearly **unbiased** ($\mathbb{E}[\widehat{\ell}_{t,i}] = \ell_{t,i}$) since $\mathbb{E}[\mathbf{1}\{i_t = i\}] = p_{t,i}$

Exp3 (**Exp**onential weight for **Exp**loration and **Exp**loitation) = feeding Hedge with loss estimator $\widehat{\ell}_t$: $p_{t,i} \propto \exp\left(-\eta \sum_{\tau<t} \widehat{\ell}_{\tau,i}\right)$

Where is the exploration?

- every time an arm is selected, its weight gets decreased

The Exp3 Algorithm

Auer-CesaBianchi-Freund-Schapire'02

Obvious issue in MAB: only one coordinate of ℓ_t is observed

Solution: construct an importance-weighted estimator $\widehat{\ell}_t$ with

$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\}$$

- non-zero only when $i = i_t$ (the selected arm), thus computable
- clearly **unbiased** ($\mathbb{E}[\widehat{\ell}_{t,i}] = \ell_{t,i}$) since $\mathbb{E}[\mathbf{1}\{i_t = i\}] = p_{t,i}$

Exp3 (**Exp**onential weight for **Exp**loration and **Exp**loitation) = feeding Hedge with loss estimator $\widehat{\ell}_t$: $p_{t,i} \propto \exp\left(-\eta \sum_{\tau<t} \widehat{\ell}_{\tau,i}\right)$

Where is the exploration?

- every time an arm is selected, its weight gets decreased
- asymmetry between "losses" and "rewards"

Regret Analysis for Exp3

Key challenge: the **variance** of the estimator can be huge

$$\mathbb{E}[\widehat{\ell}^2_{t,i}] = \frac{\ell^2_{t,i}}{p^2_{t,i}}\mathbb{E}[\mathbf{1}\{i_t = i\}] = \frac{\ell^2_{t,i}}{p_{t,i}}$$

Regret Analysis for Exp3

Key challenge: the **variance** of the estimator can be huge

$$\mathbb{E}[\widehat{\ell}_{t,i}^2] = \frac{\ell_{t,i}^2}{p_{t,i}^2}\mathbb{E}[\mathbf{1}\{i_t = i\}] = \frac{\ell_{t,i}^2}{p_{t,i}}$$

Can't avoid this, but can *control how the variance affects the regret*.

Regret Analysis for Exp3

Key challenge: the **variance** of the estimator can be huge

$$\mathbb{E}[\widehat{\ell}_{t,i}^2] = \frac{\ell_{t,i}^2}{p_{t,i}^2}\mathbb{E}[\mathbf{1}\{i_t = i\}] = \frac{\ell_{t,i}^2}{p_{t,i}}$$

Can't avoid this, but can *control how the variance affects the regret*. Recall

$$\sum_{t=1}^{T} \left\langle p_t - e_{i^\star}, \widehat{\ell}_t \right\rangle \leq \frac{\ln K}{\eta} + \eta \sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}_{t,i}^2 \qquad \text{(only need } \widehat{\ell}_{t,i} \geq 0\text{)}$$

Regret Analysis for Exp3

Key challenge: the **variance** of the estimator can be huge

$$\mathbb{E}[\widehat{\ell}_{t,i}^2] = \frac{\ell_{t,i}^2}{p_{t,i}^2}\mathbb{E}[\mathbf{1}\{i_t = i\}] = \frac{\ell_{t,i}^2}{p_{t,i}}$$

Can't avoid this, but can *control how the variance affects the regret*. Recall

$$\sum_{t=1}^{T} \left\langle p_t - e_{i^\star}, \widehat{\ell}_t \right\rangle \leq \frac{\ln K}{\eta} + \eta \sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}_{t,i}^2 \qquad \text{(only need } \widehat{\ell}_{t,i} \geq 0\text{)}$$

Taking expectation gives

$$\mathbb{E}[\mathrm{Reg}] \leq \frac{\ln K}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}_{t,i}^2\right]$$

Regret Analysis for Exp3

Key challenge: the **variance** of the estimator can be huge

$$\mathbb{E}[\widehat{\ell}^2_{t,i}] = \frac{\ell^2_{t,i}}{p^2_{t,i}}\mathbb{E}[\mathbf{1}\{i_t = i\}] = \frac{\ell^2_{t,i}}{p_{t,i}}$$

Can't avoid this, but can *control how the variance affects the regret*. Recall

$$\sum_{t=1}^{T} \left\langle p_t - e_{i^\star}, \widehat{\ell}_t \right\rangle \leq \frac{\ln K}{\eta} + \eta \sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}^2_{t,i} \qquad \text{(only need } \widehat{\ell}_{t,i} \geq 0\text{)}$$

Taking expectation gives

$$\begin{aligned}\mathbb{E}[\text{Reg}] &\leq \frac{\ln K}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}^2_{t,i}\right] \\ &= \frac{\ln K}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i} \cdot \frac{\ell^2_{t,i}}{p_{t,i}}\right] \quad \text{(magical variance cancellation)}\end{aligned}$$

Regret Analysis for Exp3

Key challenge: the **variance** of the estimator can be huge

$$\mathbb{E}[\widehat{\ell}_{t,i}^2] = \frac{\ell_{t,i}^2}{p_{t,i}^2}\mathbb{E}[\mathbf{1}\{i_t = i\}] = \frac{\ell_{t,i}^2}{p_{t,i}}$$

Can't avoid this, but can *control how the variance affects the regret*. Recall

$$\sum_{t=1}^{T}\left\langle p_t - e_{i^\star}, \widehat{\ell}_t\right\rangle \le \frac{\ln K}{\eta} + \eta\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}_{t,i}^2 \qquad \text{(only need } \widehat{\ell}_{t,i} \ge 0\text{)}$$

Taking expectation gives

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}] &\le \frac{\ln K}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}_{t,i}^2\right] \\
&= \frac{\ln K}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\cdot\frac{\ell_{t,i}^2}{p_{t,i}}\right] \qquad \text{(magical variance cancellation)} \\
&\le \frac{\ln K}{\eta} + \eta T K
\end{aligned}$$

Regret Analysis for Exp3

Key challenge: the **variance** of the estimator can be huge

$$\mathbb{E}[\widehat{\ell}_{t,i}^2] = \frac{\ell_{t,i}^2}{p_{t,i}^2}\mathbb{E}[\mathbf{1}\{i_t = i\}] = \frac{\ell_{t,i}^2}{p_{t,i}}$$

Can't avoid this, but can *control how the variance affects the regret*. Recall

$$\sum_{t=1}^{T} \left\langle p_t - e_{i^\star}, \widehat{\ell}_t \right\rangle \leq \frac{\ln K}{\eta} + \eta \sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}_{t,i}^2 \qquad \text{(only need } \widehat{\ell}_{t,i} \geq 0\text{)}$$

Taking expectation gives

$$\begin{aligned}
\mathbb{E}[\text{Reg}] &\leq \frac{\ln K}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}_{t,i}^2\right] \\
&= \frac{\ln K}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i} \cdot \frac{\ell_{t,i}^2}{p_{t,i}}\right] \quad \text{(magical variance cancellation)} \\
&\leq \frac{\ln K}{\eta} + \eta T K = \mathcal{O}(\sqrt{TK\ln K}) \qquad \text{(optimal } \eta\text{)}
\end{aligned}$$

Regret Analysis for Exp3

Key challenge: the **variance** of the estimator can be huge

$$\mathbb{E}[\widehat{\ell}_{t,i}^2] = \frac{\ell_{t,i}^2}{p_{t,i}^2}\mathbb{E}[\mathbf{1}\{i_t = i\}] = \frac{\ell_{t,i}^2}{p_{t,i}}$$

Can't avoid this, but can *control how the variance affects the regret*. Recall

$$\sum_{t=1}^{T} \left\langle p_t - e_{i^\star}, \widehat{\ell}_t \right\rangle \leq \frac{\ln K}{\eta} + \eta \sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}_{t,i}^2 \qquad \text{(only need } \widehat{\ell}_{t,i} \geq 0\text{)}$$

Taking expectation gives (**caveat**: assuming an oblivious adversary)

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}] &\leq \frac{\ln K}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\widehat{\ell}_{t,i}^2\right] \\
&= \frac{\ln K}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} p_{t,i}\cdot\frac{\ell_{t,i}^2}{p_{t,i}}\right] \quad \text{(magical variance cancellation)} \\
&\leq \frac{\ln K}{\eta} + \eta T K = \mathcal{O}(\sqrt{TK\ln K}) \quad \text{(optimal } \eta\text{)}
\end{aligned}$$

$\Omega(\sqrt{TK})$ Lower Bound

$\Omega(\sqrt{TK})$ Lower Bound

An informal argument:

- first consider $\ell_{t,i} \sim \text{Ber}(1/2)$ for all i

$\Omega(\sqrt{TK})$ Lower Bound

An informal argument:

- first consider $\ell_{t,i} \sim \text{Ber}(1/2)$ for all i
- for any algorithm, must exist $j \in [K]$ not selected more than $\frac{T}{K}$ times

$\Omega(\sqrt{TK})$ Lower Bound

An informal argument:

- first consider $\ell_{t,i} \sim \text{Ber}(1/2)$ for all i
- for any algorithm, must exist $j \in [K]$ not selected more than $\frac{T}{K}$ times
- now secretly change the loss of arm j to $\ell_{t,j} \sim \text{Ber}(1/2 - \sqrt{K/T})$

$\Omega(\sqrt{TK})$ Lower Bound

An informal argument:

- first consider $\ell_{t,i} \sim \text{Ber}(1/2)$ for all i
- for any algorithm, must exist $j \in [K]$ not selected more than $\frac{T}{K}$ times
- now secretly change the loss of arm j to $\ell_{t,j} \sim \text{Ber}(1/2 - \sqrt{K/T})$
- the same algorithm **won't realize the change** (information theoretically),

$\Omega(\sqrt{TK})$ Lower Bound

An informal argument:

- first consider $\ell_{t,i} \sim \text{Ber}(1/2)$ for all i
- for any algorithm, must exist $j \in [K]$ not selected more than $\frac{T}{K}$ times
- now secretly change the loss of arm j to $\ell_{t,j} \sim \text{Ber}(1/2 - \sqrt{K/T})$
- the same algorithm **won't realize the change** (information theoretically), so still picks arm j not often enough (e.g. $\leq \frac{T}{2}$ times)

$\Omega(\sqrt{TK})$ Lower Bound

An informal argument:

- first consider $\ell_{t,i} \sim \text{Ber}(1/2)$ for all i
- for any algorithm, must exist $j \in [K]$ not selected more than $\frac{T}{K}$ times
- now secretly change the loss of arm j to $\ell_{t,j} \sim \text{Ber}(1/2 - \sqrt{K/T})$
- the same algorithm **won't realize the change** (information theoretically), so still picks arm j not often enough (e.g. $\leq \frac{T}{2}$ times)
- every time not picking arm j, incur $\sqrt{K/T}$ regret, thus in total, $\mathbb{E}[\text{Reg}] = \Omega(\sqrt{TK})$

$\Omega(\sqrt{TK})$ Lower Bound

An informal argument:

- first consider $\ell_{t,i} \sim \text{Ber}(1/2)$ for all i
- for any algorithm, must exist $j \in [K]$ not selected more than $\frac{T}{K}$ times
- now secretly change the loss of arm j to $\ell_{t,j} \sim \text{Ber}(1/2 - \sqrt{K/T})$
- the same algorithm **won't realize the change** (information theoretically), so still picks arm j not often enough (e.g. $\leq \frac{T}{2}$ times)
- every time not picking arm j, incur $\sqrt{K/T}$ regret, thus in total, $\mathbb{E}[\text{Reg}] = \Omega(\sqrt{TK})$

Note the gap between this and Exp3's regret bound $\mathcal{O}(\sqrt{TK \ln K})$

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the $1/2$-Tsallis entropy $\psi(p) = -\sum_{i=1}^{K} \sqrt{p_i}$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the 1/2-Tsallis entropy $\psi(p) = -\sum_{i=1}^{K} \sqrt{p_i}$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p^\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the 1/2-Tsallis entropy $\psi(p) = -\sum_{i=1}^{K} \sqrt{p_i}$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p^\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \leq \sqrt{K}$

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the 1/2-Tsallis entropy $\psi(p) = -\sum_{i=1}^{K} \sqrt{p_i}$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p^\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \le \sqrt{K}$
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^{3/2} \widehat{\ell}_{t,i}^2$

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the 1/2-Tsallis entropy $\psi(p) = -\sum_{i=1}^K \sqrt{p_i}$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^T \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p^\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^T \|\widehat{\ell}_t\|_{p_t}^2$

recall: $\mathbb{E}[\widehat{\ell}_{t,i}^2] = \frac{\ell_{t,i}^2}{p_{t,i}^2}\mathbb{E}[\mathbf{1}\{i_t = i\}] = \frac{\ell_{t,i}^2}{p_{t,i}}$

- $\psi(p^\star) - \min_p \psi(p) \le \sqrt{K}$
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^{3/2}\widehat{\ell}_{t,i}^2 \xrightarrow{\mathbb{E}} \sum_i \sqrt{p_{t,i}}\ell_{t,i}^2$

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the 1/2-Tsallis entropy $\psi(p) = -\sum_{i=1}^K \sqrt{p_i}$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^T \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p^\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^T \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \leq \sqrt{K}$
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^{3/2}\widehat{\ell}_{t,i}^2 \xrightarrow{\mathbb{E}} \sum_i \sqrt{p_{t,i}}\ell_{t,i}^2 \leq \sqrt{K}$

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the 1/2-Tsallis entropy $\psi(p) = -\sum_{i=1}^{K} \sqrt{p_i}$,

	Shannon	**1/2-Tsallis**
penalty	$\ln K$	$\sqrt{K}$
stability	K	$\sqrt{K}$

Recall: $\sum_{t=}^{T}$ … $\eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \le \sqrt{K}$
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^{3/2} \widehat{\ell}_{t,i}^2 \xrightarrow{\mathbb{E}} \sum_i \sqrt{p_{t,i}} \ell_{t,i}^2 \le \sqrt{K}$

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the $1/2$-Tsallis entropy $\psi(p) = -\sum_{i=1}^K \sqrt{p_i}$,

	Shannon	**$1/2$-Tsallis**
penalty	$\ln K$	$\sqrt{K}$
stability	K	$\sqrt{K}$

Recall: $\sum_{t=}^{T}$ … $\eta \sum_{t=1}^T \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \leq \sqrt{K}$
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^{3/2}\widehat{\ell}_{t,i}^2 \xrightarrow{\mathbb{E}} \sum_i \sqrt{p_{t,i}}\ell_{t,i}^2 \leq \sqrt{K}$
- $\mathbb{E}[\text{Reg}] \lesssim \sqrt{K}\left(\frac{1}{\eta} + \eta T\right)$

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the 1/2-Tsallis entropy $\psi(p) = -\sum_{i=1}^{K} \sqrt{p_i}$,

	Shannon	1/2-Tsallis
penalty	$\ln K$	$\sqrt{K}$
stability	K	$\sqrt{K}$

Recall: $\sum_{t=}^{T}$ … $\eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \leq \sqrt{K}$
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^{3/2}\widehat{\ell}_{t,i}^2 \overset{\mathbb{E}}{\to} \sum_i \sqrt{p_{t,i}}\ell_{t,i}^2 \leq \sqrt{K}$
- $\mathbb{E}[\mathrm{Reg}] \lesssim \sqrt{K}\left(\frac{1}{\eta} + \eta T\right) = \mathcal{O}(\sqrt{TK})$ (optimal η)

Minimax Algorithm

Audibert-Bubeck'09, Abernethy-Lee-Tewari'15

Consider FTRL with the 1/2-Tsallis entropy $\psi(p) = -\sum_{i=1}^{K} \sqrt{p_i}$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p^\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \leq \sqrt{K}$
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^{3/2}\widehat{\ell}_{t,i}^2 \xrightarrow{\mathbb{E}} \sum_i \sqrt{p_{t,i}}\ell_{t,i}^2 \leq \sqrt{K}$
- $\mathbb{E}[\text{Reg}] \lesssim \sqrt{K}\left(\frac{1}{\eta} + \eta T\right) = \mathcal{O}(\sqrt{TK})$ (optimal η)

Teaser for Thodoris' talk: not only minimax optimal for adversarial losses, but (surprisingly) also *instance-optimal for stochastic losses!* (Zimmert-Seldin'19)

Beyond Minimax Optimality:
Adaptive and Problem-Dependent Regret Bounds

Robustness versus Adaptivity

Worst-case robustness ($\sqrt{TK}$-regret) might be overly pessimistic. Can we adapt to easier instances with smaller regret?

Robustness versus Adaptivity

Worst-case robustness ($\sqrt{TK}$-regret) might be overly pessimistic. Can we adapt to easier instances with smaller regret?

Measures of "easiness"	Regret bounds	References
loss of the best arm $L^\star = \sum_t \ell_{t,i^\star}$	$\widetilde{\mathcal{O}}(\sqrt{L^\star K})$	Allenberg-Auer-Györfi-Ottucsák'06 Foster-Li-Lykouris-Sridharan-Tardos'16

Robustness versus Adaptivity

Worst-case robustness ($\sqrt{TK}$-regret) might be overly pessimistic. Can we adapt to easier instances with smaller regret?

Measures of "easiness"	Regret bounds	References
loss of the best arm $L^\star = \sum_t \ell_{t,i^\star}$	$\widetilde{\mathcal{O}}(\sqrt{L^\star K})$	Allenberg-Auer-Györfi-Ottucsák'06 Foster-Li-Lykouris-Sridharan-Tardos'16
variance of losses $Q = \frac{1}{K}\sum_{t,i}(\ell_{t,i} - \frac{1}{T}L_i)$	$\widetilde{\mathcal{O}}(\sqrt{QK})$	Hazan-Kale'11, Bubeck-Cohen-Li'17

Robustness versus Adaptivity

Worst-case robustness ($\sqrt{TK}$-regret) might be overly pessimistic. Can we adapt to easier instances with smaller regret?

Measures of "easiness"	Regret bounds	References
loss of the best arm $L^\star = \sum_t \ell_{t,i^\star}$	$\widetilde{\mathcal{O}}(\sqrt{L^\star K})$	Allenberg-Auer-Györfi-Ottucsák'06 Foster-Li-Lykouris-Sridharan-Tardos'16
variance of losses $Q = \frac{1}{K}\sum_{t,i}(\ell_{t,i} - \frac{1}{T}L_i)$ $Q^\star = \sum_t(\ell_{t,i^\star} - \frac{1}{T}L^\star)$	$\widetilde{\mathcal{O}}(\sqrt{QK})$ $\widetilde{\mathcal{O}}(\sqrt{Q^\star K})$	Hazan-Kale'11, Bubeck-Cohen-Li'17 Wei-Luo'18

Robustness versus Adaptivity

Worst-case robustness ($\sqrt{TK}$-regret) might be overly pessimistic. Can we adapt to easier instances with smaller regret?

Measures of "easiness"	Regret bounds	References
loss of the best arm $L^\star = \sum_t \ell_{t,i^\star}$	$\widetilde{\mathcal{O}}(\sqrt{L^\star K})$	Allenberg-Auer-Györfi-Ottucsák'06 Foster-Li-Lykouris-Sridharan-Tardos'16
variance of losses $Q = \frac{1}{K}\sum_{t,i}(\ell_{t,i} - \frac{1}{T}L_i)$ $Q^\star = \sum_t(\ell_{t,i^\star} - \frac{1}{T}L^\star)$	$\widetilde{\mathcal{O}}(\sqrt{QK})$ $\widetilde{\mathcal{O}}(\sqrt{Q^\star K})$	Hazan-Kale'11, Bubeck-Cohen-Li'17 Wei-Luo'18
path-length of losses $V = \sum_t \|\ell_t - \ell_{t-1}\|_\infty$	$\widetilde{\mathcal{O}}(\sqrt{VK})$	Bubeck-Li-Luo-Wei'19

Robustness versus Adaptivity

Worst-case robustness ($\sqrt{TK}$-regret) might be overly pessimistic. Can we adapt to easier instances with smaller regret?

Measures of "easiness"	Regret bounds	References
loss of the best arm $L^\star = \sum_t \ell_{t,i^\star}$	$\widetilde{\mathcal{O}}(\sqrt{L^\star K})$	Allenberg-Auer-Györfi-Ottucsák'06 Foster-Li-Lykouris-Sridharan-Tardos'16
variance of losses $Q = \frac{1}{K}\sum_{t,i}(\ell_{t,i} - \frac{1}{T}L_i)$ $Q^\star = \sum_t(\ell_{t,i^\star} - \frac{1}{T}L^\star)$	$\widetilde{\mathcal{O}}(\sqrt{QK})$ $\widetilde{\mathcal{O}}(\sqrt{Q^\star K})$	Hazan-Kale'11, Bubeck-Cohen-Li'17 Wei-Luo'18
path-length of losses $V = \sum_t \|\ell_t - \ell_{t-1}\|_\infty$ $V^\star = \sum_t(\ell_{t,i^\star} - \ell_{t-1,i^\star})$	$\widetilde{\mathcal{O}}(\sqrt{VK})$ $\widetilde{\mathcal{O}}(\sqrt{V^\star K^2})$	Bubeck-Li-Luo-Wei'19 Wei-Luo'18

Robustness versus Adaptivity

Worst-case robustness ($\sqrt{TK}$-regret) might be overly pessimistic. Can we adapt to easier instances with smaller regret?

Measures of "easiness"	Regret bounds	References
loss of the best arm $L^\star = \sum_t \ell_{t,i^\star}$	$\widetilde{\mathcal{O}}(\sqrt{L^\star K})$	Allenberg-Auer-Györfi-Ottucsák'06 Foster-Li-Lykouris-Sridharan-Tardos'16
variance of losses $Q = \frac{1}{K}\sum_{t,i}(\ell_{t,i} - \frac{1}{T}L_i)$ $Q^\star = \sum_t(\ell_{t,i^\star} - \frac{1}{T}L^\star)$	$\widetilde{\mathcal{O}}(\sqrt{QK})$ $\widetilde{\mathcal{O}}(\sqrt{Q^\star K})$	Hazan-Kale'11, Bubeck-Cohen-Li'17 Wei-Luo'18
path-length of losses $V = \sum_t \|\ell_t - \ell_{t-1}\|_\infty$ $V^\star = \sum_t(\ell_{t,i^\star} - \ell_{t-1,i^\star})$	$\widetilde{\mathcal{O}}(\sqrt{VK})$ $\widetilde{\mathcal{O}}(\sqrt{V^\star K^2})$	Bubeck-Li-Luo-Wei'19 Wei-Luo'18
sparsity of rewards $s = \max_t \|\mathbf{1} - \ell_t\|_0$	$\sqrt{Ts}$	Bubeck-Cohen-Li'17

Robustness versus Adaptivity

Worst-case robustness ($\sqrt{TK}$-regret) might be overly pessimistic. Can we adapt to easier instances with smaller regret?

Measures of "easiness"	Regret bounds	References
loss of the best arm $L^\star = \sum_t \ell_{t,i^\star}$	$\widetilde{\mathcal{O}}(\sqrt{L^\star K})$	Allenberg-Auer-Györfi-Ottucsák'06 Foster-Li-Lykouris-Sridharan-Tardos'16
variance of losses $Q = \frac{1}{K}\sum_{t,i}(\ell_{t,i} - \frac{1}{T}L_i)$ $Q^\star = \sum_t(\ell_{t,i^\star} - \frac{1}{T}L^\star)$	$\widetilde{\mathcal{O}}(\sqrt{QK})$ $\widetilde{\mathcal{O}}(\sqrt{Q^\star K})$	Hazan-Kale'11, Bubeck-Cohen-Li'17 Wei-Luo'18
path-length of losses $V = \sum_t \|\ell_t - \ell_{t-1}\|_\infty$ $V^\star = \sum_t(\ell_{t,i^\star} - \ell_{t-1,i^\star})$	$\widetilde{\mathcal{O}}(\sqrt{VK})$ $\widetilde{\mathcal{O}}(\sqrt{V^\star K^2})$	Bubeck-Li-Luo-Wei'19 Wei-Luo'18
sparsity of rewards $s = \max_t \|\mathbf{1} - \ell_t\|_0$	$\sqrt{Ts}$	Bubeck-Cohen-Li'17

imply **faster convergence** in games

Achieving Small-Loss Bounds

Consider FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Achieving Small-Loss Bounds

Foster-Li-Lykouris-Sridharan-Tardos'16

Consider FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

Achieving Small-Loss Bounds

Foster-Li-Lykouris-Sridharan-Tardos'16

Consider FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \le K \ln T$ (picking $p^\star = (1 - \frac{1}{T})e_{i^\star} + \frac{1}{TK}\mathbf{1}$)

Achieving Small-Loss Bounds

Foster-Li-Lykouris-Sridharan-Tardos'16

Consider FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \le K \ln T$ (picking $p^\star = (1 - \frac{1}{T})e_{i^\star} + \frac{1}{TK}\mathbf{1}$)
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^2 \widehat{\ell}_{t,i}^2$

Achieving Small-Loss Bounds

Foster-Li-Lykouris-Sridharan-Tardos'16

Consider FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \le K \ln T$ (picking $p^\star = (1 - \frac{1}{T}) e_{i^\star} + \frac{1}{TK}\mathbf{1}$)
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^2 \widehat{\ell}_{t,i}^2 \xrightarrow{\mathbb{E}} \langle p_t, \ell_t \rangle$

Achieving Small-Loss Bounds

Foster-Li-Lykouris-Sridharan-Tardos'16

Consider FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

	Shannon	**1/2-Tsallis**	**log-barrier**
penalty	$\ln K$	$\sqrt{K}$	$K \ln T$
stability	K	$\sqrt{K}$	1

Reca ... $1 \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \le K \ln T$ (picking $p^\star = (1 - \frac{1}{T})e_{i^\star} + \frac{1}{TK}\mathbf{1}$)
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^2 \widehat{\ell}_{t,i}^2 \xrightarrow{\mathbb{E}} \langle p_t, \ell_t \rangle$

Achieving Small-Loss Bounds

Foster-Li-Lykouris-Sridharan-Tardos'16

Consider FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \leq K \ln T$ (picking $p^\star = (1 - \frac{1}{T})e_{i^\star} + \frac{1}{TK}\mathbf{1}$)
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^2 \widehat{\ell}_{t,i}^2 \xrightarrow{\mathbb{E}} \langle p_t, \ell_t \rangle$
- $\mathbb{E}[\text{Reg}] = \widetilde{\mathcal{O}}(\sqrt{K\mathbb{E}[\sum_t \langle p_t, \ell_t \rangle]})$

Achieving Small-Loss Bounds

Foster-Li-Lykouris-Sridharan-Tardos'16

Consider FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Recall: $\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t\|_{p_t}^2$

- $\psi(p^\star) - \min_p \psi(p) \leq K \ln T$ (picking $p^\star = (1 - \frac{1}{T})e_{i^\star} + \frac{1}{TK}\mathbf{1}$)
- $\|\widehat{\ell}_t\|_{p_t}^2 = \widehat{\ell}_t^\top \nabla^{-2}\psi(p_t)\widehat{\ell}_t = \sum_i p_{t,i}^2 \widehat{\ell}_{t,i}^2 \xrightarrow{\mathbb{E}} \langle p_t, \ell_t \rangle$
- $\mathbb{E}[\text{Reg}] = \widetilde{\mathcal{O}}(\sqrt{K\mathbb{E}[\sum_t \langle p_t, \ell_t \rangle]}) \Rightarrow \mathbb{E}[\text{Reg}] = \widetilde{\mathcal{O}}(\sqrt{KL^\star})$

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

- $\psi(p^\star) - \min_p \psi(p) \leq K \ln T$ as before

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

- $\psi(p^\star) - \min_p \psi(p) \leq K \ln T$ as before
- use variance-reduced estimators $\widehat{\ell}_{t,i} = \frac{\ell_{t,i} - m_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\} + m_{t,i}$

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

- $\psi(p^\star) - \min_p \psi(p) \le K \ln T$ as before
- use variance-reduced estimators $\widehat{\ell}_{t,i} = \frac{\ell_{t,i} - m_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\} + m_{t,i}$
- $\|\widehat{\ell}_t - m_t\|_{p_t}^2 = \sum_i p_{t,i}^2 (\widehat{\ell}_{t,i} - m_{t,i})^2$

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^K \ln p_i$,

$$p_t = \operatorname*{argmin}_{p\in\Delta_K} \left\langle p, m_t + \sum_{\tau<t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^T \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^T \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

- $\psi(p^\star) - \min_p \psi(p) \le K \ln T$ as before
- use variance-reduced estimators $\widehat{\ell}_{t,i} = \frac{\ell_{t,i} - m_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\} + m_{t,i}$
- $\|\widehat{\ell}_t - m_t\|_{p_t}^2 = \sum_i p_{t,i}^2 (\widehat{\ell}_{t,i} - m_{t,i})^2 = (\ell_{t,i_t} - m_{t,i_t})^2$

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

- $\psi(p^\star) - \min_p \psi(p) \leq K \ln T$ as before
- use variance-reduced estimators $\widehat{\ell}_{t,i} = \frac{\ell_{t,i} - m_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\} + m_{t,i}$
- $\|\widehat{\ell}_t - m_t\|_{p_t}^2 = \sum_i p_{t,i}^2 (\widehat{\ell}_{t,i} - m_{t,i})^2 = (\ell_{t,i_t} - m_{t,i_t})^2 \leq |\ell_{t,i_t} - m_{t,i_t}|$

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

- $\psi(p^\star) - \min_p \psi(p) \leq K \ln T$ as before
- use variance-reduced estimators $\widehat{\ell}_{t,i} = \frac{\ell_{t,i} - m_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\} + m_{t,i}$
- $\|\widehat{\ell}_t - m_t\|_{p_t}^2 = \sum_i p_{t,i}^2 (\widehat{\ell}_{t,i} - m_{t,i})^2 = (\ell_{t,i_t} - m_{t,i_t})^2 \leq |\ell_{t,i_t} - m_{t,i_t}|$
- let $m_{t,i}$ be the **most recently observed loss** for arm i,

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

- $\psi(p^\star) - \min_p \psi(p) \leq K \ln T$ as before
- use variance-reduced estimators $\widehat{\ell}_{t,i} = \frac{\ell_{t,i} - m_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\} + m_{t,i}$
- $\|\widehat{\ell}_t - m_t\|_{p_t}^2 = \sum_i p_{t,i}^2 (\widehat{\ell}_{t,i} - m_{t,i})^2 = (\ell_{t,i_t} - m_{t,i_t})^2 \leq |\ell_{t,i_t} - m_{t,i_t}|$
- let $m_{t,i}$ be the **most recently observed loss** for arm i, then $\sum_t |\ell_{t,i_t} - m_{t,i_t}| = \sum_i \sum_{t: i_t = i} |\ell_{t,i} - m_{t,i}|$

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \underset{p \in \Delta_K}{\operatorname{argmin}} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

- $\psi(p^\star) - \min_p \psi(p) \leq K \ln T$ as before
- use variance-reduced estimators $\widehat{\ell}_{t,i} = \frac{\ell_{t,i} - m_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\} + m_{t,i}$
- $\|\widehat{\ell}_t - m_t\|_{p_t}^2 = \sum_i p_{t,i}^2 (\widehat{\ell}_{t,i} - m_{t,i})^2 = (\ell_{t,i_t} - m_{t,i_t})^2 \leq |\ell_{t,i_t} - m_{t,i_t}|$
- let $m_{t,i}$ be the **most recently observed loss** for arm i, then $\sum_t |\ell_{t,i_t} - m_{t,i_t}| = \sum_i \sum_{t:i_t = i} |\ell_{t,i} - m_{t,i}| \leq \sum_{t,i} |\ell_{t,i} - \ell_{t-1,i}|$

Achieving Path-Length Bounds

Wei-Luo'18

Optimistic FTRL with the log-barrier regularizer $\psi(p) = -\sum_{i=1}^{K} \ln p_i$,

$$p_t = \operatorname*{argmin}_{p \in \Delta_K} \left\langle p, m_t + \sum_{\tau < t} \widehat{\ell}_\tau \right\rangle + \frac{1}{\eta}\psi(p)$$

$$\sum_{t=1}^{T} \left\langle p_t - p^\star, \widehat{\ell}_t \right\rangle \lesssim \frac{\psi(p_\star) - \min_p \psi(p)}{\eta} + \eta \sum_{t=1}^{T} \|\widehat{\ell}_t - m_t\|_{p_t}^2$$

- $\psi(p^\star) - \min_p \psi(p) \leq K \ln T$ as before
- use variance-reduced estimators $\widehat{\ell}_{t,i} = \frac{\ell_{t,i} - m_{t,i}}{p_{t,i}} \mathbf{1}\{i_t = i\} + m_{t,i}$
- $\|\widehat{\ell}_t - m_t\|_{p_t}^2 = \sum_i p_{t,i}^2 (\widehat{\ell}_{t,i} - m_{t,i})^2 = (\ell_{t,i_t} - m_{t,i_t})^2 \leq |\ell_{t,i_t} - m_{t,i_t}|$
- let $m_{t,i}$ be the **most recently observed loss** for arm i, then $\sum_t |\ell_{t,i_t} - m_{t,i_t}| = \sum_i \sum_{t: i_t = i} |\ell_{t,i} - m_{t,i}| \leq \sum_{t,i} |\ell_{t,i} - \ell_{t-1,i}|$
- $\mathbb{E}[\text{Reg}] = \widetilde{\mathcal{O}}\left(\sqrt{K \sum_{t,i} |\ell_{t,i} - \ell_{t-1,i}|}\right)$

More on Log-Barrier Regularizer

Surprisingly powerful for MAB and beyond:

More on Log-Barrier Regularizer

Surprisingly powerful for MAB and beyond:

- near-optimal behavior for stochastic losses (Wei-Luo'18, Ito'21)

More on Log-Barrier Regularizer

Surprisingly powerful for MAB and beyond:

- near-optimal behavior for stochastic losses (Wei-Luo'18, Ito'21)
- a tool to stabilize algorithm when combined with other regularizers

More on Log-Barrier Regularizer

Surprisingly powerful for MAB and beyond:

- near-optimal behavior for stochastic losses (Wei-Luo'18, Ito'21)
- a tool to stabilize algorithm when combined with other regularizers
  - log-barrier + Shannon entropy (Bubeck-Cohen-Li'18, Bubeck-Li-Luo-Wei'19, Lee-Luo-Zhang'20, Ito-Tsuchiya-Honda'22)
  - log-barrier + quadratic regularizer (Luo-Wei-Zheng'18)
  - log-barrier + Tsallis entropy (Pogodin-Lattimore'20, Jin-Huang-Luo'21)
  - log-barrier + Tsallis-Shannon entropy (Erez-Koren'21)

More on Log-Barrier Regularizer

Surprisingly powerful for MAB and beyond:

- near-optimal behavior for stochastic losses (Wei-Luo'18, Ito'21)
- a tool to stabilize algorithm when combined with other regularizers
  - log-barrier + Shannon entropy (Bubeck-Cohen-Li'18, Bubeck-Li-Luo-Wei'19, Lee-Luo-Zhang'20, Ito-Tsuchiya-Honda'22)
  - log-barrier + quadratic regularizer (Luo-Wei-Zheng'18)
  - log-barrier + Tsallis entropy (Pogodin-Lattimore'20, Jin-Huang-Luo'21)
  - log-barrier + Tsallis-Shannon entropy (Erez-Koren'21)
- if increase η occasionally, obtain negative regret term $-\frac{1}{\eta \min_t p_{t,i^\star}}$

More on Log-Barrier Regularizer

Surprisingly powerful for MAB and beyond:

- near-optimal behavior for stochastic losses (Wei-Luo'18, Ito'21)
- a tool to stabilize algorithm when combined with other regularizers
  - log-barrier + Shannon entropy (Bubeck-Cohen-Li'18, Bubeck-Li-Luo-Wei'19, Lee-Luo-Zhang'20, Ito-Tsuchiya-Honda'22)
  - log-barrier + quadratic regularizer (Luo-Wei-Zheng'18)
  - log-barrier + Tsallis entropy (Pogodin-Lattimore'20, Jin-Huang-Luo'21)
  - log-barrier + Tsallis-Shannon entropy (Erez-Koren'21)
- if increase η occasionally, obtain negative regret term $-\frac{1}{\eta \min_t p_{t,i^\star}}$
  - useful for combining bandit algorithms (notoriously difficult) (Agarwal-Luo-Neyshabur-Schapire'17)

More on Log-Barrier Regularizer

Surprisingly powerful for MAB and beyond:

- near-optimal behavior for stochastic losses (Wei-Luo'18, Ito'21)
- a tool to stabilize algorithm when combined with other regularizers
  - log-barrier + Shannon entropy (Bubeck-Cohen-Li'18, Bubeck-Li-Luo-Wei'19, Lee-Luo-Zhang'20, Ito-Tsuchiya-Honda'22)
  - log-barrier + quadratic regularizer (Luo-Wei-Zheng'18)
  - log-barrier + Tsallis entropy (Pogodin-Lattimore'20, Jin-Huang-Luo'21)
  - log-barrier + Tsallis-Shannon entropy (Erez-Koren'21)
- if increase η occasionally, obtain negative regret term $-\frac{1}{\eta \min_t p_{t,i^\star}}$
  - useful for combining bandit algorithms (notoriously difficult) (Agarwal-Luo-Neyshabur-Schapire'17)
  - useful for obtaining high prob. regret bounds (the only efficient and optimal way for linear bandits) (Lee-Luo-Wei-Zhang'20)

Open Questions on Adaptive Regret Bounds

Is $\mathcal{O}(\sqrt{L^\star K})$ achievable (without any logarithmic factors)?

Open Questions on Adaptive Regret Bounds

Is $\mathcal{O}(\sqrt{L^\star K})$ achievable (without any logarithmic factors)?

Is second-order path-length bound achievable?

Open Questions on Adaptive Regret Bounds

Is $\mathcal{O}(\sqrt{L^\star K})$ achievable (without any logarithmic factors)?

Is second-order path-length bound achievable?

- existing bounds are first-order, e.g. $\widetilde{\mathcal{O}}(\sqrt{K \sum_t \|\ell_t - \ell_{t-1}\|_\infty})$

Open Questions on Adaptive Regret Bounds

Is $\mathcal{O}(\sqrt{L^\star K})$ achievable (without any logarithmic factors)?

Is second-order path-length bound achievable?

- existing bounds are first-order, e.g. $\widetilde{\mathcal{O}}(\sqrt{K \sum_t \|\ell_t - \ell_{t-1}\|_\infty})$
- is $\widetilde{\mathcal{O}}(\text{poly}(K)\sqrt{\sum_t \|\ell_t - \ell_{t-1}\|_\infty^2})$ achievable?

Open Questions on Adaptive Regret Bounds

Is $\mathcal{O}(\sqrt{L^\star K})$ achievable (without any logarithmic factors)?

Is second-order path-length bound achievable?

- existing bounds are first-order, e.g. $\widetilde{\mathcal{O}}(\sqrt{K \sum_t \|\ell_t - \ell_{t-1}\|_\infty})$
- is $\widetilde{\mathcal{O}}(\text{poly}(K)\sqrt{\sum_t \|\ell_t - \ell_{t-1}\|_\infty^2})$ achievable? (yes for full-info)

Open Questions on Adaptive Regret Bounds

Is $\mathcal{O}(\sqrt{L^\star K})$ achievable (without any logarithmic factors)?

Is second-order path-length bound achievable?

- existing bounds are first-order, e.g. $\widetilde{\mathcal{O}}(\sqrt{K \sum_t \|\ell_t - \ell_{t-1}\|_\infty})$
- is $\widetilde{\mathcal{O}}(\text{poly}(K)\sqrt{\sum_t \|\ell_t - \ell_{t-1}\|_\infty^2})$ achievable? (yes for full-info)
- for games, first-order means $\frac{1}{T^{3/4}}$ convergence, second order means $\frac{1}{T}$

Open Questions on Adaptive Regret Bounds

Is $\mathcal{O}(\sqrt{L^\star K})$ achievable (without any logarithmic factors)?

Is second-order path-length bound achievable?

- existing bounds are first-order, e.g. $\widetilde{\mathcal{O}}(\sqrt{K \sum_t \|\ell_t - \ell_{t-1}\|_\infty})$
- is $\widetilde{\mathcal{O}}(\text{poly}(K)\sqrt{\sum_t \|\ell_t - \ell_{t-1}\|_\infty^2})$ achievable? (yes for full-info)
- for games, first-order means $\frac{1}{T^{3/4}}$ convergence, second order means $\frac{1}{T}$
- unknown even for $K = 2$ and $\sum_t \|\ell_t - \ell_{t-1}\|_\infty^2 = \mathcal{O}(1)$

Conclusions

Differences Between Full-Info and Bandits

	Full-info	**Bandit**

Differences Between Full-Info and Bandits

	Full-info	**Bandit**
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$

Differences Between Full-Info and Bandits

	Full-info	**Bandit**
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)

Differences Between Full-Info and Bandits

	Full-info	**Bandit**
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)
Switching costs		

- switching costs: $\mathrm{Reg} + \sum_t \mathbf{1}\{i_t \neq i_{t-1}\}$

Differences Between Full-Info and Bandits

	Full-info	**Bandit**
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)
Switching costs	$\Theta(\sqrt{T \ln K})$ Kalai-Vempala'05, Geulen-Vöcking-Winkler'10	

- switching costs: $\mathrm{Reg} + \sum_t \mathbf{1}\{i_t \neq i_{t-1}\}$

Differences Between Full-Info and Bandits

	Full-info	**Bandit**
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)
Switching costs	$\Theta(\sqrt{T \ln K})$ Kalai-Vempala'05, Geulen-Vöcking-Winkler'10	$\Theta(T^{2/3}K^{1/3})$ Dekel-Ding-Koren-Peres'14

- switching costs: $\text{Reg} + \sum_t \mathbf{1}\{i_t \neq i_{t-1}\}$

Differences Between Full-Info and Bandits

	Full-info	Bandit
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)
Switching costs	$\Theta(\sqrt{T \ln K})$ Kalai-Vempala'05, Geulen-Vöcking-Winkler'10	$\Theta(T^{2/3}K^{1/3})$ Dekel-Ding-Koren-Peres'14
Interval regret		

- switching costs: $\mathrm{Reg} + \sum_t \mathbf{1}\{i_t \neq i_{t-1}\}$
- interval regret: $\max_{i^\star} \sum_{\tau=s}^{t} (\ell_{\tau,i_\tau} - \ell_{\tau,i^\star})$ (for **unknown** $s \leq t$)

Differences Between Full-Info and Bandits

	Full-info	**Bandit**
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)
Switching costs	$\Theta(\sqrt{T \ln K})$ Kalai-Vempala'05, Geulen-Vöcking-Winkler'10	$\Theta(T^{2/3}K^{1/3})$ Dekel-Ding-Koren-Peres'14
Interval regret	$\sqrt{(t-s)\ln K},\ \forall s \le t$ Luo-Schapire'15, Daniely-Gonen-ShalevShwartz'15	

- switching costs: $\mathrm{Reg} + \sum_t \mathbf{1}\{i_t \neq i_{t-1}\}$
- interval regret: $\max_{i^\star} \sum_{\tau=s}^{t} (\ell_{\tau,i_\tau} - \ell_{\tau,i^\star})$ (for **unknown** $s \le t$)

Differences Between Full-Info and Bandits

	Full-info	**Bandit**
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)
Switching costs	$\Theta(\sqrt{T \ln K})$ Kalai-Vempala'05, Geulen-Vöcking-Winkler'10	$\Theta(T^{2/3}K^{1/3})$ Dekel-Ding-Koren-Peres'14
Interval regret	$\sqrt{(t-s)\ln K},\ \forall s \le t$ Luo-Schapire'15, Daniely-Gonen-ShalevShwartz'15	$\sqrt{(t-s)K}$ ✗ $\sqrt{TK}$ ✓ Daniely-Gonen-ShalevShwartz'15

- switching costs: $\text{Reg} + \sum_t \mathbf{1}\{i_t \neq i_{t-1}\}$
- interval regret: $\max_{i^\star} \sum_{\tau=s}^{t} (\ell_{\tau,i_\tau} - \ell_{\tau,i^\star})$ (for **unknown** $s \le t$)

Differences Between Full-Info and Bandits

	Full-info	**Bandit**
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)
Switching costs	$\Theta(\sqrt{T \ln K})$ Kalai-Vempala'05, Geulen-Vöcking-Winkler'10	$\Theta(T^{2/3}K^{1/3})$ Dekel-Ding-Koren-Peres'14
Interval regret	$\sqrt{(t-s)\ln K}, \; \forall s \le t$ Luo-Schapire'15, Daniely-Gonen-ShalevShwartz'15	$\sqrt{(t-s)K}$ ✗ $\sqrt{TK}$ ✓ Daniely-Gonen-ShalevShwartz'15
Switching regret		

- switching costs: $\mathrm{Reg} + \sum_t \mathbf{1}\{i_t \neq i_{t-1}\}$
- interval regret: $\max_{i^\star} \sum_{\tau=s}^{t} (\ell_{\tau,i_\tau} - \ell_{\tau,i^\star})$ (for **unknown** $s \le t$)
- switching regret: $\max_{i^\star_{1:T} \,:\, \sum_t \mathbf{1}\{i^\star_t \neq i^\star_{t-1}\} < S} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star_t})$ (for **unknown** S)

Differences Between Full-Info and Bandits

	Full-info	Bandit
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)
Switching costs	$\Theta(\sqrt{T \ln K})$ Kalai-Vempala'05, Geulen-Vöcking-Winkler'10	$\Theta(T^{2/3} K^{1/3})$ Dekel-Ding-Koren-Peres'14
Interval regret	$\sqrt{(t-s)\ln K}, \ \forall s \le t$ Luo-Schapire'15, Daniely-Gonen-ShalevShwartz'15	$\sqrt{(t-s)K}$ ✗ $\sqrt{TK}$ ✓ Daniely-Gonen-ShalevShwartz'15
Switching regret	$\sqrt{ST \ln K}, \ \forall S$ implication of interval regret	

- switching costs: $\mathrm{Reg} + \sum_t \mathbf{1}\{i_t \neq i_{t-1}\}$
- interval regret: $\max_{i^\star} \sum_{\tau=s}^{t} (\ell_{\tau,i_\tau} - \ell_{\tau,i^\star})$ (for **unknown** $s \le t$)
- switching regret: $\max_{i^\star_{1:T} \,:\, \sum_t \mathbf{1}\{i^\star_t \neq i^\star_{t-1}\} < S} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star_t})$ (for **unknown** S)

Differences Between Full-Info and Bandits

	Full-info	**Bandit**
Minimax regret	$\Theta(\sqrt{T \ln K})$	$\Theta(\sqrt{TK})$
Adaptive regret	small-loss, variance, path-length (second-order)	small-loss, variance, path-length (first-order)
Switching costs	$\Theta(\sqrt{T \ln K})$ Kalai-Vempala'05, Geulen-Vöcking-Winkler'10	$\Theta(T^{2/3}K^{1/3})$ Dekel-Ding-Koren-Peres'14
Interval regret	$\sqrt{(t-s)\ln K},\ \forall s \le t$ Luo-Schapire'15, Daniely-Gonen-ShalevShwartz'15	$\sqrt{(t-s)K}$ ✗ $\sqrt{TK}$ ✓ Daniely-Gonen-ShalevShwartz'15
Switching regret	$\sqrt{ST\ln K},\ \forall S$ implication of interval regret	$\sqrt{STK}$ (adaptive adversary) ✗ $\sqrt{STK}$ (oblivious adversary) ? $S\sqrt{TK}$ ✓ Marinov-Zimmert'21

- switching costs: $\mathrm{Reg} + \sum_t \mathbf{1}\{i_t \neq i_{t-1}\}$
- interval regret: $\max_{i^\star} \sum_{\tau=s}^{t} (\ell_{\tau,i_\tau} - \ell_{\tau,i^\star})$ (for **unknown** $s \le t$)
- switching regret: $\max_{i^\star_{1:T} \,:\, \sum_t \mathbf{1}\{i^\star_t \neq i^\star_{t-1}\} < S} \sum_{t=1}^{T} (\ell_{t,i_t} - \ell_{t,i^\star_t})$ (for **unknown** S)

Summary

Central techniques for adversarial MAB:

Summary

Central techniques for adversarial MAB:

- design algorithms for the full-info case first (using classical framework e.g. FTRL, Online Mirror Descent, or Follow-the-Perturbed-Leader)

Summary

Central techniques for adversarial MAB:

- design algorithms for the full-info case first (using classical framework e.g. FTRL, Online Mirror Descent, or Follow-the-Perturbed-Leader)
- design loss estimators for the bandit case

Summary

Central techniques for adversarial MAB:

- design algorithms for the full-info case first (using classical framework e.g. FTRL, Online Mirror Descent, or Follow-the-Perturbed-Leader)
- design loss estimators for the bandit case
- find the right combination of estimator and regularizer to **control variance** (using the local norm in the stability term)

Summary

Central techniques for adversarial MAB:

- design algorithms for the full-info case first (using classical framework e.g. FTRL, Online Mirror Descent, or Follow-the-Perturbed-Leader)
- design loss estimators for the bandit case
- find the right combination of estimator and regularizer to **control variance** (using the local norm in the stability term)

Applicable to many other online learning problems with partial info:

Summary

Central techniques for adversarial MAB:

- design algorithms for the full-info case first (using classical framework e.g. FTRL, Online Mirror Descent, or Follow-the-Perturbed-Leader)
- design loss estimators for the bandit case
- find the right combination of estimator and regularizer to **control variance** (using the local norm in the stability term)

Applicable to many other online learning problems with partial info:

- bandits with structures: combinatorial bandits, linear bandits, graph bandits, contextual bandits, convex bandits

Summary

Central techniques for adversarial MAB:

- design algorithms for the full-info case first (using classical framework e.g. FTRL, Online Mirror Descent, or Follow-the-Perturbed-Leader)
- design loss estimators for the bandit case
- find the right combination of estimator and regularizer to **control variance** (using the local norm in the stability term)

Applicable to many other online learning problems with partial info:

- bandits with structures: combinatorial bandits, linear bandits, graph bandits, contextual bandits, convex bandits
- partial monitoring (e.g. apple tasting, dynamic pricing)

Summary

Central techniques for adversarial MAB:

- design algorithms for the full-info case first (using classical framework e.g. FTRL, Online Mirror Descent, or Follow-the-Perturbed-Leader)
- design loss estimators for the bandit case
- find the right combination of estimator and regularizer to **control variance** (using the local norm in the stability term)

Applicable to many other online learning problems with partial info:

- bandits with structures: combinatorial bandits, linear bandits, graph bandits, contextual bandits, convex bandits
- partial monitoring (e.g. apple tasting, dynamic pricing)
- reinforcement learning (Markov decision processes)