Minimax Policies for Adversarial and Stochastic Bandits

Sébastien Bubeck[1]

joint work with Jean-Yves Audibert[2,3]

[1] INRIA Lille, SequeL team
[2] Univ. Paris Est, Imagine
[3] CNRS/ENS/INRIA, Willow project

Bandit game

Parameters: the number of arms (or actions) K and the number of rounds n.

For each round $t = 1, 2, \dots, n$

1. The forecaster chooses an arm $I_t \in \{1, \dots, K\}$, possibly with the help of an external randomization.
2. Simultaneously the adversary chooses a gain vector $g_t = (g_{1,t}, \dots, g_{K,t}) \in [0, 1]^K$.
3. The forecaster receives (and observes) the gain $g_{I_t,t}$.

Goal: Maximize the cumulative gains obtained. We consider the regret:

$$R_n = \max_{i=1,\dots,K} \mathbb{E} \sum_{t=1}^{n} g_{i,t} - \mathbb{E} \sum_{t=1}^{n} g_{I_t,t}.$$

Bandit game

Parameters: the number of arms (or actions) K and the number of rounds n.

For each round $t = 1, 2, \ldots, n$

1. The forecaster chooses an arm $I_t \in \{1, \ldots, K\}$, possibly with the help of an external randomization.
2. Simultaneously the adversary chooses a gain vector $g_t = (g_{1,t}, \ldots, g_{K,t}) \in [0,1]^K$.
3. The forecaster receives (and observes) the gain $g_{I_t,t}$.

Goal: Maximize the cumulative gains obtained. We consider the regret:

$$R_n = \max_{i=1,\ldots,K} \mathbb{E} \sum_{t=1}^{n} g_{i,t} - \mathbb{E} \sum_{t=1}^{n} g_{I_t,t}.$$

Bandit game

Parameters: the number of arms (or actions) K and the number of rounds n.

For each round $t = 1, 2, \ldots, n$

1. The forecaster chooses an arm $I_t \in \{1, \ldots, K\}$, possibly with the help of an external randomization.
2. Simultaneously the adversary chooses a gain vector $g_t = (g_{1,t}, \ldots, g_{K,t}) \in [0, 1]^K$.
3. The forecaster receives (and observes) the gain $g_{I_t,t}$.

Goal: Maximize the cumulative gains obtained. We consider the regret:

$$R_n = \max_{i=1,\ldots,K} \mathbb{E} \sum_{t=1}^{n} g_{i,t} - \mathbb{E} \sum_{t=1}^{n} g_{I_t,t}.$$

Bandit game

Parameters: the number of arms (or actions) K and the number of rounds n.

For each round $t = 1, 2, \dots, n$

1. The forecaster chooses an arm $I_t \in \{1, \dots, K\}$, possibly with the help of an external randomization.
2. Simultaneously the adversary chooses a gain vector $g_t = (g_{1,t}, \dots, g_{K,t}) \in [0, 1]^K$.
3. The forecaster receives (and observes) the gain $g_{I_t,t}$.

Goal: Maximize the cumulative gains obtained. We consider the regret:

$$R_n = \max_{i=1,\dots,K} \mathbb{E} \sum_{t=1}^{n} g_{i,t} - \mathbb{E} \sum_{t=1}^{n} g_{I_t,t}.$$

Bandit game

Parameters: the number of arms (or actions) K and the number of rounds n.

For each round $t = 1, 2, \ldots, n$

1. The forecaster chooses an arm $I_t \in \{1, \ldots, K\}$, possibly with the help of an external randomization.
2. Simultaneously the adversary chooses a gain vector $g_t = (g_{1,t}, \ldots, g_{K,t}) \in [0,1]^K$.
3. The forecaster receives (and observes) the gain $g_{I_t,t}$.

Goal: Maximize the cumulative gains obtained. We consider the regret:

$$R_n = \max_{i=1,\ldots,K} \mathbb{E} \sum_{t=1}^{n} g_{i,t} - \mathbb{E} \sum_{t=1}^{n} g_{I_t,t}.$$

Stochastic and adversial settings

- **Adversarial setting:** The adversary can freely choose the gain vector at each time step, that is the adversary is non-oblivious.
- **Stochastic setting:** The adversary samples the gain vector from an unknown product distribution $(\nu_1, \ldots, \nu_K)$ on $[0,1]^K$, that is $g_{i,t} \sim \nu_i$.

Stochastic and adversial settings

- **Adversarial setting:** The adversary can freely choose the gain vector at each time step, that is the adversary is non-oblivious.
- **Stochastic setting:** The adversary samples the gain vector from an unknown product distribution $(\nu_1, \dots, \nu_K)$ on $[0, 1]^K$, that is $g_{i,t} \sim \nu_i$.

Known results

- **Lower bound:** For both settings and for any strategy, $\sup R_n \geq \frac{1}{20}\sqrt{nK}$, [Auer et al. 02a].
- **Adversarial setting:** Exp3 satisfies $R_n \leq \sqrt{2nK \log K}$, [Auer et al. 02a].
- **Stochastic setting:** UCB satisfies $R_n \leq \sqrt{10nK \log n}$ and $R_n \leq 10 \sum_{i:\Delta_i>0} \frac{1}{\Delta_i} \log n$ where Δ_i is the difference between the mean of the best arm and the mean of arm i, [Auer et al. 02b].

Known results

- **Lower bound:** For both settings and for any strategy, $\sup R_n \geq \frac{1}{20}\sqrt{nK}$, [Auer et al. 02a].
- **Adversarial setting:** Exp3 satisfies $R_n \leq \sqrt{2nK \log K}$, [Auer et al. 02a].
- **Stochastic setting:** UCB satisfies $R_n \leq \sqrt{10nK \log n}$ and $R_n \leq 10 \sum_{i:\Delta_i>0} \frac{1}{\Delta_i} \log n$ where Δ_i is the difference between the mean of the best arm and the mean of arm i, [Auer et al. 02b].

Known results

- **Lower bound:** For both settings and for any strategy, $\sup R_n \geq \frac{1}{20}\sqrt{nK}$, [Auer et al. 02a].
- **Adversarial setting:** Exp3 satisfies $R_n \leq \sqrt{2nK \log K}$, [Auer et al. 02a].
- **Stochastic setting:** UCB satisfies $R_n \leq \sqrt{10nK \log n}$ and $R_n \leq 10 \sum_{i:\Delta_i>0} \frac{1}{\Delta_i} \log n$ where Δ_i is the difference between the mean of the best arm and the mean of arm i, [Auer et al. 02b].

Known results

- **Lower bound:** For both settings and for any strategy, $\sup R_n \geq \frac{1}{20}\sqrt{nK}$, [Auer et al. 02a].
- **Adversarial setting:** Exp3 satisfies $R_n \leq \sqrt{2nK \log K}$, [Auer et al. 02a].
- **Stochastic setting:** UCB satisfies $R_n \leq \sqrt{10nK \log n}$ and $R_n \leq 10 \sum_{i:\Delta_i>0} \frac{1}{\Delta_i} \log n$ where Δ_i is the difference between the mean of the best arm and the mean of arm i, [Auer et al. 02b].

MOSS (Minimax Optimal Strategy in the Stochastic setting)

- $T_i(t)$ = the number of pulls of arm i up to time t.
- $\widehat{X}_{i,t}$ = the empirical mean estimate of arm i at time t (that is based on $T_i(t)$ pulls).
- Classical UCB:

$$I_t = \arg\max_{i \in \{1,\ldots,K\}} \widehat{X}_{i,t-1} + \sqrt{\frac{2 \log t}{T_i(t-1)}}.$$

- MOSS:

$$I_t = \arg\max_{i \in \{1,\ldots,K\}} \widehat{X}_{i,t-1} + \sqrt{\frac{\max\left(\log \left(\frac{n}{K T_i(t-1)} \right), 0 \right)}{T_i(t-1)}}.$$

MOSS (Minimax Optimal Strategy in the Stochastic setting)

- $T_i(t)$ = the number of pulls of arm i up to time t.
- $\widehat{X}_{i,t}$ = the empirical mean estimate of arm i at time t (that is based on $T_i(t)$ pulls).
- Classical UCB:

$$I_t = \arg\max_{i \in \{1,\dots,K\}} \widehat{X}_{i,t-1} + \sqrt{\frac{2\log t}{T_i(t-1)}}.$$

- MOSS:

$$I_t = \arg\max_{i \in \{1,\dots,K\}} \widehat{X}_{i,t-1} + \sqrt{\frac{\max\left(\log\left(\frac{n}{KT_i(t-1)}\right), 0\right)}{T_i(t-1)}}.$$

MOSS (Minimax Optimal Strategy in the Stochastic setting)

- $T_i(t)$ = the number of pulls of arm i up to time t.
- $\widehat{X}_{i,t}$ = the empirical mean estimate of arm i at time t (that is based on $T_i(t)$ pulls).
- Classical UCB:

$$I_t = \arg \max_{i \in \{1,\dots,K\}} \widehat{X}_{i,t-1} + \sqrt{\frac{2 \log t}{T_i(t-1)}} .$$

- MOSS:

$$I_t = \arg \max_{i \in \{1,\dots,K\}} \widehat{X}_{i,t-1} + \sqrt{\frac{\max\left(\log\left(\frac{n}{K T_i(t-1)} \right), 0 \right)}{T_i(t-1)}} .$$

MOSS (Minimax Optimal Strategy in the Stochastic setting)

- $T_i(t)$ = the number of pulls of arm i up to time t.
- $\widehat{X}_{i,t}$ = the empirical mean estimate of arm i at time t (that is based on $T_i(t)$ pulls).
- Classical UCB:

$$I_t = \arg \max_{i \in \{1,\dots,K\}} \widehat{X}_{i,t-1} + \sqrt{\frac{2 \log t}{T_i(t-1)}} .$$

- MOSS:

$$I_t = \arg \max_{i \in \{1,\dots,K\}} \widehat{X}_{i,t-1} + \sqrt{\frac{\max\left(\log\left(\frac{n}{K T_i(t-1)} \right), 0 \right)}{T_i(t-1)}} .$$

Regret bound for MOSS

Theorem

In the stochastic setting MOSS satisfies $R_n \leq 49\sqrt{nK}$.

Theorem

In the stochastic setting MOSS satisfies

$$R_n \leq 23K \sum_{i:\Delta_i>0} \frac{\max\left(\log\left(\frac{n\Delta_i^2}{K}\right), 1\right)}{\Delta_i}.$$

Regret bound for MOSS

Theorem

In the stochastic setting MOSS satisfies $R_n \leq 49\sqrt{nK}$.

Theorem

In the stochastic setting MOSS satisfies

$$R_n \leq 23K \sum_{i:\Delta_i>0} \frac{\max\left(\log\left(\frac{n\Delta_i^2}{K}\right), 1\right)}{\Delta_i}.$$

INF (Implicitly Normalized Forecaster)

Parameter: function $\psi : \mathbb{R}_-^* \to \mathbb{R}_+^*$ increasing, convex, twice continuously differentiable, and such that

$$\lim_{x \to -\infty} \psi(x) < 1/K, \qquad \text{and} \qquad \lim_{x \to 0} \psi(x) \geq 1.$$

Let p_1 be the uniform distribution over $\{1, \ldots, K\}$.

For each round $t = 1, 2, \ldots,$

1. $I_t \sim p_t$.
2. Compute $\tilde{g}_{i,t} = \frac{g_{i,t}}{p_{i,t}} \mathbb{1}_{I_t = i}$ and $\tilde{G}_{i,t} = \sum_{s=1}^{t} \tilde{g}_{i,s}$.
3. Compute the new probability distribution:

$$p_{i,t+1} = \psi(\tilde{G}_{i,t} - C_t)$$

where C_t is the unique real number such that $\sum_{i=1}^{K} p_{i,t+1} = 1$.

INF (Implicitly Normalized Forecaster)

Parameter: function $\psi : \mathbb{R}_-^* \to \mathbb{R}_+^*$ increasing, convex, twice continuously differentiable, and such that

$$\lim_{x \to -\infty} \psi(x) < 1/K, \qquad \text{and} \qquad \lim_{x \to 0} \psi(x) \geq 1.$$

Let p_1 be the uniform distribution over $\{1, \ldots, K\}$.

For each round $t = 1, 2, \ldots,$

1. $I_t \sim p_t$.
2. Compute $\tilde{g}_{i,t} = \frac{g_{i,t}}{p_{i,t}} \mathbb{1}_{I_t = i}$ and $\tilde{G}_{i,t} = \sum_{s=1}^{t} \tilde{g}_{i,s}$.
3. Compute the new probability distribution:

$$p_{i,t+1} = \psi(\tilde{G}_{i,t} - C_t)$$

where C_t is the unique real number such that $\sum_{i=1}^{K} p_{i,t+1} = 1$.

INF (Implicitly Normalized Forecaster)

Parameter: function $\psi : \mathbb{R}_-^* \to \mathbb{R}_+^*$ increasing, convex, twice continuously differentiable, and such that

$$\lim_{x \to -\infty} \psi(x) < 1/K, \qquad \text{and} \qquad \lim_{x \to 0} \psi(x) \geq 1.$$

Let p_1 be the uniform distribution over $\{1, \ldots, K\}$.

For each round $t = 1, 2, \ldots,$

1. $I_t \sim p_t$.
2. Compute $\tilde{g}_{i,t} = \frac{g_{i,t}}{p_{i,t}} \mathbb{1}_{I_t = i}$ and $\tilde{G}_{i,t} = \sum_{s=1}^{t} \tilde{g}_{i,s}$.
3. Compute the new probability distribution:

$$p_{i,t+1} = \psi(\tilde{G}_{i,t} - C_t)$$

where C_t is the unique real number such that $\sum_{i=1}^{K} p_{i,t+1} = 1$.

INF (Implicitly Normalized Forecaster)

Parameter: function $\psi : \mathbb{R}_-^* \to \mathbb{R}_+^*$ increasing, convex, twice continuously differentiable, and such that

$$\lim_{x \to -\infty} \psi(x) < 1/K, \qquad \text{and} \qquad \lim_{x \to 0} \psi(x) \geq 1.$$

Let p_1 be the uniform distribution over $\{1, \ldots, K\}$.

For each round $t = 1, 2, \ldots,$

1. $I_t \sim p_t$.
2. Compute $\tilde{g}_{i,t} = \frac{g_{i,t}}{p_{i,t}} \mathbb{1}_{I_t = i}$ and $\tilde{G}_{i,t} = \sum_{s=1}^{t} \tilde{g}_{i,s}$.
3. Compute the new probability distribution:

$$p_{i,t+1} = \psi(\tilde{G}_{i,t} - C_t)$$

where C_t is the unique real number such that $\sum_{i=1}^{K} p_{i,t+1} = 1$.

Examples

1. $\psi(x) = \exp(\eta x) + \frac{\gamma}{K}$ with $\eta > 0$ and $\gamma \in [0, 1)$; this corresponds exactly to the EXP3 strategy.
2. $\psi(x) = \left(\frac{\eta}{x}\right)^q + \frac{\gamma}{K}$ with $q > 1$, $\eta > 0$ and $\gamma \in [0, 1)$; this is a new forecaster which will be proved to be minimax optimal for appropriate parameters.

Examples

1. $\psi(x) = \exp(\eta x) + \frac{\gamma}{K}$ with $\eta > 0$ and $\gamma \in [0,1)$; this corresponds exactly to the EXP3 strategy.
2. $\psi(x) = \left(\frac{\eta}{x}\right)^q + \frac{\gamma}{K}$ with $q > 1$, $\eta > 0$ and $\gamma \in [0,1)$; this is a new forecaster which will be proved to be minimax optimal for appropriate parameters.

Regret bound for INF

Theorem

For any real $q > 1$, *the Implicitly Normalized Forecaster with* $\psi(x) = \frac{1}{K}\left(\frac{9\sqrt{qnK}}{-x}\right)^q + \frac{q^{q/(2q-2)}}{\sqrt{nK}}$ *satisfies*

$$R_n \leq \frac{37}{1 - 1/q}\sqrt{qnK}.$$

In particular for $q = 3$ *we get* $R_n \leq 100\sqrt{nK}$.

Regret bound for INF

Theorem

For any real $q > 1$*, the Implicitly Normalized Forecaster with* $\psi(x) = \frac{1}{K}\left(\frac{9\sqrt{qnK}}{-x}\right)^q + \frac{q^{q/(2q-2)}}{\sqrt{nK}}$ *satisfies*

$$R_n \leq \frac{37}{1 - 1/q}\sqrt{qnK}.$$

In particular for $q = 3$ *we get* $R_n \leq 100\sqrt{nK}$*.*

Proof

By an Abel transform we shift the focus from:

$$\sum_{t=1}^{n} g_{I_t,t} = \sum_{t=1}^{n} \sum_{i=1}^{K} p_{i,t}(\tilde{G}_{i,t} - \tilde{G}_{i,t-1})$$

to

$$\sum_{t=1}^{n-1} \sum_{i=1}^{K} \tilde{G}_{i,t}(p_{i,t+1} - p_{i,t}) = \sum_{i=1}^{K} \sum_{t=1}^{n-1} \psi^{-1}(p_{i,t+1})(p_{i,t+1} - p_{i,t}).$$

Then a Taylor expansion gives us:

$$(p_{i,t+1} - p_{i,t})\psi^{-1}(p_{i,t+1}) = -\int_{p_{i,t+1}}^{p_{i,t}} \psi^{-1}(u)du + \frac{(p_{i,t} - p_{i,t+1})^2}{2\psi'(\psi^{-1}(\tilde{p}_{i,t+1}))}.$$

The first resulting term: $-\sum_{i=1}^{K} \int_{p_{i,n+1}}^{1/K} \psi^{-1}(u)du$ is easy to control. On the other hand for the second term we need to do a multivariate Taylor expansion on

Proof

By an Abel transform we shift the focus from:

$$\sum_{t=1}^{n} g_{I_t,t} = \sum_{t=1}^{n}\sum_{i=1}^{K} p_{i,t}(\tilde{G}_{i,t} - \tilde{G}_{i,t-1})$$

to

$$\sum_{t=1}^{n-1}\sum_{i=1}^{K} \tilde{G}_{i,t}(p_{i,t+1} - p_{i,t}) = \sum_{i=1}^{K}\sum_{t=1}^{n-1} \psi^{-1}(p_{i,t+1})(p_{i,t+1} - p_{i,t}).$$

Then a Taylor expansion gives us:

$$(p_{i,t+1} - p_{i,t})\psi^{-1}(p_{i,t+1}) = -\int_{p_{i,t+1}}^{p_{i,t}} \psi^{-1}(u)du + \frac{(p_{i,t} - p_{i,t+1})^2}{2\psi'(\psi^{-1}(\tilde{p}_{i,t+1}))}.$$

The first resulting term: $-\sum_{i=1}^{K}\int_{p_{i,n+1}}^{1/K} \psi^{-1}(u)du$ is easy to control. On the other hand for the second term we need to do a multivariate Taylor expansion on

Proof

By an Abel transform we shift the focus from:

$$\sum_{t=1}^{n} g_{I_t,t} = \sum_{t=1}^{n} \sum_{i=1}^{K} p_{i,t}(\tilde{G}_{i,t} - \tilde{G}_{i,t-1})$$

to

$$\sum_{t=1}^{n-1} \sum_{i=1}^{K} \tilde{G}_{i,t}(p_{i,t+1} - p_{i,t}) = \sum_{i=1}^{K} \sum_{t=1}^{n-1} \psi^{-1}(p_{i,t+1})(p_{i,t+1} - p_{i,t}).$$

Then a Taylor expansion gives us:

$$(p_{i,t+1} - p_{i,t})\psi^{-1}(p_{i,t+1}) = -\int_{p_{i,t+1}}^{p_{i,t}} \psi^{-1}(u)du + \frac{(p_{i,t} - p_{i,t+1})^2}{2\psi'(\psi^{-1}(\tilde{p}_{i,t+1}))}.$$

The first resulting term: $-\sum_{i=1}^{K} \int_{p_{i,n+1}}^{1/K} \psi^{-1}(u)du$ is easy to control. On the other hand for the second term we need to do a multivariate Taylor expansion on

$$p_{i,t} - p_{i,t+1} = \psi(\tilde{G}_{i,t} - C_t) - \psi(\tilde{G}_{i,t+1} - C_{t+1})$$

as well as a careful treatment of the "shift" introduced by $\tilde{p}$

Proof

By an Abel transform we shift the focus from:

$$\sum_{t=1}^{n} g_{I_t,t} = \sum_{t=1}^{n} \sum_{i=1}^{K} p_{i,t}(\tilde{G}_{i,t} - \tilde{G}_{i,t-1})$$

to

$$\sum_{t=1}^{n-1} \sum_{i=1}^{K} \tilde{G}_{i,t}(p_{i,t+1} - p_{i,t}) = \sum_{i=1}^{K} \sum_{t=1}^{n-1} \psi^{-1}(p_{i,t+1})(p_{i,t+1} - p_{i,t}).$$

Then a Taylor expansion gives us:

$$(p_{i,t+1} - p_{i,t})\psi^{-1}(p_{i,t+1}) = -\int_{p_{i,t+1}}^{p_{i,t}} \psi^{-1}(u)du + \frac{(p_{i,t} - p_{i,t+1})^2}{2\psi'(\psi^{-1}(\tilde{p}_{i,t+1}))}.$$

The first resulting term: $-\sum_{i=1}^{K} \int_{p_{i,n+1}}^{1/K} \psi^{-1}(u)du$ is easy to control. On the other hand for the second term we need to do a multivariate Taylor expansion on

$$p_{i,t} - p_{i,t+1} = \psi(\tilde{G}_{i,t} - C_t) - \psi(\tilde{G}_{i,t+1} - C_{t+1})$$

as well as a careful treatment of the "shift" introduced by $\tilde{p}_{i,t+1}$.

Proof

By an Abel transform we shift the focus from:

$$\sum_{t=1}^{n} g_{I_t,t} = \sum_{t=1}^{n} \sum_{i=1}^{K} p_{i,t}(\tilde{G}_{i,t} - \tilde{G}_{i,t-1})$$

to

$$\sum_{t=1}^{n-1} \sum_{i=1}^{K} \tilde{G}_{i,t}(p_{i,t+1} - p_{i,t}) = \sum_{i=1}^{K} \sum_{t=1}^{n-1} \psi^{-1}(p_{i,t+1})(p_{i,t+1} - p_{i,t}).$$

Then a Taylor expansion gives us:

$$(p_{i,t+1} - p_{i,t})\psi^{-1}(p_{i,t+1}) = -\int_{p_{i,t+1}}^{p_{i,t}} \psi^{-1}(u)du + \frac{(p_{i,t} - p_{i,t+1})^2}{2\psi'(\psi^{-1}(\tilde{p}_{i,t+1}))}.$$

The first resulting term: $-\sum_{i=1}^{K} \int_{p_{i,n+1}}^{1/K} \psi^{-1}(u)du$ is easy to control. On the other hand for the second term we need to do a multivariate Taylor expansion on

$$p_{i,t} - p_{i,t+1} = \psi(\tilde{G}_{i,t} - C_t) - \psi(\tilde{G}_{i,t+1} - C_{t+1})$$

as well as a careful treatment of the "shift" introduced by $\tilde{p}_{i,t+1}$.

Other notions of regret

1. In this work we considered

$$R_n = \max_{i=1,\dots,K} \mathbb{E} \sum_{t=1}^{n} g_{i,t} - \mathbb{E} \sum_{t=1}^{n} g_{I_t,t}.$$

However ultimately we want to control $\max_i \sum_{t=1}^n g_{i,t} - \sum_{t=1}^n g_{I_t,t}$ with high probability as well as $\mathbb{E} \max_i \sum_{t=1}^n g_{i,t} - \mathbb{E} \sum_{t=1}^n g_{I_t,t}$.

2. In fact if the adversary is oblivious then

$$\mathbb{E} \max_i \sum_{t=1}^{n} g_{i,t} - \max_i \mathbb{E} \sum_{t=1}^{n} g_{i,t} \leq \sqrt{\frac{n \log K}{2}}.$$

3. For non-oblivious adversary we set $\tilde{g}_{i,t} = \frac{g_{i,t} \mathbb{1}_{i=I_t} + \beta}{p_{i,t}}$. Then high probability bounds on $\max_i \sum_{t=1}^n g_{i,t} - \sum_{t=1}^n g_{I_t,t}$ follow as well as bounds on this quantity in expectation.

Other notions of regret

1. In this work we considered

$$R_n = \max_{i=1,\dots,K} \mathbb{E} \sum_{t=1}^{n} g_{i,t} - \mathbb{E} \sum_{t=1}^{n} g_{I_t,t}.$$

   However ultimately we want to control $\max_i \sum_{t=1}^{n} g_{i,t} - \sum_{t=1}^{n} g_{I_t,t}$ with high probability as well as $\mathbb{E} \max_i \sum_{t=1}^{n} g_{i,t} - \mathbb{E} \sum_{t=1}^{n} g_{I_t,t}$.

2. In fact if the adversary is oblivious then

$$\mathbb{E} \max_i \sum_{t=1}^{n} g_{i,t} - \max_i \mathbb{E} \sum_{t=1}^{n} g_{i,t} \leq \sqrt{\frac{n \log K}{2}}.$$

3. For non-oblivious adversary we set $\tilde{g}_{i,t} = \frac{g_{i,t} \mathbb{1}_{i=I_t} + \beta}{p_{i,t}}$. Then high probability bounds on $\max_i \sum_{t=1}^{n} g_{i,t} - \sum_{t=1}^{n} g_{I_t,t}$ follow as well as bounds on this quantity in expectation.

Other notions of regret

1. In this work we considered

$$R_n = \max_{i=1,\dots,K} \mathbb{E} \sum_{t=1}^{n} g_{i,t} - \mathbb{E} \sum_{t=1}^{n} g_{I_t,t}.$$

   However ultimately we want to control $\max_i \sum_{t=1}^n g_{i,t} - \sum_{t=1}^n g_{I_t,t}$ with high probability as well as $\mathbb{E} \max_i \sum_{t=1}^n g_{i,t} - \mathbb{E} \sum_{t=1}^n g_{I_t,t}$.

2. In fact if the adversary is oblivious then

$$\mathbb{E} \max_i \sum_{t=1}^{n} g_{i,t} - \max_i \mathbb{E} \sum_{t=1}^{n} g_{i,t} \leq \sqrt{\frac{n \log K}{2}}.$$

3. For non-oblivious adversary we set $\tilde{g}_{i,t} = \frac{g_{i,t} \mathbb{1}_{i=I_t} + \beta}{p_{i,t}}$. Then high probability bounds on $\max_i \sum_{t=1}^n g_{i,t} - \sum_{t=1}^n g_{I_t,t}$ follow as well as bounds on this quantity in expectation.

Extensions of INF to other games

1. The INF forecaster can be generalized to work in the classical full information game and the label efficient game (with bandit or full information).
2. One can also compute bounds on the regret in a "tracking the best expert" setting, that is we compare ourselves to a strategy allowed to switch S times between different arms (in this talk we considered the case $S = 0$).
3. All the proofs follow the same scheme !

Extensions of INF to other games

1. The INF forecaster can be generalized to work in the classical full information game and the label efficient game (with bandit or full information).
2. One can also compute bounds on the regret in a "tracking the best expert" setting, that is we compare ourselves to a strategy allowed to switch S times between different arms (in this talk we considered the case $S = 0$).
3. All the proofs follow the same scheme !

Extensions of INF to other games

1. The INF forecaster can be generalized to work in the classical full information game and the label efficient game (with bandit or full information).
2. One can also compute bounds on the regret in a "tracking the best expert" setting, that is we compare ourselves to a strategy allowed to switch S times between different arms (in this talk we considered the case $S = 0$).
3. All the proofs follow the same scheme !

Summary

	$\max_i \mathbb{E} \sum_{t=1}^{n} \left(g_{i,t} - g_{I_t,t}\right)$		$\mathbb{E} \max_i \sum_{t=1}^{n} \left(g_{i,t} - g_{I_t,t}\right)$	
	L.B.	U.B.	L.B.	U.B.
Full Information				
Label Efficient F.I.				
Oblivious Bandit	$\sqrt{nK}$	$\sqrt{nK}$	$\sqrt{nK}$	$\sqrt{nK}$
Non-Oblivious Bandit	$\sqrt{nK}$	$\sqrt{nK}$		
Label Efficient Bandit				
Tracking the Best Expert				

Summary

	$\max_i \mathbb{E} \sum_{t=1}^n (g_{i,t} - g_{I_t,t})$		$\mathbb{E} \max_i \sum_{t=1}^n (g_{i,t} - g_{I_t,t})$	
	L.B.	U.B.	L.B.	U.B.
Full Information	$\sqrt{n \log K}$	$\sqrt{n \log K}$	$\sqrt{n \log K}$	$\sqrt{n \log K}$
Label Efficient F.I.	$n\sqrt{\frac{\log K}{m}}$	$n\sqrt{\frac{\log K}{m}}$	$n\sqrt{\frac{\log K}{m}}$	$n\sqrt{\frac{\log K}{m}}$
Oblivious Bandit	$\sqrt{nK}$	$\sqrt{nK}$	$\sqrt{nK}$	$\sqrt{nK}$
Non-Oblivious Bandit	$\sqrt{nK}$	$\sqrt{nK}$	?	$\sqrt{nK \log K}$
Label Efficient Bandit	?	$n\sqrt{\frac{K}{m}}$	?	$n\sqrt{\frac{K \log K}{m}}$
Tracking the Best Expert	?	$\sqrt{nKS \log \frac{nK}{S}}$	?	$\sqrt{nKS \log \frac{nK}{S}}$